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

Assessing differential expression profiles and modeling allele-specific expression in leaves of *Saccharum* accessions contrasting in biomass production

Fernando Henrique Correr

Thesis presented to obtain the degree of Doctor in Science. Area: Genetics and Plant Breeding

Piracicaba
2021

Fernando Henrique Correr
Bachelor in Biotechnology

Assessing differential expression profiles and modeling allele-specific expression in leaves of *Saccharum* accessions contrasting in biomass production

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

Thesis presented to obtain the degree of Doctor in Science.
Area: Genetics and Plant Breeding

Piracicaba
2021

DEDICATORY

To my parents, Claudia and Valdecir,
and my girlfriend Maria Clara

ACKNOWLEDGEMENTS

I thank the University of São Paulo for my professional qualification. To the "Luiz de Queiroz" College of Agriculture (ESALQ), especially to the Genetics Department for supporting me during the graduate courses.

To the Brazilian National Council for Scientific and Technological Development (CNPq) and to the Coordination for the Improvement of Higher Education Personnel (CAPES) for providing the financial support to conduct my research.

Special acknowledgments go to my advisor Prof. Gabriel Rodrigues Alves Margarido for his guidance and support during these years. Thank you for teaching me not only about genetics, statistics and bioinformatics, but for showing me how to be an ethical professional. I deeply appreciate the effort you spent for keeping me engaged. Thank you for being an example of professor and scientist! I also extend my gratitude to his family, Maidy and Beatriz, for all attention you gave me in Brisbane.

My special thanks go to my friends from the Laboratory of Bioinformatics Applied to Bioenergy for the scientific discussion, friendship, talks and funny moments: Amanda Avelar de Oliveira, Amanda Ghelfi Dumit, Ana Letycia Basso Garcia, André Luís Patrício, Diane de Fátima Sgariboldi, Guilherme Bovi Ambrosano, Guilherme Kenichi Hosaka, Leonardo Sartori Menegatto, Lorena Guimarães Batista and Victor Hugo de Mello Pessoa. It is a pleasure to be your friend!

Special acknowledgement to Prof. Monalisa Sampaio Carneiro, who contributed to all the biological material, laboratory infrastructure and insights. Thank you for showing the beautiful of sugarcane genetics!

Additionally, I express my gratitude to the University of Queensland and the Queensland Alliance for Agriculture and Food Innovation (QAAFI), for all the support provided during my internship. Special acknowledgment goes to Prof. Robert Henry for the opportunity to work in his group. Thank you for your kindness in helping me and for the valuable discussions we had. I also express my gratitude to Dr. Agnelo Furtado, for his help and friendship. I extend my gratitude to all my friends in Robert's group: Adhini, Angela, Ardy, Bader, Joseph, Kalpani, Katrina, Onkar, Othman, Patrick, Prameela, Priyanka, Sharmin, Vallari and Virginie. I also thank Prof. David Braun for sharing his knowledge on sucrose transporters and for his friendship during his stay in Brisbane. My gratitude to all QAAFI staff, especially to Annie Morley, Annie Cox, and Maria Caldeira. I am also thankful to Tarnya, Dave and Restu for all the good moments at the "Trenton Trio". My gratitude to Julia, Elliot, Carolina, Diana, Fernanda, Wei-An Tsai and William (Bill), who were incredible friends during this period in Australia. I also wish to acknowledge The University of Queensland's Research Computing Centre (RCC) for its support in this research.

My sincere thanks go to all professors of the Department of Genetics, especially to Claudia Barros Monteiro Vitorello, Gerhard Bandel, Giancarlo Conde Xavier Oliveira, Margarida Lopes Rodrigues de Aguiar-Perecin and Maria Carolina Quecine Verdi. Thank you for all the lectures and for the important talks regarding not only science, but also for showing the history of genetics in Brazil. My special gratitude to Prof. Augusto Franco Garcia, who contributed enormously to this work and with my academic development. Sincere thanks to the Department's staff: Antonio de Padua Gorga (Berdan), Cândida Vanderléia de Oliveira, Carlos Roberto Macedonio, Fernando Leopoldino, Maídia Maria Thomaziello, Silvia Zanatta and Valdir Próspero.

Additionally, I would like to thank all members of GVENCK and GENt for providing an incredible and creative environment within the department. To my friends Maria Bonatelli, Júlia Morosini and Fernando Espolador. To Cristiane Taniguti, for your long-term friendship.

I am also grateful to all the collaborators during my PhD: Prof. Anete Pereira, Carla Silva, Danilo Sforça and Mariana Cia. To Eliana Garcia and Rodrigo Pessanha for all help with the thesis

submission. I would like also to thank Prof. Clícia Grativol Gaspar de Matos, Prof. Marcelo Carazzolle, Dr. Augusto Lima Diniz, Prof. Paulo José Pereira Lima Teixeira and Prof. Tiago Antônio de Oliveira Mendes for accepting the invitation to participate in my defense. And to all the friends, teachers and professors who have contributed to my academic and personal growth.

Thanks to my family, especially to my parents, Claudia and Valdecir and my grandparents, Maria Cecília and Orlando. To my girlfriend Maria Clara. Thank you for your PATIENCE, encouragement and unconditional love! I love you!

Finally, I thank God for the gift of life, for never letting me weaken or giving up during this journey.

If a thing be ordained to another as to its end, its last end cannot consist in the preservation of its being. Hence a captain does not intend as a last end, the preservation of the ship entrusted to him, since a ship is ordained to something else as its end, viz. to navigation.

St. Thomas Aquinas

RESUMO

Avaliação de perfis de expressão diferencial e modelagem da expressão alelo-específica em folhas de acessos de *Saccharum* contrastantes na produção de biomassa

A cana-de-açúcar é uma das mais importantes culturas agrícolas mundiais devido a seus principais produtos - açúcar e álcool -, o reuso de seus subprodutos e a capacidade de inovação de sua agroindústria. Apresenta um potencial para uma produção mais rentável e sustentável, que pode ser obtida pelo desenvolvimento de cultivares de alta produtividade. Por esse motivo, características além do teor de sacarose nos colmos devem ser exploradas. Recentemente, a chamada cana-energia fez com que os programas de melhoramento contemplassem características relacionadas à biomassa, como o conteúdo de fibra e a capacidade de perfilhamento. A variação genética associada a essas características pode ser melhorada pela inclusão de outros acessos de *Saccharum*, os quais ainda não foram explorados pelos melhoristas. Além disso, os estudos sobre os perfis de expressão gênica em diferentes grupos de genótipos ainda são limitados na literatura. Portanto, objetivou-se a avaliação dos transcriptomas das folhas de dois grupos de genótipos - alta e baixa biomassa - a fim de identificar genes ou alelos potencialmente envolvidos com o conteúdo de biomassa. Para esse objetivo, genótipos foram selecionados pela similaridade fenotípica, independentemente de suas classificações como cultivados ou selvagens. O estudo foi dividido em dois capítulos. No primeiro, o objetivo foi a identificação de genes diferencialmente expressos entre os grupos de biomassa e investigação dos perfis de expressão de genes co-expressos. Os resultados mostraram que a expressão gênica permitiu não só estudar a variabilidade entre os grupos, como também a variabilidade dentro de cada grupo. Apesar da similaridade fenotípica, o grupo de alta biomassa mostrou uma alta variabilidade entre seus acessos, o que resultou em número expressivo de genes diferencialmente expressos, muito maior do que a comparação intergrupo. Genes que codificam a sacarose sintase e proteínas relacionadas à síntese de sacarose foram ligeiramente mais expressas no grupo de baixa biomassa, enquanto que aqueles envolvidos com a síntese de compostos da parede celular foram significativamente menos expressos. Curiosamente, a análise de co-expressão revelou que a expressão de genes relacionados com a fotossíntese foi maior em todos os genótipos híbridos e em *Saccharum officinarum*. Mostrou-se, também, que diferentes níveis de quantificação possuem certa influência nas considerações biológicas desse tipo de estudo. No segundo capítulo, testou-se a expressão alelo-específica (ASE) em um subconjunto de amostras de *Saccharum*. Esses acessos - três híbridos, uma *S. officinarum* e duas *S. spontaneum* - foram genotipados através da técnica de genotipagem por sequenciamento, seguida das estimativas da ploidia e dosagens alélicas. Modelou-se, para cada polimorfismo, a probabilidade da expressão do alelo de referência por um modelo Beta-Binomial hierárquico, no qual as dosagens alélicas serviram de informação *a priori*. Os resultados revelaram que ASE afeta parte dos *loci* avaliados em *Saccharum*. Entretanto, nenhum termo funcional foi enriquecido com os genes que demonstram ASE. Este estudo foi a primeira visão geral da ocorrência de expressão alelo-específica em múltiplos genótipos de cana-de-açúcar. Ademais, o modelo hierárquico pode ser usado para avaliar ASE em outros organismos de ploidia mista.

Palavras-chave: *Saccharum*, Transcriptomas, Biomassa, Desbalanço alélico

ABSTRACT

Assessing differential expression profiles and modeling allele-specific expression in leaves of *Saccharum* accessions contrasting in biomass production

Sugarcane is one of the most important crops worldwide due to its main products - sugar and ethanol -, the reuse of byproducts and the innovation capability of the agroindustry. It offers the potential for a more profitable and sustainable production, which can be accomplished by developing high-yielding cultivars. For that reason, traits other than the sucrose content in culms should be explored. The so-called energy cane has recently moved the attention of breeding programs towards biomass-related traits such as fiber content and tillering capacity. The genetic variation associated with these traits can be enhanced with other *Saccharum* accessions that have not yet been explored by breeders. In addition, studies regarding gene expression profiles in diverse groups of genotypes still limited in the literature. Therefore, we aimed to assess the transcriptomes from leaves of two groups of genotypes - high and low biomass - to identify genes or alleles potentially involved with the biomass content. To achieve such goal, genotypes were selected based on their similar phenotypes, regardless of their classification as cultivated or wild. We divided this study into two chapters. In the first chapter, the aim was to identify differentially expressed genes between the biomass groups and to investigate the expression profiles of coexpressed genes. Our results showed that gene expression allowed to study beyond the variability between the contrasting groups, the variability within each group. Despite the phenotypic similarity, the high biomass group showed an impressive variability among its accessions, resulting in many differentially expressed genes (DEGs), much more than the intergroup comparison. Genes coding for sucrose synthase and proteins related to sucrose synthesis were slightly more expressed in the low biomass group, whereas genes involved with the synthesis of cell wall compounds were significantly less expressed. Interestingly, the coexpression analysis revealed that the expression of genes related to photosynthesis was higher in all hybrids and *Saccharum officinarum* genotypes. We also showed that different quantification levels have certain influence on the biological insights provided by this kind of study. In the second chapter, we tested for allele-specific expression (ASE) in a subset of the *Saccharum* samples. These accessions - three hybrids, a *S. officinarum* and two *S. spontaneum* - were genotyped via genotyping-by-sequencing, followed by the estimation of ploidy and allelic dosages. We then modeled, for each polymorphism, the probability of expressing the reference allele using a hierarchical Beta-Binomial model, where allelic dosages served as prior information. Results revealed that ASE affects part of the loci assessed in *Saccharum*. However, any functional term was enriched with genes showing ASE. This study was the first global view of allele-specific expression in multiple genotypes of sugarcane. Furthermore, the hierarchical model can be used to evaluate ASE in other mixed-ploidy organisms.

Keywords: *Saccharum*, Transcriptome, Biomass, Allelic imbalance

1 INTRODUCTION

Sugarcane is an important crop in Brazil since the Portuguese colonization to produce sugar and, for approximately 50 years, to produce ethanol. Recently, data from the Brazilian Sugarcane Industry Association (UNICA) shows that sugarcane is planted in more than 5.5 million hectares in the State of São Paulo (http://www.unicadata.com.br/ - year 2018). According to the National Supply Company (CONAB), the Brazilian sugarcane production in the 2020/21 harvest is expected to increase in comparison with the previous year, reaching roughly 665.1 million tonnes [5]. While the total ethanol production will be reduced by 7.9%, sugar production is estimated to increase by 40.4% (41.8 million tonnes). Progress in the sugarcane industry was partially achieved through breeding high-performance cultivars. Briefly, the sugarcane breeding process relies on crossing parental genotypes, selecting superior genotypes for traits with high variability, then evaluating clones in proper experimental designs for lower heritability traits and, finally, assessing the genotype-environment interactions in competition trials [20]. Breeders have focused on increasing plant productivity to supply the industrial needs of raw material. At the same time, a more effective production in the same cultivable area is desired for a more sustainable agriculture. Scortecci and colleagues [49] stress the importance of leveraging the genetic potential of cultivars to achieve high yields and reduce the natural resources consumed by the plant. Moreover, we should explore not only the variability of sugarcane cultivars, but also from other *Saccharum* species.

Sugarcane is taxonomically classified as belonging to the genus *Saccharum*, subtribe *Saccharinae*, of the Poaceae family. Six species have been studied for understanding the evolution in the genus. Among them, four can be classified as cultivable: *Saccharum officinarum* L., *S. barberi* Jeswiet, *S. sinense* Roxb. and *S. edule* Hassk [40, 55, 39]. The same authors classify the two remaining species as wild: *S. spontaneum* L. and *S. robustum* Brandes & Jeswiet ex Grassl. Due to their proximity and the possibility of intergeneric crossings, *Erianthus*, *Miscanthus*, *Narenga*, *Saccharum* and *Sclerostachya* form the *Saccharum* complex [40, 55, 39]. Historically, the main objective of sugarcane breeding was sucrose accumulation in culms using mostly *S. officinarum* accessions. Later, crossings with *S. spontaneum* were performed to introgress traits related to stress tolerance [53]. The recent development of a group of high-productivity cultivars - energy canes - directed the breeders' attention to biomass [22, 13, 21]. As stated in studies dating from the 80s [8, 31], energy canes should achieve high yields of both sugar and biomass. The development of such new genotypes demands genetic resources in terms of biomass-related traits, such as fiber content in culms and tillering capacity. Breeding programs can thus benefit from enhanced knowledge about the molecular basis of desired traits, obtained via molecular markers and genomic sequences [7].

The association between genotypic and phenotypic data is not trivial in sugarcane. All *Saccharum* are polyploids showing a large number of chromosomes, which is variable in different accessions of the same species [55, 51]. As a consequence of the interspecific hybridization and successive backcrosses with *S. officinarum*, the modern cultivars have a very complex genome. Most of the basic chromosome architectures ($x = 10$) are represented by approximately eight *S. officinarum* homologs, *S. spontaneum* chromosomes and a small proportion of recombinants between the two species [51]. During sugarcane breeding, other *Saccharum* species - *S. barberi*, *S. sinense* and *S. robustum* - had a minimum contribution [38, 52]. Multiple strategies were used to unravel its genome sequence [50, 58, 33]. Recently, Garsmeur and colleagues [16] published a mosaic genome assembly of a commercial hybrid; Zhang and colleagues [26] published the sequence of a tetraploid *S. spontaneum* genome; and Souza and colleagues [4] published the gene space assembly of a Brazilian hybrid. However, analyzing the sugarcane genome is still a difficult task when different *Saccharum* accessions are being studied. Approaches using transcriptomes are useful to investigate likely cellular functions of putative genes, aiming to obtain molecular markers from functional genomic regions. Pioneering initiatives paved the way for functional genomics in sugar-

cane. First, Carson and colleagues [43] assessed gene expression in sugarcane leaf rolls using expressed sequence tags (ESTs). Two years later, after assessing the transcriptome of sugarcane leaves, they found genes functionally associated with the control and maintenance of cellular metabolism, transport and response to stresses [41]. Afterwards, researchers in the SUCEST project obtained more than 200 thousand ESTs from different samples [57]. Differentially expressed genes related to cell wall, cellulose and lignin biosynthesis were identified among different stages of culm development via transcriptome profiling [9].

These functional genomics and physiological studies in sugarcane provided evidence of important genes related to sucrose accumulation and synthesis of structural compounds. Along with advances described in the literature for other plants, efforts have also been made to connect genes in pathways to understand carbon partitioning in sugarcane. Wang and colleagues [48] showed the main steps for this process, from sucrose synthesis to its distribution to the sink cells. They showed that after photosynthesis on sugarcane leaves, sucrose is translocated in the phloem and reaches the stem parenchyma cells through both symplast and apoplast. These authors also reported key enzymes for sucrose accumulation: i) sucrose phosphate synthase (SPS) synthesizing sucrose-P from fructose-6-P and UDP-glucose; ii) sucrose phosphate phosphatase (SPP) producing sucrose from sucrose-P; iii) sucrose synthase (SuSy) being responsible for a reversible reaction converting fructose and UDP-glucose to sucrose; iv) cell wall invertase hydrolyzing sucrose into hexoses in the apoplast. There are also other classes of invertases and transporters that participate in transferring hexoses and sucrose into the cellular compartments. In addition to the transport via symplast, hexoses are transported by carriers and resynthesized into sucrose in the cytoplasm. Curiously, *Saccharum* species accumulate similar levels of symplastic and apoplastic solutes [2]. However, in general, high fiber species - *S. robustum* and *S. spontaneum* - show higher percentages of insoluble solids than sucrose-rich *Saccharum*, which in turn present a higher content of soluble solids [46, 2]. It is worth mentioning that *S. spontaneum* has a higher content of starch in mature culms to probably meet metabolic demands, serving as a resource for tillering and when the plant is submitted to stress [46].

Attention has been devoted to understand the synthesis of cell wall compounds, as the fibrous part can now be used as raw material by the sugarcane industry. The cell wall can be used in diverse manners, such as a prime source of energy, as feedstock and to develop cellulose-based materials [1]. Regarding the structure, primary and secondary walls of grasses are formed mostly by cellulose, followed by hemicellulose - arabinan- and xylan-derived compounds -, phenolic compounds, pectins, proteins and silica [1, 47]. The composition varies in different developmental stages of the culm. While the hemicellulose content is higher in younger internodes, cellulose is higher in mature internodes [46]. The synthesis of these elements requires the action of enzymes coordinated in different molecular pathways. More than a hundred candidate genes were found to be significantly associated with different fiber composition traits [47]. For cellulose, UDP-glucose from SuSy reaction is used by a complex set of cellulose synthase proteins to synthesize the glucan chain [45, 48]. This is corroborated by the significant association of both SuSy and UDP-glucosyl transferase with cellulose [47]. The biosynthesis of lignin is carried out by many enzymes of the phenylpropanoid pathway. In this pathway, Jardim-Messeder and colleagues [44] defined a core set of genes involved in lignin biosynthesis from the following families: phenylalanine/tyrosine ammonia-lyase, 4-(hydroxy) cinnamoyl CoA ligase, cinnamate 4-hydroxylase, hydroxycinnamoyl CoA shikimate:quinate hydroxycinnamoyltransferase, ρ-coumaroyl shikimate:quinate 3´-hydroxylase, caffeoyl CoA O-methyltransferase, caffeic acid/5-hydroxyferulic acid O-methyltransferase, ferulic acid/coniferaldehyde/coniferyl alcohol 5-hydroxylase, (hydroxy)cinnamoyl CoA reductase and (hydroxy)cinnamyl alcohol dehydrogenase. Authors reported that the expression of genes of the biosynthesis of monolignols have both genotype- and tissue-specificity [46]. High-fiber *Saccharum* species - *S. robustum* and *S. spontaneum* - show more diverse lignin oligomers [46]. The set of 15 phenylpropanoid core genes showed increased expression levels according to culm development [44].

These authors also analyzed the haplotypes of these genes, revealing an uneven distribution in the *S. spontaneum* genome. However, they could identify similar distribution of *cis*-elements in the upstream region of different haplotypes of a gene. Transcription factors can bind to such regions an regulate the expression of members of the phenylpropanoid pathway. In fact, biosynthesis of secondary cell wall can be regulated by myeloblastosis (MYB) and NAC transcription factors, as they are correlated to genes acting on the synthesis of lignin, tricin and hemicellulose [45].

New sequencing technologies, the possibility of assembling transcriptomes *de novo* and the development of statistical methods led to a revolution in the analysis of transcriptomes. The so-called RNA-Sequencing [30] has allowed an increase in the number of characterized sugarcane transcripts, as well as the comparison between contrasting conditions. In 2014, Cardoso-Silva and colleagues [56] assembled the transcriptomes of six cultivars, discovering 5,272 new putative genes not found in the SUCEST database. These authors found genes related to sucrose accumulation and responses to diseases. In the same year, the transcriptomes of the cultivar SP80-3280, accessions of *S. officinarum* and *S. spontaneum* were investigated [10]. These authors showed a high number of *S. spontaneum*-specific transcripts related to stress, signal transduction and transcription factors in sugarcane leaves. They also found that 78.28% of the transcripts were expressed in all genotypes and suggested that major phenotypic differences may be due to reasons other than expression variation at the gene level, such as isoforms, allelic variation and polymorphisms. Later, more than 500 transcripts associated to carbohydrate metabolism and transport were identified in the transcriptome of a high-sucrose cultivar [17].

The advance of sequencing methods has allowed the identification of isoforms, their occurrence in different tissues, development stages or growth conditions. In sugarcane, libraries from different organs were combined: i) first, second and third visible dewlap leaves; ii) immature and mature roots; and iii) the third internode from the top and the third internode from the base [59]. They generated a *de novo* transcriptome using Illumina sequencing on samples of (iii) and the isoform sequencing (Iso-Seq) from Pacific Biosciences on (i), (ii) and (iii) to identify isoforms. The *de novo* assembled transcriptome had a higher percentage of read alignment, more predicted proteins with homology to Viridiplantae and allowed the discovery of a larger number of KEGG pathways. Iso-Seq, on the other hand, recovered more complete transcripts, which aligned better to the *Sorghum bicolor* genome [59]. These results indicate the potential of Iso-Seq for comparative analyses.

Gene expression can be quantified after mapping reads to the transcripts from which they were originated. RNA-Sequencing has the potential to capture the dynamism of expressed genes from a population of cells, in a given experimental condition, creating the base for differential expression studies [28, 19, 30]. Gene expression data were also used to compare genotypes with different biomass content, aiming to identify transcripts related to carbon partitioning and to precursors of fiber components. Vicentini and collaborators [54] compared two cultivars showing 4% difference in lignin content. They identified more than 2,000 differentially expressed genes (DEGs), with four main distinct expression profiles and more than 100 groups of genes with similar expression. Among the DEGs, authors reported enrichment of the phenylpropanoid pathway, glutathione-S-transferases, trehalose metabolism, cell-wall proteins, response to biotic stresses and plant hormones. Instead of using clonal replicates of single genotypes to represent a given phenotypic group, Kasirajan and colleagues [32] compared two groups of genotypes with contrasting lignin content. They found DEGs more expressed in the high-fiber genotypes that were present in the phenylpropanoid pathway - lignin precursors - and associated with carbohydrate metabolism. However, by only using elite germplasm these articles exploit little existing variability for fiber content and, consequently, for biomass yield.

There are also other approaches to use the expression data provided by these high-throughput methods. One strategy is not to focus on expression at the gene level, but to look for differentially expressed transcripts and characterize splicing events. A second procedure is to assess the variation in

expression levels among the alleles of a gene. In that case, differences in the expression magnitude of two alleles can indicate allele-specific expression (ASE). This phenomenon can be explained by *cis*-regulation on promoter regions, frameshift mutations and epigenetic modifications that result on higher expression of one allele [24]. To evaluate ASE, polymorphisms have to be detected and allelic quantification should be obtained from RNA-Seq reads [14, 35, 23]. Then, for each polymorphism, a statistical test can be performed to detect allelic imbalance, by checking for deviations from equivalent expression between the alleles [12, 36]. ASE has been commonly assessed in large scale projects, mostly in human genetics [27, 37, 36, 35]. For example, a higher genic dosage caused by structural variations resulting from tumors was directly associated to increased allelic imbalance [36]. Recently, Lee and collaborators [37] found genes with allele-specific expression related to autism spectrum disorder risk. This approach has been used also in plants [14, 18, 42] and can be explored in other species.

As stated previously, ASE studies jointly use genotypic and expression data, which is feasible for sugarcane. Mancini and collaborators [7] discuss the main advances in sugarcane genetics and genomics. One of the most important is the use of SNPs to estimate the doses of the sugarcane alleles [11]. The high abundance of such markers is important for detecting a large number of polymorphisms, which are used to build genetic maps, discover QTLs and genomic regions associated with a given trait. It also opens the possibility for integration with expression data. A diverse set of *Saccharum* accessions was established in the Federal University of São Carlos (UFSCar), where researchers of the sugarcane breeding program laid out the Brazilian Panel of Sugarcane Genotypes [29, 3]. It is composed by 254 genotypes, representing wild species, cultivars with historic relevance and more recent cultivars. Some authors have already benefited from the genotyping of the panel [15, 29, 11, 3]. Using quantitative genotyping pipelines to obtain SNPs, the relative allelic proportions can be also estimated in this complex crop [11, 25, 34]. Then, the combination of such data with RNA-Sequencing provides enough information to evaluate ASE in sugarcane. However, a careful examination of the data is needed, as biases in the procedures - mapping and genotyping - can result in false ASE [24].

In this context, we point that it is feasible to understand, at the transcript level, differences between groups of accessions contrasting in their biomass content. In addition, investigating allelic imbalance can provide complementary results to the conventional analyses of gene profiles [36]. We explored gene expression data from leaves of twelve *Saccharum* accessions, phenotypically clustered in high- and low-biomass groups. First, we aimed to explore the variation between and within the groups in terms of differential gene expression. Next, we investigated the extent to which ASE occurred in both wild and cultivated accessions. We present and discuss our main findings regarding these objectives in two thesis chapters. The first chapter contains the investigation of differential gene expression, which was published in BMC Genomics [6]. We kept the integrity of all sections of this manuscript, including all the main and supplementary information. The second chapter focuses on the development of a model to test for ASE in complex polyploids such as sugarcane. It is also organized as a manuscript to be submitted.

References

[1] de Oliveira Buanafina MM, Cosgrove DJ. Cell Walls: Structure and Biogenesis. In: Sugarcane: Physiology, Biochemistry, and Functional Biology. Chichester, UK: John Wiley & Sons Ltd; 2013. p. 307–329. Available from: http://doi.wiley.com/10.1002/9781118771280.ch13.

[2] Welbaum GE. Water Relations and Cell Expansion of Storage Tissue. In: Sugarcane: Physiology, Biochemistry, and Functional Biology. Chichester, UK: John Wiley & Sons Ltd; 2013. p. 197–220. Available from: http://doi.wiley.com/10.1002/9781118771280.ch9.

[3] Medeiros C, Balsalobre TWA, Carneiro MS. Molecular diversity and genetic structure of Saccharum complex accessions. PLOS ONE. 2020 may;15(5):e0233211. Available from: http://dx.doi.org/10.1371/journal.pone.0233211https://dx.plos.org/10.1371/journal.pone.0233211.

[4] Souza GM, Van Sluys MA, Lembke CG, Lee H, Margarido GRA, Hotta CT, et al. Assembly of the 373k gene space of the polyploid sugarcane genome reveals reservoirs of functional diversity in the world's leading biomass crop. GigaScience. 2019 dec;8(12):1–18. Available from: https://academic.oup.com/gigascience/article/doi/10.1093/gigascience/giz129/5647371.

[5] Companhia Nacional de Abastecimento (CONAB). Acompanhamento da Safra Brasileira de Cana-de-Açúcar – Terceiro Levantamento da safra 2020/21. Monitoramento agrícola – Cana-de-açúcar. 2020;7(3):1–62.

[6] Correr FH, Hosaka GK, Barreto FZ, Valadão IB, Balsalobre TWA, Furtado A, et al. Differential expression in leaves of Saccharum genotypes contrasting in biomass production provides evidence of genes involved in carbon partitioning. BMC Genomics. 2020 dec;21(1):673. Available from: https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-020-07091-y.

[7] Mancini MC, Cardoso-Silva CB, Costa EA, Marconi TG, Garcia AAF, De Souza AP. New Developments in Sugarcane Genetics and Genomics. In: Buckeridge MS, De Souza AP, editors. Advances of Basic Science for Second Generation Bioethanol from Sugarcane. Cham: Springer International Publishing; 2017. p. 159–174. Available from: http://link.springer.com/10.1007/978-3-319-49826-3http://link.springer.com/10.1007/978-3-319-49826-3{_}9.

[8] Alexander AG. The energy cane alternative. Amsterdam, Netherlands: Elsevier Science Publishers B.V.; 1985.

[9] Casu RE, Jarmey JM, Bonnett GD, Manners JM. 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. 2007 feb;7(2):153–167. Available from: http://link.springer.com/10.1007/s10142-006-0038-z.

[10] Nishiyama MY, Ferreira SS, Tang PZ, Becker S, Pörtner-Taliana A, Souza GM. Full-length enriched cDNA libraries and ORFeome analysis of sugarcane hybrid and ancestor genotypes. PLoS ONE. 2014 sep;9(9):e107351. Available from: http://dx.plos.org/10.1371/journal.pone.0107351.

[11] Garcia AAF, Mollinari M, Marconi TG, Serang OR, Silva RR, Vieira MLC, et al. SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. Scientific Reports. 2013 dec;3(1):3399. Available from: http://www.nature.com/articles/srep03399.

[12] Wood DLA, Nones K, Steptoe A, Christ A, Harliwong I, Newell F, et al. Recommendations for accurate resolution of Gene and isoform allele-specific expression in RNA-seq data. PLoS ONE. 2015;10(5):1–27.

[13] Jackson PA. Breeding for improved sugar content in sugarcane. Field Crops Research. 2005 jun;92(2-3):277–290. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0378429005000365.

[14] Hu X, Wang H, Diao X, Liu Z, Li K, Wu Y, et al. Transcriptome profiling and comparison of maize ear heterosis during the spikelet and floret differentiation stages. BMC Genomics. 2016;17(1):1–18. Available from: http://dx.doi.org/10.1186/s12864-016-3296-8.

[15] Balsalobre TWA, da Silva Pereira G, Margarido GRA, Gazaffi R, Barreto FZ, Anoni CO, et al. GBS-based single dosage markers for linkage and QTL mapping allow gene mining for yield-related traits in sugarcane. BMC Genomics. 2017 dec;18(1):72. Available from: http://dx.doi.org/10.1186/s12864-016-3383-xhttp://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-3383-x.

[16] Garsmeur O, Droc G, Antonise R, Grimwood J, Potier B, Aitken K, et al. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. Nature Communications. 2018;9(1). Available from: http://dx.doi.org/10.1038/s41467-018-05051-5.

[17] Huang DL, Gao YJ, Gui YY, Chen ZL, Qin CX, Wang M, et al. Transcriptome of High-Sucrose Sugarcane Variety GT35. Sugar Tech. 2016;18(5):520–528.

[18] Ereful NC, Liu LY, Tsai E, Kao SM, Dixit S, Mauleon R, et al. Analysis of Allelic Imbalance in Rice Hybrids Under Water Stress and Association of Asymmetrically Expressed Genes with Drought-Response QTLs. Rice. 2016;9(1). Available from: http://dx.doi.org/10.1186/s12284-016-0123-4.

[19] Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett DE, et al. Characterization of the Yeast Transcriptome. Cell. 1997 jan;88(2):243–251. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0092867400818450.

[20] Gazaffi R, Oliveira KM, de Souza AP, Garcia AAF. Melhoramento genético e mapeamento da cana-de-açúcar. In: Cortez LAB, editor. Bioetanol de Cana-de-Açúcar: P&D para produtividade e sustentabilidade; 2010. p. 333–344.

[21] Cavalett O, Chagas MF, Junqueira TL, Watanabe MDB, Bonomi A. Environmental impacts of technology learning curve for cellulosic ethanol in Brazil. Industrial Crops and Products. 2017 nov;106:31–39. Available from: http://dx.doi.org/10.1016/j.indcrop.2016.11.025http://linkinghub.elsevier.com/retrieve/pii/S0926669016307695.

[22] Creste S, Xavier MA, Landell MGA. Importância do germoplasma no desenvolvimento de cultivares de cana-de-açúcar com perfil agroenergético. In: Cortez LAB, editor. Bioetanol de Cana-de-Açúcar: P&D para produtividade e sustentabilidade; 2010. p. 313–317.

[23] Romanel A, Lago S, Prandi D, Sboner A, Demichelis F. ASEQ: Fast allele-specific studies from next-generation sequencing data. BMC Medical Genomics. 2015;8(1):1–12.

[24] Castel SE, Levy-Moonshine A, Mohammadi P, Banks E, Lappalainen T. Tools and best practices for data processing in allelic expression analysis. Genome Biology. 2015;16(1):1–12. Available from: http://dx.doi.org/10.1186/s13059-015-0762-6.

[25] Pereira GS, Garcia AAF, Margarido GRA. A fully automated pipeline for quantitative genotype calling from next generation sequencing data in autopolyploids. BMC Bioinformatics. 2018;19(1):1–10.

[26] Zhang J, Zhang X, Tang H, Zhang Q, Hua X, Ma X, et al. Allele-defined genome of the autopolyploid sugarcane Saccharum spontaneum L. Nature Genetics. 2018;50(11):1565–1573.

[27] Degner JF, Marioni JC, Pai AA, Pickrell JK, Nkadori E, Gilad Y, et al. Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. Bioinformatics. 2009;25(24):3207–3212.

[28] Soneson C, Delorenzi M. A comparison of methods for differential expression analysis of RNA-seq data. BMC Bioinformatics. 2013;14(1):91. Available from: http://www.biomedcentral.com/1471-2105/14/91/abstract{\%}5Cnhttp://www.biomedcentral.com/1471-2105/14/91{\%}5Cnhttp://www.biomedcentral.com/content/pdf/1471-2105-14-91.pdfhttp://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-91.

[29] Barreto FZ, Rosa JRBF, Balsalobre TWA, Pastina MM, Silva RR, Hoffmann HP, et al. A genome-wide association study identified loci for yield component traits in sugarcane (Saccharum spp.). PLOS ONE. 2019 jul;14(7):e0219843. Available from: http://dx.plos.org/10.1371/journal.pone.0219843.

[30] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics. 2009;10(1):57–63.

[31] Matsuoka S, Kennedy AJ, dos Santos EGD, Tomazela AL, Rubio LCS. Energy Cane: Its Concept, Development, Characteristics, and Prospects. Advances in Botany. 2014;2014:1–13. Available from: http://www.hindawi.com/archive/2014/597275/.

[32] Kasirajan L, Hoang NV, Furtado A, Botha FC, Henry RJ. Transcriptome analysis highlights key differentially expressed genes involved in cellulose and lignin biosynthesis of sugarcane genotypes varying in fiber content. Scientific Reports. 2018;8(1):1–16.

[33] Riaño-Pachón DM, Mattiello L. Draft genome sequencing of the sugarcane hybrid SP80-3280. F1000Research. 2017;6(0):861. Available from: https://f1000research.com/articles/6-861/v1.

[34] Serang O, Mollinari M, Garcia AAF. Efficient exact maximum a posteriori computation for Bayesian SNP genotyping in polyploids. PLoS ONE. 2012;7(2):1–13.

[35] Rivas MA, Pirinen M, Conrad DF, Lek M, Tsang EK, Karczewski KJ, et al. Effect of predicted protein-truncating genetic variants on the human transcriptome. Science. 2015 may;348(6235):666–669. Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.1261877.

[36] Tuch BB, Laborde RR, Xu X, Gu J, Chung CB, Monighetti CK, et al. Tumor transcriptome sequencing reveals allelic expression imbalances associated with copy number alterations. PLoS ONE. 2010;5(2).

[37] Lee C, Kang EY, Gandal MJ, Eskin E, Geschwind DH. Profiling allele-specific gene expression in brains from individuals with autism spectrum disorder reveals preferential minor allele usage. Nature Neuroscience. 2019;22(9):1521–1532. Available from: http://dx.doi.org/10.1038/s41593-019-0461-9.

[38] Grivet L, Arruda P. Sugarcane genomics: Depicting the complex genome of an important tropical crop. Current Opinion in Plant Biology. 2002;5(2):122–127.

[39] Moore PH, Paterson AH, Tew T. Sugarcane: The Crop, the Plant, and Domestication. In: Sugarcane: Physiology, Biochemistry, and Functional Biology. Chichester, UK: John Wiley & Sons Ltd; 2013. p. 1–17. Available from: http://doi.wiley.com/10.1002/9781118771280.ch1.

[40] Daniels J, Roach BT. Taxonomy and evolution. In: Heinz DJ, editor. Developments in Crop Science. vol. 11 of Developments in Crop Science. Elsevier; 1987. p. 7–84. Available from: http://www.sciencedirect.com/science/article/pii/B9780444427694500072.

[41] Carson DL, Huckett BI, Botha FC. Differential gene expression in sugarcane leaf and internodal tissues of varying maturity. South African Journal of Botany. 2002 dec;68(4):434–442. Available from: c:{\%}5CDocumentsandSettings{\%}5Ccas128{\%}5CMyDocuments{\%}5CDownloadedpapers{\%}5CSAJB-2002-68-434-442.pdf{\%}5Cnhttp://linkinghub.elsevier.com/retrieve/pii/S0254629915303707https://linkinghub.elsevier.com/retrieve/pii/S0254629915303707.

[42] Pham GM, Newton L, Wiegert-Rininger K, Vaillancourt B, Douches DS, Buell CR. Extensive genome heterogeneity leads to preferential allele expression and copy number-dependent expression in cultivated potato. Plant Journal. 2017;92(4):624–637.

[43] Carson DL, Botha FC. Preliminary Analysis of Expressed Sequence Tags for Sugarcane. Crop Science. 2000;40(6):1769. Available from: https://www.crops.org/publications/cs/abstracts/40/6/1769.

[44] Jardim-Messeder D, Felix-Cordeiro T, Barzilai L, de Souza-Vieira Y, Galhego V, Bastos GA, et al. Genome-wide analysis of general phenylpropanoid and monolignol-specific metabolism genes in sugarcane. Functional & Integrative Genomics. 2021 jan;21(1):73–99. Available from: http://link.springer.com/10.1007/s10142-020-00762-9.

[45] Simões MS, Ferreira SS, Grandis A, Rencoret J, Persson S, Floh EIS, et al. Differentiation of Tracheary Elements in Sugarcane Suspension Cells Involves Changes in Secondary Wall Deposition and Extensive Transcriptional Reprogramming. Frontiers in Plant Science. 2020 dec;11(December):1–19. Available from: https://www.frontiersin.org/articles/10.3389/fpls.2020.617020/full.

[46] Llerena JPP, Figueiredo R, Brito MdS, Kiyota E, Mayer JLS, Araujo P, et al. Deposition of lignin in four species of Saccharum. Scientific Reports. 2019 dec;9(1):5877. Available from: http://dx.doi.org/10.1038/s41598-019-42350-3http://www.nature.com/articles/s41598-019-42350-3.

[47] Yang X, Todd J, Arundale R, Binder JB, Luo Z, Islam MS, et al. Identifying loci controlling fiber composition in polyploid sugarcane (Saccharum spp.) through genome-wide association study. Industrial Crops and Products. 2019 apr;130(January):598–605. Available from: https://doi.org/10.1016/j.indcrop.2019.01.023https://linkinghub.elsevier.com/retrieve/pii/S0926669019300305.

[48] Wang J, Nayak S, Koch K, Ming R. Carbon partitioning in sugarcane (Saccharum species). Frontiers in Plant Science. 2013;4(June):2005–2010. Available from: http://journal.frontiersin.org/article/10.3389/fpls.2013.00201/abstract.

[49] Scortecci KC, Creste S, Jr TC, Xavier MA, Landell MGA, Figueira A, et al. Challenges, Opportunities and Recent Advances in Sugarcane Breeding. In: Plant Breeding. InTech; 2012. p. 352. Available from: http://www.intechopen.com/books/plant-breeding/challenges-opportunities-and-recent-advances-in-sugarcane-breeding.

[50] Grativol C, Regulski M, Bertalan M, McCombie WR, Da Silva FR, Zerlotini Neto A, et al. Sugarcane genome sequencing by methylation filtration provides tools for genomic research in the genus Saccharum. Plant Journal. 2014 jul;79(1):162–172. Available from: http://doi.wiley.com/10.1111/tpj.12539.

[51] Piperidis N, D'Hont A. Sugarcane genome architecture decrypted with chromosome-specific oligo probes. The Plant Journal. 2020 jul:tpj.14881. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/tpj.14881.

[52] Piperidis G, Piperidis N, D'Hont A. Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. Molecular Genetics and Genomics. 2010 jul;284(1):65–73. Available from: http://link.springer.com/10.1007/s00438-010-0546-3.

[53] Matsuoka S, Ferro J, Arruda P. The Brazilian experience of sugarcane ethanol industry. In Vitro Cellular & Developmental Biology - Plant. 2009 jun;45(3):372–381. Available from: http://link.springer.com/10.1007/978-1-4419-7145-6{_}9http://link.springer.com/10.1007/s11627-009-9220-z.

[54] Vicentini R, Bottcher A, Dos Santos Brito M, Dos Santos AB, Creste S, De Andrade Landell MG, et al. Large-scale transcriptome analysis of two sugarcane genotypes contrasting for lignin content. PLoS ONE. 2015 aug;10(8):e0134909. Available from: http://dx.plos.org/10.1371/journal.pone.0134909.

[55] Irvine JE. Saccharum species as horticultural classes. Theoretical and Applied Genetics. 1999 feb;98(2):186–194. Available from: http://dx.doi.org/10.1007/s001220051057http://link.springer.com/10.1007/s001220051057.

[56] Cardoso-Silva CB, Costa EA, Mancini MC, Balsalobre TWA, Costa Canesin LE, Pinto LR, et al. De novo assembly and transcriptome analysis of contrasting sugarcane varieties. PLoS ONE. 2014;9(2).

[57] Vettore AL, da Silva FR, Kemper EL, Souza GM, da Silva AM, Ferro MIT, et al. Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. Genome Research. 2003 dec;13(12):2725–2735. Available from: http://www.genome.org/cgi/doi/10.1101/gr.1532103.

[58] Okura VK, de Souza RSC, de Siqueira Tada SF, Arruda P. BAC-Pool Sequencing and Assembly of 19 Mb of the Complex Sugarcane Genome. Frontiers in Plant Science. 2016 mar;7(March):342. Available from: http://journal.frontiersin.org/Article/10.3389/fpls.2016.00342/abstract.

[59] Hoang NV, Furtado A, Mason PJ, Marquardt A, Kasirajan L, Thirugnanasambandam PP, et al. A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. BMC Genomics. 2017;18(1):395. Available from: http://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-017-3757-8.

2 DIFFERENTIAL EXPRESSION IN LEAVES OF *Saccharum* GENOTYPES CONTRASTING IN BIOMASS PRODUCTION PROVIDES EVIDENCE OF GENES INVOLVED IN CARBON PARTITIONING

Abstract

Background: The development of biomass crops aims to meet industrial yield demands, in order to optimize profitability and sustainability. Achieving these goals in an energy crop like sugarcane relies on breeding for sucrose accumulation, fiber content and stalk number. To expand the understanding of the biological pathways related to these traits, we evaluated gene expression of two groups of genotypes contrasting in biomass composition.

Results: First visible dewlap leaves were collected from 12 genotypes, six per group, to perform RNA-Seq. We found a high number of differentially expressed genes, showing how hybridization in a complex polyploid system caused extensive modifications in genome functioning. We found evidence that differences in transposition and defense related genes may arise due to the complex nature of the polyploid *Saccharum* genomes. Genotypes within both biomass groups showed substantial variability in genes involved in photosynthesis. However, most genes coding for photosystem components or those coding for *phosphoenolpyruvate carboxylases* (PEPCs) were upregulated in the high biomass group. *Sucrose synthase* (SuSy) coding genes were upregulated in the low biomass group, showing that this enzyme class can be involved with sucrose synthesis in leaves, similarly to *sucrose phosphate synthase* (SPS) and *sucrose phosphate phosphatase* (SPP). Genes in pathways related to biosynthesis of cell wall components and *expansins* coding genes showed low average expression levels and were mostly upregulated in the high biomass group.

Conclusions: Together, these results show differences in carbohydrate synthesis and carbon partitioning in the source tissue of distinct phenotypic groups. Our data from sugarcane leaves revealed how hybridization in a complex polyploid system resulted in noticeably different transcriptomic profiles between contrasting genotypes.

Keywords: Sugarcane; Gene expression; Transcriptomics; RNA-Seq; Polyploid.

2.1 Conclusion

This work presented a broad view of the expression of many coding genes in sugarcane leaves of different genotypes. With regard to cell wall, most genes were upregulated in the high biomass group, but in general with low average expression levels. On the other hand, highly expressed genes involved in sucrose synthesis were upregulated in hybrids and *S. officinarum* genotypes. These results agree with current knowledge about the partitioning of carbohydrate to sucrose storage and maintenance of plant structure and metabolism in wild genotypes and modern cultivars. In addition, our research shows that investigating expression profiles in wild genotypes can enhance the understanding of genes selected through domestication and breeding. Expression profiles in other plant parts of wild and cultivated accessions are needed to provide knowledge about the action of the genes involved in carbohydrate metabolism and biomass production. Our data from sugarcane leaves revealed how hybridization in a complex polyploid system resulted in noticeably different transcriptomic profiles between contrasting genotypes.

References

[1] Piperidis N, D'Hont A. Sugarcane genome architecture decrypted with chromosome-specific oligo probes. The Plant Journal. 2020 jul:tpj.14881. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/tpj.14881.

[2] Osborn TC, Chris Pires J, Birchler JA, Auger DL, Chen ZJ, Lee HS, et al. Understanding mechanisms of novel gene expression in polyploids. Trends in Genetics. 2003;19(3):141–147.

[3] Qi X, Wang H, Song A, Jiang J, Chen S, Chen F. Genomic and transcriptomic alterations following intergeneric hybridization and polyploidization in the Chrysanthemum nankingense×Tanacetum vulgare hybrid and allopolyploid (Asteraceae). Horticulture Research. 2018 dec;5(1):5. Available from: http://dx.doi.org/10.1038/s41438-017-0003-0http://www.nature.com/articles/s41438-017-0003-0.

[4] Diniz AL, Ferreira SS, Ten-Caten F, Margarido GRA, dos Santos JM, Barbosa GVdS, et al. Genomic resources for energy cane breeding in the post genomics era. Computational and Structural Biotechnology Journal. 2019;17:1404–1414. Available from: https://doi.org/10.1016/j.csbj.2019.10.006.

[5] Chiniquy D, Sharma V, Schultink A, Baidoo EE, Rautengarten C, Cheng K, et al. XAX1 from glycosyltransferase family 61 mediates xylosyltransfer to rice xylan. Proceedings of the National Academy of Sciences. 2012 oct;109(42):17117–17122. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.1202079109.

[6] Driouich A, Follet-Gueye ML, Bernard S, Kousar S, Chevalier L, Vicré-Gibouin M, et al. Golgi-Mediated Synthesis and Secretion of Matrix Polysaccharides of the Primary Cell Wall of Higher Plants. Frontiers in Plant Science. 2012;3. Available from: http://journal.frontiersin.org/article/10.3389/fpls.2012.00079/abstract.

[7] McClintock B. The significance of responses of the genome to challenge. Science. 1984;226(4676):792–801.

[8] Chen ZJ. Genetic and Epigenetic Mechanisms for Gene Expression and Phenotypic Variation in Plant Polyploids. Annual Review of Plant Biology. 2007;58(1):377–406.

[9] Fedoroff NV, Bennetzen JL. Transposons, Genomic Shock, and Genome Evolution. In: Plant Transposons and Genome Dynamics in Evolution. Oxford, UK: Wiley-Blackwell; 2013. p. 181–201. Available from: http://doi.wiley.com/10.1002/9781118500156.ch10.

[10] Barreto FZ, Rosa JRBF, Balsalobre TWA, Pastina MM, Silva RR, Hoffmann HP, et al. A genome-wide association study identified loci for yield component traits in sugarcane (Saccharum spp.). PLOS ONE. 2019 jul;14(7):e0219843. Available from: http://dx.plos.org/10.1371/journal.pone.0219843.

[11] Ferreira SS, Hotta CT, Poelking VGdC, Leite DCC, Buckeridge MS, Loureiro ME, et al. Co-expression network analysis reveals transcription factors associated to cell wall biosynthesis in sugarcane. Plant Molecular Biology. 2016 may;91(1-2):15–35. Available from: http://link.springer.com/10.1007/s11103-016-0434-2.

[12] Sampedro J, Guttman M, Li LC, Cosgrove DJ. Evolutionary divergence of β-expansin structure and function in grasses parallels emergence of distinctive primary cell wall traits. Plant Journal. 2015;81(1):108–120.

[13] Labrou NE, Papageorgiou AC, Pavli O, Flemetakis E. Plant GSTome: structure and functional role in xenome network and plant stress response. Current Opinion in Biotechnology. 2015 apr;32:186–194. Available from: http://dx.doi.org/10.1016/j.copbio.2014.12.024https://linkinghub.elsevier.com/retrieve/pii/S0958166914002390.

[14] Fasano C, Diretto G, Aversano R, D'Agostino N, Di Matteo A, Frusciante L, et al. Transcriptome and metabolome of synthetic Solanum autotetraploids reveal key genomic stress events following polyploidization. New Phytologist. 2016 jun;210(4):1382–1394. Available from: http://doi.wiley.com/10.1111/nph.13878.

[15] Jackson S, Chen ZJ. Genomic and expression plasticity of polyploidy. Current Opinion in Plant Biology. 2010 apr;13(2):153–159. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1369526609001757.

[16] Yan W, Wu X, Li Y, Liu G, Cui Z, Jiang T, et al. Cell Wall Invertase 3 Affects Cassava Productivity via Regulating Sugar Allocation From Source to Sink. Frontiers in Plant Science. 2019;10(April):1–16.

[17] Vicient CM, Casacuberta JM. Impact of transposable elements on polyploid plant genomes. Annals of Botany. 2017;120(2):195–207.

[18] Alexander AG. The energy cane alternative. Amsterdam, Netherlands: Elsevier Science Publishers B.V.; 1985.

[19] Leal MRLV, Galdos MV, Scarpare FV, Seabra JEA, Walter A, Oliveira COF. Sugarcane straw availability, quality, recovery and energy use: A literature review. Biomass and Bioenergy. 2013;53:11–19.

[20] Rubin EM. Genomics of cellulosic biofuels. Nature. 2008;454(7206):841–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18704079.

[21] Yuan JS, Tiller KH, Al-Ahmad H, Stewart NR, Stewart CN. Plants to power: bioenergy to fuel the future. Trends in Plant Science. 2008;13(8):421–429.

[22] Keidar-Friedman D, Bariah I, Kashkush K. Genome-wide analyses of miniature inverted-repeat transposable elements reveals new insights into the evolution of the triticum-Aegilops group. PLoS ONE. 2018;13(10):1–23.

[23] Singh R, Jones T, Wai CM, Jifon J, Nagai C, Ming R, et al. Transcriptomic analysis of transgressive segregants revealed the central role of photosynthetic capacity and efficiency in biomass accumulation in sugarcane. Scientific Reports. 2018;8(1):1–10. Available from: http://dx.doi.org/10.1038/s41598-018-22798-5.

[24] Santiago TR, Pereira VM, de Souza WR, Steindorff AS, Cunha BADB, Gaspar M, et al. Genome-wide identification, characterization and expression profile analysis of expansins gene family in sugarcane (Saccharum spp.). PLoS ONE. 2018;13(1):1–18.

[25] Thirugnanasambandam PP, Mason PJ, Hoang NV, Furtado A, Botha FC, Henry RJ. Analysis of the diversity and tissue specificity of sucrose synthase genes in the long read transcriptome of sugarcane. BMC Plant Biology. 2019;19(1):160. Available from: https://bmcplantbiol.biomedcentral.com/articles/10.1186/s12870-019-1733-y.

[26] Verma I, Roopendra K, Sharma A, Chandra A, Kamal A. 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. 2019;25(1):207–220. Available from: https://doi.org/10.1007/s12298-018-0627-z.

[27] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics. 2008;9:559. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19114008{\%}0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2631488.

[28] Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research. 2016 feb;4:1521. Available from: https://f1000research.com/articles/4-1521/v2.

[29] Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. Nature Methods. 2017;14(4):417–419.

[30] Hoang NV, Furtado A, Donnan L, Keeffe EC, Botha FC, Henry RJ. High-Throughput Profiling of the Fiber and Sugar Composition of Sugarcane Biomass. Bioenergy Research. 2017;10(2):400–416.

[31] Carson DL, Botha FC. Preliminary Analysis of Expressed Sequence Tags for Sugarcane. Crop Science. 2000;40(6):1769. Available from: https://www.crops.org/publications/cs/abstracts/40/6/1769.

[32] Thirugnanasambandam PP, Hoang NV, Furtado A, Botha FC, Henry RJ. Association of variation in the sugarcane transcriptome with sugar content. BMC Genomics. 2017;18(1):1–22.

[33] Schwacke R, Ponce-Soto GY, Krause K, Bolger AM, Arsova B, Hallab A, et al. MapMan4: A Refined Protein Classification and Annotation Framework Applicable to Multi-Omics Data Analysis. Molecular Plant. 2019 jun;12(6):879–892. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1674205219300085.

[34] Burbano CS, Liu Y, Rösner KL, Reis VM, Caballero-Mellado J, Reinhold-Hurek B, et al. Predominant nifH transcript phylotypes related to Rhizobium rosettiformans in field-grown sugarcane plants and in Norway spruce. Environmental Microbiology Reports. 2011;3(3):383–389.

[35] da Silveira LCI, Brasileiro BP, Kist V, Daros E, Peternelli LA. Genetic diversity and coefficient of kinship among potential genitors for obtaining cultivars of energy cane. Revista Ciencia Agronomica. 2015;46(2):358–368.

[36] Urquiaga S, Xavier RP, de Morais RF, Batista RB, Schultz N, Leite JM, et al. Evidence from field nitrogen balance and 15N natural abundance data for the contribution of biological N2 fixation to Brazilian sugarcane varieties. Plant and Soil. 2012 jul;356(1-2):5–21. Available from: http://link.springer.com/10.1007/s11104-011-1016-3.

[37] Swapna M, Sivaraju K, Sharma RK, Singh NK, Mohapatra T. Single-Strand Conformational Polymorphism of EST-SSRs: A Potential Tool for Diversity Analysis and Varietal Identification in Sugarcane. Plant Molecular Biology Reporter. 2011 sep;29(3):505–513. Available from: http://link.springer.com/10.1007/s11105-010-0254-5.

[38] Pan YB, Burner DM, Legendre BL, Grisham MP, White WH. An assessment of the genetic diversity within a collection of Saccharum spontaneum L. with RAPD-PCR. Genetic Resources and Crop Evolution. 2005;51(8):895–903.

[39] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. 2005;102(43):15545–15550. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16199517{\%}0Ahttp://www.pnas.org/cgi/doi/10.1073/pnas.0506580102.

[40] Acevedo A, Tejedor MT, Erazzú LE, Cabada S, Sopena R. Pedigree comparison highlights genetic similarities and potential industrial values of sugarcane cultivars. Euphytica. 2017;213(6).

[41] Thirugnanasambandam PP, Hoang NV, Henry RJ. The Challenge of Analyzing the Sugarcane Genome. Frontiers in Plant Science. 2018;9(May):1–18. Available from: http://journal.frontiersin.org/article/10.3389/fpls.2018.00616/full.

[42] Schäfer WE, Rohwer JM, Botha FC. Partial purification and characterisation of sucrose synthase in sugarcane. Journal of Plant Physiology. 2005;162(1):11–20.

[43] Wang J, Nayak S, Koch K, Ming R. Carbon partitioning in sugarcane (Saccharum species). Frontiers in Plant Science. 2013;4(June):2005–2010. Available from: http://journal.frontiersin.org/article/10.3389/fpls.2013.00201/abstract.

[44] Kasirajan L, Hoang NV, Furtado A, Botha FC, Henry RJ. Transcriptome analysis highlights key differentially expressed genes involved in cellulose and lignin biosynthesis of sugarcane genotypes varying in fiber content. Scientific Reports. 2018;8(1):1–16.

[45] Lingle SE, Dyer JM. Cloning and expression of sucrose synthase-1 cDNA from sugarcane. Journal of Plant Physiology. 2001;158(1):129–131.

[46] Hoffmann-Thoma G, Hinkel K, Nicolay P, Willenbrink J. Sucrose accumulation in sweet sorghum stem internodes in relation to growth. Physiologia Plantarum. 1996;97(2):277–284.

[47] Zhang J, Zhang X, Tang H, Zhang Q, Hua X, Ma X, et al. Allele-defined genome of the autopolyploid sugarcane Saccharum spontaneum L. Nature Genetics. 2018;50(11):1565–1573.

[48] Garsmeur O, Droc G, Antonise R, Grimwood J, Potier B, Aitken K, et al. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. Nature Communications. 2018;9(1). Available from: http://dx.doi.org/10.1038/s41467-018-05051-5.

[49] Grivet L, Daniels C, Glaszmann JCC, Hont aD, D'Hont A. A review of recent molecular genetics evidence for sugarcane evolution and domestication. Ethnobotany Research & Applications. 2004;2(0):9–17.

[50] Aitken K, Li J, Piperidis G, Qing C, Yuanhong F, Jackson P. Worldwide Genetic Diversity of the Wild Species and Level of Diversity Captured within Sugarcane Breeding Programs. Crop Science. 2018;58(1):218. Available from: https://dl.sciencesocieties.org/publications/cs/abstracts/58/1/218.

[51] Matsuoka S, Ferro J, Arruda P. The Brazilian experience of sugarcane ethanol industry. In Vitro Cellular & Developmental Biology - Plant. 2009 jun;45(3):372–381. Available from: http://link.springer.com/10.1007/978-1-4419-7145-6{_}9http://link.springer.com/10.1007/s11627-009-9220-z.

[52] Creste S, Xavier MA, Landell MGA. Importância do germoplasma no desenvolvimento de cultivares de cana-de-açúcar com perfil agroenergético. In: Cortez LAB, editor. Bioetanol de Cana-de-Açúcar: P&D para produtividade e sustentabilidade; 2010. p. 313–317.

[53] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010 jan;26(1):139–140. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btp616.

[54] Irvine JE. Saccharum species as horticultural classes. Theoretical and Applied Genetics. 1999 feb;98(2):186–194. Available from: http://dx.doi.org/10.1007/s001220051057http://link.springer.com/10.1007/s001220051057.

[55] Casu RE, Jarmey JM, Bonnett GD, Manners JM. 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. 2007 feb;7(2):153–167. Available from: http://link.springer.com/10.1007/s10142-006-0038-z.

[56] Carson DL, Huckett BI, Botha FC. Differential gene expression in sugarcane leaf and internodal tissues of varying maturity. South African Journal of Botany. 2002 dec;68(4):434–442. Available from: c:{\%}5CDocumentsandSettings{\%}5Ccas128{\%}5CMyDocuments{\%}5CDownloadedpapers{\%}5CSAJB-2002-68-434-442.pdf{\%}5Cnhttp://linkinghub.elsevier.com/retrieve/pii/S0254629915303707https://linkinghub.elsevier.com/retrieve/pii/S0254629915303707.

[57] Grivet L, Arruda P. Sugarcane genomics: Depicting the complex genome of an important tropical crop. Current Opinion in Plant Biology. 2002;5(2):122–127.

[58] Vicentini R, Bottcher A, Dos Santos Brito M, Dos Santos AB, Creste S, De Andrade Landell MG, et al. Large-scale transcriptome analysis of two sugarcane genotypes contrasting for lignin content. PLoS ONE. 2015 aug;10(8):e0134909. Available from: http://dx.plos.org/10.1371/journal.pone.0134909.

[59] Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biology. 2010;11(2):R14. Available from: http://genomebiology.biomedcentral.com/articles/10.1186/gb-2010-11-2-r14.

[60] Zhang J, Nagai C, Yu Q, Pan YB, Ayala-Silva T, Schnell RJ, et al. Genome size variation in three Saccharum species. Euphytica. 2012;185(3):511–519.

[61] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. 1995;57(1):289–300. Available from: http://www.jstor.org/stable/2346101{\%}5Cnhttp://about.jstor.org/terms.

[62] de Setta N, Monteiro-Vitorello C, Metcalfe C, Cruz GM, Del Bem L, Vicentini R, et al. Building the sugarcane genome for biotechnology and identifying evolutionary trends. BMC Genomics. 2014;15(1):540. Available from: http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-15-540.

[63] Cardoso-Silva CB, Costa EA, Mancini MC, Balsalobre TWA, Costa Canesin LE, Pinto LR, et al. De novo assembly and transcriptome analysis of contrasting sugarcane varieties. PLoS ONE. 2014;9(2).

3 A HIERARCHICAL BAYESIAN MODEL TO ASSESS ALLELE-SPECIFIC EXPRESSION IN MIXED-PLOIDY SPECIES REVEALS EXPRESSION BIASES IN SUGARCANE

Abstract

Allele-specific expression (ASE) represents differences in the magnitude of expression between alleles of the same gene. Allelic imbalance in diploids occurs if the ratio of expression between both alleles shows deviations from the expected equivalent expression. However, this is not straightforward for polyploids, especially autopolyploids, as knowledge about the dosage of each allele is required for accurate estimation of ASE. This is the case for the genomically complex *Saccharum* species, characterized by high levels of ploidy and aneuploidy. We propose a model to test for allelic imbalance in *Saccharum* that can be easily expanded to other polyploids. As a test case we used genotyping data and RNA-Sequencing libraries from leaves of six sugarcane accessions. We used a hierarchical Beta-Binomial model to test if allele expression followed the expectation based on genomic allele dosage. The doses of the alleles were used in a prior Beta distribution for modeling the proportion of the reference allele from RNA counts. This proportion was then used in a Binomial distribution to model the number of RNA-seq reads showing this allele. We used the Bayesian Markov chain Monte Carlo procedure to draw samples from the *a posteriori* distribution. We called a polymorphism as showing ASE when the relative genomic dose was outside the highest density interval of the posterior distribution in a certain genotype. Part of the genes evaluated in each accession showed ASE and were related to a broad range of processes, mostly associated to the general metabolism, organelles, responses to stress and responses to stimuli. In addition, the frequency of genes with ASE in high-level functional terms was similar among the genotypes. Because the highest frequencies of ASE occurred in sugarcane hybrids, we fancy some influence of the interspecific hybridization in these genotypes. Although the number of polymorphisms we evaluated is still somewhat limited, our study is the first to assess genome-wide ASE in a high- and mixed-ploidy system using estimated doses of the alleles.

Keywords: Allelic imbalance; Polyploid; Allele dosage; Bayes; *Saccharum*

References

[1] Benjamini, Y. e Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57(1):289–300.

[2] Cai, M., Lin, J., Li, Z., Lin, Z., Ma, Y., Wang, Y., e Ming, R. (2020). Allele specific expression of Dof genes responding to hormones and abiotic stresses in sugarcane. *PLOS ONE*, 15(1):e0227716.

[3] Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., e Riddell, A. (2017). Stan : A Probabilistic Programming Language. *Journal of Statistical Software*, 76(1).

[4] Castel, S. E., Levy-Moonshine, A., Mohammadi, P., Banks, E., e Lappalainen, T. (2015). Tools and best practices for data processing in allelic expression analysis. *Genome Biology*, 16(1):1–12.

[5] Correr, F. H., Hosaka, G. K., Barreto, F. Z., Valadão, I. B., Balsalobre, T. W. A., Furtado, A., Henry, R. J., Carneiro, M. S., e Margarido, G. R. A. (2020). Differential expression in leaves of Saccharum genotypes contrasting in biomass production provides evidence of genes involved in carbon partitioning. *BMC Genomics*, 21(1):673.

[6] da Silva, J. A. (2017). The Importance of the Wild Cane Saccharum spontaneum for Bioenergy Genetic Breeding. *Sugar Tech*, 19(3):229–240.

[7] de C. Lara, L. A., Santos, M. F., Jank, L., Chiari, L., Vilela, M. d. M., Amadeu, R. R., dos Santos, J. P. R., Pereira, G. d. S., Zeng, Z.-B., e Garcia, A. A. F. (2019). Genomic Selection with Allele Dosage in Panicum maximum Jacq. *G3: Genes|Genomes|Genetics*, 9(8):2463–2475.

[8] De Mendonça Vilela, M., Del Bem, L. E., Van Sluys, M. A., De Setta, N., Kitajima, J. P., Cruz, G. M. Q., Sforça, D. A., De Souza, A. P., Ferreira, P. C. G., Grativol, C., Cardoso-Silva, C. B., Vicentini, R., e Vincentz, M. (2017). Analysis of Three Sugarcane Homo/Homeologous Regions Suggests Independent Polyploidization Events of Saccharum officinarum and Saccharum spontaneum. *Genome Biology and Evolution*, 9(2):266–278.

[9] DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., e Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5):491–498.

[10] Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., e Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, 6(5):e19379.

[11] Ereful, N. C., Liu, L. Y., Tsai, E., Kao, S. M., Dixit, S., Mauleon, R., Malabanan, K., Thomson, M., Laurena, A., Lee, D., Mackay, I., Greenland, A., Powell, W., e Leung, H. (2016). Analysis of Allelic Imbalance in Rice Hybrids Under Water Stress and Association of Asymmetrically Expressed Genes with Drought-Response QTLs. *Rice*, 9(1).

[12] Garcia, A. A. F., Mollinari, M., Marconi, T. G., Serang, O. R., Silva, R. R., Vieira, M. L. C., Vicentini, R., Costa, E. A., Mancini, M. C., Garcia, M. O. S., Pastina, M. M., Gazaffi, R., Martins, E. R. F., Dahmer, N., Sforça, D. A., Silva, C. B. C., Bundock, P., Henry, R. J., Souza, G. M., van Sluys, M.-A., Landell, M. G. A., Carneiro, M. S., Vincentz, M. A. G., Pinto, L. R., Vencovsky, R., e Souza, A. P. (2013). SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. *Scientific Reports*, 3(1):3399.

[13] Gemenet, D. C., da Silva Pereira, G., De Boeck, B., Wood, J. C., Mollinari, M., Olukolu, B. A., Diaz, F., Mosquera, V., Ssali, R. T., David, M., Kitavi, M. N., Burgos, G., Felde, T. Z., Ghislain, M., Carey, E., Swanckaert, J., Coin, L. J. M., Fei, Z., Hamilton, J. P., Yada, B., Yencho, G. C., Zeng, Z.-B., Mwanga, R. O. M., Khan, A., Gruneberg, W. J., e Buell, C. R. (2020). Quantitative trait loci and differential gene expression analyses reveal the genetic basis for negatively associated β-carotene and starch content in hexaploid sweetpotato [Ipomoea batatas (L.) Lam.]. *Theoretical and Applied Genetics*, 133(1):23–36.

[14] Grover, C. E., Gallagher, J. P., Szadkowski, E. P., Yoo, M. J., Flagel, L. E., e Wendel, J. F. (2012). Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytologist*, 196(4):966–971.

[15] Hu, X., Wang, H., Diao, X., Liu, Z., Li, K., Wu, Y., Liang, Q., Wang, H., e Huang, C. (2016). Transcriptome profiling and comparison of maize ear heterosis during the spikelet and floret differentiation stages. *BMC Genomics*, 17(1):1–18.

[16] Kim, D., Langmead, B., e Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, 12(4):357–360.

[17] Langmead, B. e Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–359.

[18] Lavania, U. C. (2013). Polyploidy, body size, and opportunities for genetic enhancement and fixation of heterozygosity in plants. *The Nucleus*, 56(1):1–6.

[19] McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., e DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9):1297–1303.

[20] Medeiros, C., Balsalobre, T. W. A., e Carneiro, M. S. (2020). Molecular diversity and genetic structure of Saccharum complex accessions. *PLOS ONE*, 15(5):e0233211.

[21] N'Diaye, A., Haile, J. K., Cory, A. T., Clarke, F. R., Clarke, J. M., Knox, R. E., e Pozniak, C. J. (2017). Single marker and haplotype-based association analysis of semolina and pasta colour in elite durum wheat breeding lines using a high-density consensus map. *PLoS ONE*, 12(1):1–24.

[22] Nguyen, T. K. e Lim, J.-H. (2019). Tools for Chrysanthemum genetic research and breeding: Is genotyping-by-sequencing (GBS) the best approach? *Horticulture, Environment, and Biotechnology*, 60(5):625–635.

[23] Osborn, T. C., Chris Pires, J., Birchler, J. A., Auger, D. L., Chen, Z. J., Lee, H. S., Comai, L., Madlung, A., Doerge, R. W., Colot, V., e Martienssen, R. A. (2003). Understanding mechanisms of novel gene expression in polyploids. *Trends in Genetics*, 19(3):141–147.

[24] Pereira, G. S., Garcia, A. A. F., e Margarido, G. R. (2018). A fully automated pipeline for quantitative genotype calling from next generation sequencing data in autopolyploids. *BMC Bioinformatics*, 19(1):1–10.

[25] Pham, G. M., Newton, L., Wiegert-Rininger, K., Vaillancourt, B., Douches, D. S., e Buell, C. R. (2017). Extensive genome heterogeneity leads to preferential allele expression and copy number-dependent expression in cultivated potato. *Plant Journal*, 92(4):624–637.

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

[27] Piperidis, N. e D'Hont, A. (2020). Sugarcane genome architecture decrypted with chromosome-specific oligo probes. *The Plant Journal*, page tpj.14881.

[28] Powell, J. J., Fitzgerald, T. L., Stiller, J., Berkman, P. J., Gardiner, D. M., Manners, J. M., Henry, R. J., e Kazan, K. (2017). The defence-associated transcriptome of hexaploid wheat displays homoeolog expression and induction bias. *Plant Biotechnology Journal*, 15(4):533–543.

[29] Romanel, A., Lago, S., Prandi, D., Sboner, A., e Demichelis, F. (2015). ASEQ: Fast allele-specific studies from next-generation sequencing data. *BMC Medical Genomics*, 8(1):1–12.

[30] Sehgal, D. e Dreisigacker, S. (2019). Haplotypes-based genetic analysis: Benefits and challenges. *Vavilovskii Zhurnal Genetiki i Selektsii*, 23(7):803–808.

[31] Serang, O., Mollinari, M., e Garcia, A. A. F. (2012). Efficient exact maximum a posteriori computation for Bayesian SNP genotyping in polyploids. *PLoS ONE*, 7(2):1–13.

[32] Sforça, D. A., Vautrin, S., Cardoso-Silva, C. B., Mancini, M. C., Romero-da Cruz, M. V., Pereira, G. d. S., Conte, M., Bellec, A., Dahmer, N., Fourment, J., Rodde, N., Van Sluys, M.-A., Vicentini, R., Garcia, A. A. F., Forni-Martins, E. R., Carneiro, M. S., Hoffmann, H. P., Pinto, L. R., Landell, M. G. d. A., Vincentz, M., Berges, H., e de Souza, A. P. (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(May).

[33] Soneson, C. e Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*, 14(1):91.

[34] Spoelhof, J. P., Soltis, P. S., e Soltis, D. E. (2017). Pure polyploidy: Closing the gaps in autopolyploid research. *Journal of Systematics and Evolution*, 55(4):340–352.

[35] Stan Development Team (2018). RStan: the R interface to Stan.

[36] Tuch, B. B., Laborde, R. R., Xu, X., Gu, J., Chung, C. B., Monighetti, C. K., Stanley, S. J., Olsen, K. D., Kasperbauer, J. L., Moore, E. J., Broomer, A. J., Tan, R., Brzoska, P. M., Muller, M. W., Siddiqui, A. S., Asmann, Y. W., Sun, Y., Kuersten, S., Barker, M. A., De La Vega, F. M., e Smith, D. I. (2010). Tumor transcriptome sequencing reveals allelic expression imbalances associated with copy number alterations. *PLoS ONE*, 5(2).

[37] Vieira, M. L. C., Almeida, C. B., Oliveira, C. A., Tacuatiá, L. O., Munhoz, C. F., Cauz-Santos, L. A., Pinto, L. R., Monteiro-Vitorello, C. B., Xavier, M. A., e Forni-Martins, E. R. (2018). Revisiting meiosis in sugarcane: Chromosomal irregularities and the prevalence of bivalent configurations. *Frontiers in Genetics*, 9(JUN):1–12.

[38] Wang, Z., Gerstein, M., e Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1):57–63.

[39] Wood, D. L., Nones, K., Steptoe, A., Christ, A., Harliwong, I., Newell, F., Bruxner, T. J., Miller, D., Cloonan, N., e Grimmond, S. M. (2015). Recommendations for accurate resolution of Gene and isoform allele-specific expression in RNA-seq data. *PLoS ONE*, 10(5):1–27.

[40] Yoo, M. J., Szadkowski, E., e Wendel, J. F. (2013). Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity*, 110(2):171–180.

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

[42] Zhang, J., Nagai, C., Yu, Q., Pan, Y. B., Ayala-Silva, T., Schnell, R. J., Comstock, J. C., Arumuganathan, A. K., e Ming, R. (2012). Genome size variation in three Saccharum species. *Euphytica*, 185(3):511–519.

[43] Zhang, J., Sharma, A., Yu, Q., Wang, J., Li, L., Zhu, L., Zhang, X., Chen, Y., e Ming, R. (2016a). Comparative structural analysis of Bru1 region homeologs in Saccharum spontaneum and S. officinarum. *BMC Genomics*, 17(1).

[44] Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., Zhu, F., Jones, T., Zhu, X., Bowers, J., Wai, C. M., Zheng, C., Shi, Y., Chen, S., Xu, X., Yue, J., Nelson, D. R., Huang, L., Li, Z., Xu, H., Zhou, D., Wang, Y., Hu, W., Lin, J., Deng, Y., Pandey, N., Mancini, M., Zerpa, D., Nguyen, J. K., Wang, L., Yu, L., Xin, Y., Ge, L., Arro, J., Han, J. O., Chakrabarty, S., Pushko, M., Zhang, W., Ma, Y., Ma, P., Lv, M., Chen, F., Zheng, G., Xu, J., Yang, Z., Deng, F., Chen, X., Liao, Z., Zhang, X., Lin, Z., Lin, H., Yan, H., Kuang, Z., Zhong, W., Liang, P., Wang, G., Yuan, Y., Shi, J., Hou, J., Lin, J., Jin, J., Cao, P., Shen, Q., Jiang, Q., Zhou, P., Ma, Y., Zhang, X., Xu, R., Liu, J., Zhou, Y., Jia, H., Ma, Q., Qi, R., Zhang, Z., Fang, J., Fang, H., Song, J., Wang, M., Dong, G., Wang, G., Chen, Z., Ma, T., Liu, H., Dhungana, S. R., Huss, S. E., Yang, X., Sharma, A., Trujillo, J. H., Martinez, M. C., Hudson, M., Riascos, J. J., Schuler, M., Chen, L. Q., Braun, D. M., Li, L., Yu, Q., Wang, J., Wang, K., Schatz, M. C., Heckerman, D., Van Sluys, M. A., Souza, G. M., Moore, P. H., Sankoff, D., VanBuren, R., Paterson, A. H., Nagai, C., e Ming, R. (2018). Allele-defined genome of the autopolyploid sugarcane Saccharum spontaneum L. *Nature Genetics*, 50(11):1565–1573.

4 CONCLUSIONS

In this thesis, we aimed to investigate differences among *Saccharum* genotypes phenotypically contrasting in their biomass content. In the first chapter we assessed gene expression profiles of twelve sugarcane genotypes grouped into high and low biomass groups. The gene expression data correctly represented the difference between the groups and revealed substantial variability among the high biomass accessions. The groups showed significant differences in the expression of genes involved in carbon partitioning, mostly sucrose synthesis and degradation. Within the groups we could identify the enrichment of defense and carbohydrate-related terms. In addition, we explored the expression and co-expression profiles of groups of genes that were members of pathways of interest. Finally, we also showed how expression profiles at the transcript level can bring new insights when assessing differences between the biomass groups.

We devoted the second chapter to investigate if genes showing allele imbalance could be related to distinct functional processes. As we aimed to investigate whether alleles were expressed accordingly to their estimated dosages, we proposed a model to account for prior knowledge of this information. We used a hierarchical Bayesian approach to go from a prior distribution of the allele proportion, based on genotyping information, to a posterior considering the relative expression of the allele. Our results reveal that allele-specific expression affects part of the investigated loci in *Saccharum* genotypes. However, we could not find clear functional patterns among genes showing allele-specific expression. Despite the innate limitations of the genotyping-by-sequencing approach, we successfully developed and applied a model to drive insights about allele-specific expression in the complex polyploid sugarcane.