

University of São Paulo  
Luiz de Queiroz College of Agriculture

Metabolomics and transcriptomics analyses of sugarcane variety SP80-3280  
throughout development in the field

**Maryke Wijma**

Thesis presented to obtain the degree of Doctor in Science  
Area: Bioenergy

Piracicaba  
2020

Maryke Wijma  
PhD in Bioenergy

**Metabolomics and transcriptomics analyses of sugarcane variety SP80-3280 throughout  
development in the field**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Advisor:

Prof<sup>a</sup> Dr<sup>a</sup> **GLAUCIA MENDES SOUZA**

Thesis presented to obtain the degree of Doctor in Science  
Area: Bioenergy

Piracicaba  
2020

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

Wijma, Maryke

Metabolomics and transcriptomics analyses of sugarcane variety SP80-3280 throughout development in the field / Maryke Wijma - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2020.

110 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura "Luiz de Queiroz". Universidade Estadual de Campinas. Universidade Estadual Paulista "Julio de Mesquita Filho"

1. Cana-de-açúcar 2. SP80-3280 3. Metabolômica 4. Transcriptômica 5. Biologia de sistemas I. Título

## ACKNOWLEDGEMENTS

To CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for the financial support during the period of 03/2016 to 12/2016 – Obrigada.

To FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo), process n<sup>o</sup> 2016/21008-8, for the financing of this project during the period of 01/2017 to 05/2020 – Obrigada.

To the secretaries of the PhD in Bioenergy program, specifically RODRIGO PESSANHA and ALEXANDRE JOVINIANO DOS SANTOS, for administrative aid – Obrigada.

To the principal investigator of this PhD thesis and my academic advisor, PROF DR GLAUCIA MENDES SOUZA – Obrigada.

To my unofficial co-advisors and academic mentors during the development of this PhD thesis, PROF DR PIO COLEPICOLO NETO, DR LEONARDO ZAMBOTTI VILLELA and DR AUGUSTO LIMA DINIZ – Obrigada.

To my previous academic advisors and inspirers from my bachelor's degree, DR SHAUN PETERS and DR PAUL HILLS from the University of Stellenbosch, South Africa – Dankie.

To my colleagues from the Sugarcane Signal Transduction Laboratory, and the Biochemistry and Molecular Biology of Algae Laboratory from the Institute of Chemistry from the University of São Paulo, Brazil – Obrigada.

To my organism of study, sugarcane (*Saccharum* sp. – SP80-3280, *S. officinarum*, *S. spontaneum*), for challenging me every step of the way – Obrigada.

To my Brazilian family and friends – Obrigada.

To my South African family and friends – Dankie.

To my international family and friends – Thank you. Gracias. Merci.

To you Brazil, for receiving me with open arms, allowing me to develop my academic, social, and personal skills – Obrigada.

## TABLE OF CONTENTS

RESUMO .....	6
ABSTRACT .....	67
LIST OF FIGURES .....	8
LIST OF TABLES .....	10
1 INTRODUCTION .....	11
1.1. Motivation and justification of this study .....	11
1.2. Objectives.....	12
1.2.1. Main objective.....	12
1.2.2. Specific objectives .....	12
1.3. Supporting literature review.....	13
1.3.1. Sugarcane development, phenology and cultivation.....	13
1.3.2. Untargeted and targeted metabolomics studies.....	14
1.3.3. Sugarcane metabolomics studies .....	16
1.4. Previous results obtained from studies directly associated with the PhD project presented here carried out by the Sugarcane Signal Transduction Laboratory .....	17
2. MATERIALS AND METHODS.....	21
2.1. Plant material .....	21
2.2. Experimental design, field conditions and sample collection.....	21
2.3. Sample preparation .....	23
2.4. Untargeted metabolomics analyses of four anatomically different sugarcane tissues..	23
2.4.1. Samples used and global metabolite extractions.....	23
2.4.2. HPLC-MS analysis and data acquisition .....	24
2.4.3. Untargeted metabolomics data generated with HPLC-MS - data processing.....	25
2.4.4. Normality testing, data dimensionality reduction and model validation .....	25
2.4.5. Metabolomic tissue profiles – discriminant analyses .....	26
2.5. Transcriptomics analyses of four anatomically different tissues from sugarcane .....	26
2.5.1. DNA microarray experiments and data acquisition.....	26
2.5.2. The usage of the processed microarray data and multi-dimensional scaling (MDS) analyses .....	27
2.6. The integration between two data modalities: metabolomics and transcriptomics data	27
2.6.1. Pathway Activity Profiling (PAPi) .....	27
2.6.2. Selection of the input data for the Multi-Omics Factor Analysis (MOFA) .....	28
2.6.3. The implementation of the Multi-Omics Factor Analysis (MOFA).....	28
2.7. The selection of the targeted metabolic pathways and possible genes of interest .....	29
3. RESULTS .....	31
3.1. Untargeted metabolomics analyses of four anatomically different sugarcane tissues..	31
3.1.1. HPLC-MS data processing, annotation and the generation of “clean” metabolomics data for downstream analyses .....	31
3.1.2. Data dimensionality reduction and discriminant analyses – PCA and PLS-DA models .....	34
3.1.2.1. Between tissue comparisons – all tissues, fields and collection points joined for discriminant analyses .....	34

3.1.2.2. Between field comparison – tissues separated; fields and collection points joined .....	36
3.2. Transcriptomics analyses of four anatomically different tissues from SP80-3280 .....	39
3.3. MOFA – the tool leading the unsupervised integration between two data modalities ..	43
3.4. Narrowing down the metabolic pathways and the selection of target genes within these pathways .....	48
4. DISCUSSION.....	57
4.1. Sugarcane phenology and morphological changes in response to precipitation and temperature alterations throughout development under field conditions .....	57
4.2. A metabolomics pipeline for the successful exploration of the sugarcane metabolome: from the analytical tool to the metabolite name .....	59
4.3. The exploration of the sugarcane metabolome highlights the metabolic differences between four anatomical tissues and identify economically valuable compounds .....	61
4.4. Metabolite accumulation in different sugarcane tissues allow to distinguish between the different developmental profiles of “one-year” and “one-and-a-half-year” sugarcane crops .....	64
4.5. The exploration of the sugarcane transcriptome proves overlaps between the metabolite profiles drawn for each anatomical tissue .....	67
4.6. The direct combination of two “omics” in an unsupervised manner pinpoints the main metabolic pathways being altered throughout sugarcane development .....	68
4.7. Omics integration and pathway exploration led to the final identification and selection of target genes within the candidate pathways .....	71
5. MAIN CONCLUSIONS .....	75
REFERENCES .....	77110
APPENDICES – SUPPLEMENTARY FIGURES .....	77103
APPENDICES – SUPPLEMENTARY TABLES.....	108108

## RESUMO

**Análises metabolômicas e transcriptômicas da variedade de cana-de-açúcar SP80-3280 ao longo do desenvolvimento no campo**

A cana-de-açúcar (*Saccharum* sp.) é considerada uma das culturas mais eficientes na conversão de energia solar em energia química, especificamente na forma de açúcar e biomassa, e tem sido explorada principalmente como fonte primária para a produção de açúcar e bioetanol. Melhorias no rendimento da cana são necessárias para atender às demandas crescentes de produção de açúcar e bioetanol. Para tanto, se faz necessário compreender como a planta percebe e responde às mudanças em seu ambiente externo. Em trabalhos anteriores, nosso grupo observou que a precipitação e a temperatura influenciam no rendimento e o perfil de maturação da variedade comercial SP80-3280. Experimentos de campo mostraram que a cana pode suportar o déficit hídrico da fase inicial de desenvolvimento, mantendo a fotossíntese e acúmulo de carbono na forma de fibra, especificamente produzindo colmos mais grossos, em detrimento do acúmulo de açúcar. A produção de biomassa e açúcar são caracteres complexos e necessitam de abordagens integradas para ser estudada. Optamos por desenvolver uma abordagem de biologia de sistemas integrando dados de transcriptoma e metaboloma para verificar se e quais vias metabólicas podem ser alteradas durante o ciclo de crescimento e maturação desta cultura. Desenvolvemos um pipeline analítico e estatístico para a realização de estudos de metabolômica não direcionada e sua integração, de uma forma não supervisionada, com dados transcriptômicos, destacando assim a importância da integração ao trabalhar com conjuntos de dados altamente complexos e heterogêneos. Aqui, discutimos quatro vias principais associadas ao desenvolvimento da variedade SP80-3280: via da pentose fosfato; biossíntese de fenilalanina, tirosina e triptofano; biossíntese de flavonóides; e biossíntese de fenilpropanóides, principalmente ativos em folha (fonte) e entrenó maduro (dreno). Genes candidatos (CAD, PER, PAL, F3H) dentro dessas vias são identificados como possíveis alvos para estudos futuros para melhor compreender os diferentes perfis de desenvolvimento e sua demanda de carbono. Este estudo também lança luz sobre as diferenças metabolômica e transcriptômica entre quatro tecidos anatômicos diferentes. Além disso, foram identificados compostos de valor econômico e agrônômico, tais como os fenólicos e os glucosinolatos, ampliando o conhecimento sobre o metaboloma da cana e sugerindo novos usos para as partes da planta atualmente pouco exploradas.

Palavras-chave: Cana-de-açúcar, SP80-3280, Metabolômica, Transcriptômica, Biologia de sistemas

## ABSTRACT

**Metabolomics and transcriptomics analyses of sugarcane variety SP80-3280 throughout development in the field**

Sugarcane (*Saccharum* sp.) is considered one of the world's most efficient crops in the conversion of solar to chemical energy in the form of sugar and biomass, and it has mostly been exploited as a primary source for the production of sugar and bioethanol. Yield improvements are required to meet the increasing demands for sugar and bioethanol production; however, it is necessary to understand how sugarcane perceives and responds to changes in its external environment to meet these demands. Previous work from the group observed that precipitation and temperature influence SP80-3280's yield and maturation profile. Field experiments showed that sugarcane can support early phase water deficit by continuing photosynthate production and carbon assimilation in the form of fiber production, specifically producing thicker culms, instead of accumulating sugar. Yield is a complex trait that needs integrated approaches for being studied. We developed and took a systems biology approach integrating data from the sugarcane metabolome and transcriptome to verify if and which pathways are altered during the growth and maturation cycle of this crop. We developed an analytical and statistical pipeline for conducting untargeted metabolomics studies and integrating it in an unsupervised manner with transcriptomics data, thus highlighting the importance of integration when working with highly complex and heterogeneous datasets. We discuss four main pathways associated with SP80-3280's development: pentose phosphate pathway; phenylalanine, tyrosine and tryptophan biosynthesis; flavonoid biosynthesis; and phenylpropanoid biosynthesis, mostly active in the main source (leaf) and sink (mature internode) tissues. Candidate genes (*CAD*, *PER*, *PAL*, *F3H*) within these pathways are identified as possible targets for further studies to better understand the different developmental profiles and their carbon demand. This study also sheds light on the metabolomics and transcriptomics differences between four different anatomical tissues. Additionally, economically and agronomically valuable compounds, such as phenolics and glucosinolates were identified which contributed to the knowledge gaps regarding the sugarcane metabolome and suggested new uses for currently poorly utilized parts of the sugarcane plant.

Keywords: Sugarcane, SP80-3280, Metabolomics, Transcriptomics, Systems biology

## LIST OF FIGURES

- Figure 1. Morphological and degrees Brix results (Inada, 2016). (A) Plant height in meters; (B) the number of internodes; (C) diameter in the middle in millimetres; (D) diameter at the base in millimetres; (E) soluble solid content in degrees Brix. (F) Climatic conditions for the field experiments. Fields 1 (F1) and 2 (F2) refer to the plants that were planted in April 2012 and October 2012, respectively. C1, C2, C3, and C4 refer to collection points 1, 2, 3 and 4 specifically 4, 8, 11 and 13 months after planting. Different letters on the graphs indicate statistically significant differences between the different collection points based on ANOVA and Fisher's LSD ( $p < 0.05$ ). The asterisks indicate statistically significant differences between the different fields based on Student's t-tests ( $p < 0.05$ ); (F) Climatic conditions which sugarcane variety SP80-3280 experienced throughout development in the two field experiments. The average precipitation and temperature measurements during the two field experiments are plotted as continuous lines, respectively. .... 18
- Figure 2. Schematic representation of the field experiments. (A) The two field experiments and the eight plots within each field. Fields 1 (F1) and 2 (F2) refer to the plants that were planted in April 2012 and October 2012, respectively; (B) Climatic conditions which sugarcane variety SP80-3280 experienced throughout development in the two field experiments. The average precipitation and temperature measurements during the two field experiments are plotted as continuous lines, respectively; (C) The plots from which the tissues were collected. The arrows indicate the rows from which the tissues were collected; (D) One plot indicating that the plants were planted in four rows of 10 m with 1.35 m between the furrows; (E) Tissue samples that were collected namely leaf +1 (L1), upper (I1), young (I5) and mature (I9) internode..... 22
- Figure 3. Global metabolite extractions. (A) Representation of the metabolite extraction which separates the extract into three phases, the top layer represents the nonpolar phase, the bottom layer represents the polar phase and the pellet represents the biomass and protein excess; (B) Phase separations after the addition of both extraction buffers MTBE:MeOH (3:1) and H<sub>2</sub>O:MeOH (3:1); (C) The polar phase after the phase separation step. .... 24
- Figure 4. Workflow of Pathway Activity Profiling (PAPi) and its inputs adapted from Aggio et al., 2010. (A) Raw data from HPLC-MS is processed to render (B) the most probable metabolite, their annotations and in turn their (C) KEGG compound IDs and relative abundances which are (D) the inputs for PAPi. Step 1: All pathways for which each metabolite is known to play a part are collected from the KEGG database. Step 2: Each identified pathway then receives a score based on the abundance/relative abundance of the metabolite to which it is linked. Step 3: The total number of metabolites associated with each pathway is recorded and the pathways are then ranked according to the number of metabolites with which they are associated. The percentage of detected metabolic intermediates is then calculated for each listed pathway. Step 4: Finally, we sum over the scores for each pathway to obtain the total pathway score, SA. Step 5: The SA is normalized by dividing by the proportion of metabolites detected from its respective pathway. Step 6: The final data frame is obtained which includes all the pathways and their SA within each condition. .... 28
- Figure 5. Venn diagram of metabolite numbers from the metabolite profiles generated via HPLC-MS of each tissue collected throughout sugarcane SP80-3280 development in the field. .... 32
- Figure 6. QQ-plots of the log<sub>10</sub> transformed metabolomics data in each tissue. (A) Leaf +1 (L1); (B) upper (I1) internodes; (C) young (I5) internodes and (D) mature (I9) internodes. .... 33
- Figure 7. Discriminant models for between tissue comparisons. (A) PCA and (B) PLS-DA models generated from dry weight (DW) normalized and log<sub>10</sub> transformed metabolomics data, combining all tissues (L1, I1, I5 and I9), collection points (C1, C2, C3 and C4) and experimental fields (F1 and F2). (C) Heatmap representation of the 79 main metabolites responsible for the separations (VIP scores  $\geq 1.0$  from PLS-DA analysis). .... 35
- Figure 8. Discriminant models for between field comparisons generated from the normalized and log<sub>10</sub> transformed metabolomics data from the leaf +1 (L1) tissues, combining collection points (C1, C2, C3 and C4) and experimental fields (F1 and F2). (A) PCA and (B) PLS-DA models for L1; (C) Heatmap representation of the 32 main metabolites responsible for the separations (VIP scores  $\geq 1.0$  from PLS-DA analysis) and NA refers to non-detected metabolites. .... 36
- Figure 9. Discriminant models for between field comparisons generated from the normalized and log<sub>10</sub> transformed metabolomics data from the upper internode (I1) tissues, combining collection points (C1, C2, C3 and C4) and experimental fields (F1 and F2). (A) PCA and (B) PLS-DA models for I1; (C) Heatmap representation of the 9 main metabolites responsible for the separations (VIP scores  $\geq 1.0$  from PLS-DA analysis) and NA refers to non-detected metabolites. .... 37

- Figure 10. Discriminant models for between field comparisons generated from the normalized and log<sub>10</sub> transformed metabolomics data from the young internode (I5) tissues, combining collection points (C1, C2, C3 and C4) and experimental fields (F1 and F2). (A) PCA and (B) PLS-DA models for I5; (C) Heatmap representation of the 9 main metabolites responsible for the separations (VIP scores  $\geq 1.0$  from PLS-DA analysis) and NA refers to non-detected metabolites. .... 38
- Figure 11. Discriminant models for between field comparisons generated from the normalized and log<sub>10</sub> transformed metabolomics data from the mature internode (I9) tissues, combining collection points (C1, C2, C3 and C4) and experimental fields (F1 and F2). (A) PCA and (B) PLS-DA models for I9; (C) Heatmap representation of the 22 main metabolites responsible for the separations (VIP scores  $\geq 1.0$  from PLS-DA analysis) and NA refers to non-detected metabolites. .... 39
- Figure 12. Multidimensional scaling (MDS) analyses of the transcriptomics data. (A) Dim1 vs Dim2; (B) Dim1 vs Dim3. L1, I1, I5 and I9 refer to leaf +1 and upper, young, and mature internodes. Field 1 (F1) and 2 (F2) refer to the plants that were planted in April 2012 and October 2012, respectively. .... 40
- Figure 13. Heatmap constructed using top 500 most variable genes responsible for the variations in the different tissues (L1, I1, I5 and I9) from a transcriptomics point of view. Field 1 (F1) and 2 (F2) refer to the plants that were planted in April 2012 and October 2012, respectively. C1, C2, C3 and C4 refer to collection points 1, 2, 3 and 4 specifically 4, 8, 11 and 13 months after planting. .... 41
- Figure 14. Doughnut graphs representing the main KEGG metabolic pathways identified to be altered throughout development in the (A) leaf +1 (L1); (B) upper (I1); (C) young (I5) and mature (I9) internodal tissues. .... 43
- Figure 15. Pathway activities based on the Pathway Activity Profiling (PAPi) tool for leaf +1 (L1) from (A) F1 and (B) F2. C1, C2, C3, and C4 refer to collection points 1, 2, 3 and 4 specifically 4, 8, 11 and 13 months after planting. Different letters under the bars indicate statistically significant differences between the different collection points based on ANOVA and Fisher's LSD ( $p < 0.05$ ). (C) Heatmap analysis for the statistically significant expression of the selected sugarcane assembled sequences (SAS) transcripts shown as "classes". When a transcript was not expressed, it is coloured as black in the heatmap analysis. (D) Climatic conditions for the field experiments. Fields 1 (F1) and 2 (F2) refer to the plants that were planted in April 2012 and October 2012, respectively. C1, C2, C3, and C4 refer to collection points 1, 2, 3 and 4 specifically 4, 8, 11 and 13 months after planting. .... 52
- Figure 16. Pathway activities based on the Pathway Activity Profiling (PAPi) tool for mature internode (I9) from (A) F1 (April 2012 plantation) and (B) F2 (October 2013 plantation). C1, C2, C3, and C4 refer to collection points 1, 2, 3 and 4 specifically 4, 8, 11 and 13 months after planting. Different letters indicate statistically significant differences between the different collection points based on ANOVA and Fisher's LSD ( $p < 0.05$ ). (C) Heatmap analysis for the statistically significant expression of the selected sugarcane assembled sequences (SAS) transcripts shown as "classes". When a transcript was not expressed, it is coloured as black in the heatmap analysis. (D) Climatic conditions for the field experiments. Fields 1 (F1) and 2 (F2) refer to the plants that were planted in April 2012 and October 2012, respectively. C1, C2, C3, and C4 refer to collection points 1, 2, 3 and 4 specifically 4, 8, 11 and 13 months after planting. .... 54

## LIST OF TABLES

Table 1.Gradient information for the high-performance liquid chromatography coupled mass spectrometry (HPLC-MS) runs using a pentafluorophenyl propyl ligand column (Phenomenex Luna PFP, 100 mm x 4.60 mm x 2.6 $\mu$ m).....	25
Table 2.Number of features detected from XCMS in both positive and negative ionization modes; the corresponding amount of XCMS features with at least 1 KEGG ID; the total possible amount of KEGG IDs detected and the unique KEGG IDs detected in each tissue namely leaf +1 (L1), upper internode (I1), young internode (I5) and mature internode (I9). .....	31
Table 3.Percentages of the differently transformed metabolomics data generated for the leaf +1 (L1), upper (I1), young (I5) and mature internodal (I9) tissue samples in which the Shapiro Wilk's test for normality were $p > 0.05$ . .....	32
Table 4.Percentages of the log10 transformed metabolomics data generated for the leaf +1 (L1), upper (I1), young (I5) and mature internodal (I9) tissue samples, in conjunction with auto scaled and pareto scaled metabolomics data, in which the Shapiro Wilk's test for normality were $p > 0.05$ . .....	33
Table 5.Principal Component Analysis (PCA) results generated for the log10 transformed metabolomics data of the leaf +1 (L1), upper (I1), young (I5) and mature internodal (I9) tissue samples when no scaling, auto scaling and pareto scaling techniques were applied on the metabolomics data. ....	34
Table 6.The Gene Ontology (GO) IDs and terms represented by the genes from each of the clusters identified by the heatmap analysis. ....	41
Table 7.An integrative view of the combined output results from MOFA for the leaf +1 (L1), upper (I1), young (I5) and mature internodal (I9) tissues. For the transcriptomics data modality, the GO IDs and GO terms are listed. For the metabolomics data modality, the KEGG map IDS and KEGG map names are listed.....	44
Table 8.The four main selected pathways from the KEGG database, the pathway map ID, the number of enzymes within the particular pathway and the number of SAS in the particular pathway.....	48
<b>TABLE 9.THE NINE NARROWED DOWN SAS TRANSCRIPTS FROM THE FOUR SELECTED PATHWAYS FROM THE KEGG DATABASE AND THEIR RESPECTIVE ENZYME CODES AND ENZYME NAMES. ....</b>	<b>50</b>

## 1 INTRODUCTION

### 1.1. Motivation and justification of this study

The continuous use and misuse of fossil fuels have led to the risk of its depletion, and severe and potentially irreversible environmental threats (Goldemberg, 2007; Gupta & Demirbas, 2012). Critical reductions in global greenhouse gas (GHG) emissions are required to achieve the objective of keeping the increase in the average global temperature to well below 2°C in the context of sustainable development and poverty eradication (Giannakidis et al., 2018; Lukas, 2015). One way of achieving this goal is to increase the use of renewable energy sources, such as biofuels like bioethanol produced from sugarcane, which can be used to fuel society as well as reduce GHG emissions (Cortez, 2014).

Sugarcane is currently regarded as one of the most efficient feedstocks for bioethanol production (Goldemberg et al., 2014). The commercial sugarcane varieties (*Saccharum* sp.), members of the Poaceae family, are interspecific hybrids between 2 ancestral species namely *S. officinarum*, also referred to as “noble cane” due to its ability to produce high amounts of sucrose in the culms, and *S. spontaneum*, which is known for its tolerance to biotic and abiotic stresses throughout the developmental cycle (Zhou, 2017). Sugarcane plants present high biomass (Coelho et al., 2012; Santchurn et al., 2012) and sugar productivity (Shrivastava et al., 2015; Tiawari et al., 2009), thus making them economically and agronomically valuable crops (Costa et al., 2014; McKay et al., 2016).

In recent years, the improvement of sugarcane has reached a plateau (Dal-Bianco et al., 2012; Dwivedi, 2003) and it is known that it has not reached its full yield potential just yet. In addition, due to global climate change, significant alterations in climatic patterns are being observed and its effects are expected to increase even more by the turn of the century (Hertel et al., 2010; Mekonnen, 2019; Mendelsohn & Dinar, 2009). These changes will affect sugarcane production and yield, and plants in general, thus, a great amount of research is now being directed towards understanding how changes in precipitation and temperature alter sugarcane developmental patterns and metabolism (Cardozo et al., 2015; Legendre, 1975; Wiedenfeld, 2000).

During plant development, plants need to respond to developmental and environmental signals in their internal and external environments by relying on the activity of a variety of cellular components, such as for example metabolites, that trigger the appropriate metabolic responses needed for survival (Krzyzaniak et al., 2018; Matilla, 2018). It has been well described that plants have the ability to produce large amounts and a wide variety of metabolites that are important role players in their development and in response to changes in their environment (Bhatla, 2018; Bhatla & Lal, 2018; Mitchell et al., 2020), be it internal or external. These small molecules (< 1.800 Da) form the basis of crop yield and quality (Gatehouse, 2004).

Metabolites are either involved in primary metabolism, which refers to processes indispensable for growth and development, or secondary metabolism, which is crucial for the adaptation to changes in the internal and external environments in order to remain in homeostasis (Hüsemann et al., 1989; Luckner, 2013).

The study of plant metabolomics not only provides information regarding the number of these small molecules referred to as metabolites but also the correlations between them and important agronomic traits (Herrmann & Schauer, 2013; Lima et al., 2018), thus it could possibly lead to the development of more rational models to link specific metabolites and/or pathways to yield or quality associated phenotypes. These phenotypes depend on the biosynthetic pathways and accumulation of certain metabolites in specific organs, at specific developmental stages and upon internal and external environmental signals (Roldan et al., 2014).

Plant development strongly relies on several signalling systems that provide the plant with information with regards to its internal and external environments (Fosket, 1994). Sugarcane development is largely governed by environmental factors such as precipitation and temperature (Moore et al., 2013). In addition, the yield is mainly controlled by the rate of biomass accumulation and the duration of growth (Blum, 2018; Wallace, 2011), the latter being dependent on the season in which the sugarcane is initially planted.

Since metabolites are the end products of gene expression and regulation, protein translation and enzymatic activities (Allwood et al., 2008; Rochfort, 2005), the comprehensive and unbiased study of these small, endogenous metabolites called metabolomics (Fiehn, 2002; Hall et al., 2002) can provide us with valuable information and a comprehensive understanding of a biological system on a molecular level. This is of immense importance seeing as it is favourable to understand a system prior to its modification, and current conventional breeding is too complex due to the polyploid nature of the sugarcane genome (Grivet & Arruda, 2002; Hoang et al., 2017), thus alternative methods for unravelling sugarcane are being researched to a great extent.

When integrating metabolomics data with gene expression information, obtained from transcriptomics studies (Kleessen et al., 2015), and conducting in-depth data mining it would be possible to find connections between specific traits, the genes that code for them and the metabolites and/or metabolic pathways that play a role in the regulation of important agronomic and economically valuable processes (Caldana et al., 2012). In addition, important conclusions regarding carbon partitioning can be drawn and biomarkers can be identified in order to improve sugarcane crops for the bioenergy industry, as well as identify value-added products which could be used by other industries and biorefineries. This is referred to as taking a systems biology approach which aims to relate genes, metabolites and the respective metabolic pathways in which they take part (Sheth & Thaker, 2014).

By understanding how sugarcane responds to changes in their external environment and by studying the dynamic metabolic changes occurring throughout development, insights can be gained which could aid in crop management, crop engineering, and overall crop development. As mentioned, metabolomics can be used to elucidate these factors of interest (Gupta & Singh, 2013; Jorge & António, 2018; Kliebenstein, 2007), however more research is needed to achieve a better understanding of the sugarcane metabolome and how it changes throughout development, in response to environmental changes, as well as between different tissues and genotypes by focusing on metabolites, and the genes and the enzymes they code for (Cardozo & Sentelhas, 2013).

## **1.2. Objectives**

### **1.2.1. Main objective**

Study sugarcane variety SP80-3280 throughout development on a morphological and molecular level to gather insights with regards to processes that govern its development, and to determine the impact of changes in climatic conditions on its cultivation.

### **1.2.2. Specific objectives**

- Analyze the morphological and agro-technical traits of SP80-3280 planted in two different seasons;

- Establish and optimize protocols for the extraction of soluble sugars and metabolites from leaf (L1), upper internode (I1), young internode (I5) and mature internode (I9) tissue from SP80-3280;
- Establish and optimize a method for the global analysis of the SP80-3280 metabolome (untargeted metabolomics analysis);
- Generate global metabolic profiles of 4 tissues (L1, I1, I5, I9) from SP80-3280 planted in two different seasons;
- Analyze transcriptomics data (microarray) of 4 tissues (L1, I1, I5, I9) from SP80-3280 planted in two different seasons;
- Implement Multi-Omics Factor Analysis (MOFA) for the integration and validation of metabolomics and transcriptomics data;
- Identify which and how metabolic pathways are altered throughout development in SP80-3280 planted in two different seasons.

### 1.3. Supporting literature review

#### 1.3.1. Sugarcane development, phenology and cultivation

Sugarcane development is correlated with the amount of rain and different temperatures that they are exposed to (Moore et al., 2013), however the main studies regarding the effects of different climatic conditions on sugarcane development were published a fairly long time ago (Alexander & Samuels, 1968; Clements, 1962; Glasziou et al., 1965; Legendre, 1975). The information within these studies have not been consolidated in ways that allow their use in crop planning and decision-making based on the current conditions of the sugar-ethanol sector in Brazil (Cardozo & Sentelhas, 2013).

To ensure proper development and high sugarcane crop yields, the planting and the production environment need to be taken into account throughout all of the different phenological stages. The germination of sugarcane seeds occurs when temperatures reach approximately 36 °C, with high moisture and humidity in the surrounding air and soil. For sprouting, the ideal temperature is between 28-30 °C and a moist environment is required. It is characterized by a rapid increase in respiratory activity and the active transport of solutes to the growing extremities of the plant. It is however also possible for buds to sprout from ratoon crops where they emerge faster since they are already attached to a functioning root system (Moore et al., 2013). This referred to as asexual propagation and it is the most common form of cultivating sugarcane across the globe (Getnet, 2017).

Vegetative growth occurs in several stages, starting from a primary shoot where leaves emerge, followed by culm elongation and the formation of phytomers consisting of a node (to which a leaf is attached), internodes and axillary buds (Moore et al., 2013; Simões et al., 2005). Tillering commences from the side shoots forming from the axillary buds on internodes with little expansive growth, with the optimal temperature for tillering being 30 °C. Temperatures below 20 °C slow down this process and tillers formed earlier in development are usually thicker. Culm elongation occurs producing cells that subsequently expand in the intercalary meristem when the leaf attached to the base of an internode has fully expanded, with cooler temperatures and reduced moisture contents leading to shorter internodes (Ebrahim et al., 1998; Moore & Botha, 2013). Sucrose initially accumulates in the lower internodes while the top internodes of the culms are still expanding (Uys et al., 2007). Finally, the reduction of the emergence of leaf

sheaths and internodes indicates that flowering is about to commence. The transition between vegetative and reproductive growth at the apex occurs when there is a reduction in the photoperiod or under stress conditions when the culms have already reached maturation (Coleman & Others, 1969; Melloni et al., 2015). This is however unfavourable in large scale sugarcane cultivation since the carbon flow is directed away from biomass and/or sucrose production, towards inflorescence development (Botha, 2007).

In Brazil, sugarcane can be planted all year round, however, the water availability and the characteristics of the specific variety should be taken into account. Sugarcane plants are mainly planted in two different periods in South-Eastern Brazil and the cultivation times vary for each (de Oliveira et al., 2019). When planting is carried out between September and early December, the term “one-year” sugarcane is used since the crops are harvested around 12 months after planting. These crops present productivity estimates of less than 100 t/ha. On the other hand, when planting is carried out between early January and March or April, the crops are grown for a longer time period hence the name “one-and-a-half-year” sugarcane and higher yields (above 120 t/ha) are usually observed, and it is considered the most common planting and harvesting strategy currently employed in Brazil (de Medeiros Barbosa, 2015). These higher yields are attributed to the longer growing periods. However, the mechanisms that regulate sugarcane development and the shifts between the different phenological stages still require more in-depth research.

The variety used here, SP80-3280, a variety derived from interspecific crosses between the ancestral species *S. officinarum* and *S. spontaneum*, is a widely cultivated variety that has been planted in various regions of Brazil, rendering a cultivation area of about 300.000 ha in total. It presents a good ratooning capacity and requires humid and fertile soil for optimal productivity (Margarido & Santos, 2015).. It has been the variety of choice for conducting genomics (Figueira et al., 2012; Souza et al., 2011) and transcriptomics (Nishiyama et al., 2013; Souza et al., 2019) studies over the past decade, and more importantly, a recently published gene-space assembly has been conducted (Souza et al., 2019). Due to these above-mentioned reasons, this variety has been bred for its adaptability and yield stability traits.

### **1.3.2. Untargeted and targeted metabolomics studies**

The field of metabolomics, which refers to the study of the global constituents called “metabolites” within a biological system (e.g. cell, tissue, organ or organism), is considered a relatively new “omics” study as compared to the other “omics” such as genomics, transcriptomics and proteomics (Allwood et al., 2008). Thanks to the surge in the interest in this field in recent years, metabolomics has contributed greatly to the understanding of various physiological and biological processes of biological systems, specifically from a biochemical point of view which reflects the final outcome of all biological activities within a biological system (Fiehn, 2002; Rochfort, 2005; Roldan et al., 2014). Metabolites have been defined as the total low molecular weight (< 1,000 Da) endogenous compounds present within a biological system (Fiehn, 2002; Roldan et al., 2014). They are the end products of all cellular regulatory processes within a biological system, thus they provide a functional readout or snapshot of the metabolic state of a biological system at a specific point in time (Fiehn, 2002). These metabolites constitute the ‘metabolome’ of the biological system and are considered the furthest downstream product of the genome, and the closest link to the observed phenotype (Forsberg et al., 2018). The study of the metabolome is referred to as metabolomics and this field of “omics” is divided into two categories: untargeted and targeted metabolomics studies (Fiehn, 2016; García-Villalba et al., 2015), each with their own advantages and disadvantages. An untargeted metabolomics approach is taken to measure the relative abundance of as many metabolites as possible within the biological system under study, in an unbiased way (Schrimpe-

Rutledge et al., 2016). Most untargeted metabolomic studies use a separation method prior to the mass spectrometric analysis (Griffiths et al., 2010; Koal & Deigner, 2010). The current method of choice for conducting untargeted metabolomics studies is liquid chromatography coupled to high resolution mass spectrometry (LC-MS) due to its high throughput, soft ionization, and good coverage of a wide range of chemically diverse metabolites (Li et al., 2016; Schrimpe-Rutledge et al., 2016).

High-performance liquid chromatography (HPLC) has the ability to separate metabolites of a wide range of polarity (Ardrey, 2003) through either isocratic elution, where the liquid solvents remain constant during the separation (for non-complex sample matrices), or a gradient elution, where the liquid solvent compositions change during the separation (for more complex samples matrices) leading to improved analysis and results, as was the case in this study.

In mass spectrometry, a mass spectrometer is used that mainly consists of three in tandem components: an ion source, a mass analyzer, and a detector. The ion source ionizes (adds a charge to) the metabolites that reach it after the chromatographic separation step, the mass analyzer then resolves them prior to the measurement by the detector (Ardrey, 2003; Fiehn, 2016). High-resolution mass spectrometers such as time-of-flight analyzers coupled to an electrospray ionization (ESI) source (which was used in this study) improve metabolite identification and provide accurate metabolite quantitation. It is currently the ionization method of choice in LC-MS based studies due to its “soft ionization” ability, providing intact metabolites for detection and facilitated metabolite identification (Chen et al., 2002; da Silva et al., 2017; Jakabová et al., 2012).

During LC-MS analyses, thousands of peaks are detected from a single biological sample. Most of them are overlapped and after segregation by computer data processing, each detected peak is referred to as a “metabolite feature” which corresponds to an ion with a unique mass-to-charge ratio ( $m/z$  value) obtained from the mass spectrometer and retention time (rt) obtained from the chromatographic separation (Ardrey, 2003). This leads to extremely large and complex datasets that require multiple data pre-processing steps prior to biological interpretation (Lämmerhofer & Weckwerth, 2013; Schrimpe-Rutledge et al., 2016).

The pre-processing steps are as follows (Supplementary Figure 1): (i) outlier screening is carried out to eliminate the LC-MS runs that deviate from the majority of the samples. Once the final sample set is determined, (ii) a filtering step is conducted that suppresses the noise while preserving the peaks in the data. (iii) A baseline correction algorithm is then selected to subtract the baseline from the raw signals. (iv) Peak detection is carried out so that each ion in each sample is represented as a peak. Since chromatographic column degradation and changes in the pH and temperatures of the environments and/or mobile phases can lead to rt drifts across samples, (v) peak matching and retention time alignment is an essential step that needs to be carried out to compare metabolomics data across samples. Following the identification of ions originating from the same metabolite, (vi) they are grouped together in the ion annotation step, and (vii) metabolite identification and annotation can then be done. The relative abundance of each metabolite needs to go through (viii) a normalization step to reduce the systematic variation of LC-MS data. The final step in the preparation of the metabolomics data for subsequent statistical analysis and biological interpretation is a step where (ix) data transformation and scaling techniques are implemented to produce more normally-distributed datasets with reduced heteroscedasticity.

The major drawback of using LC-MS for untargeted metabolomics studies is the inability to distinguish between structurally similar metabolites within the sample of interest (Ardrey, 2003). Thus, untargeted metabolomics studies are also known as hypothesis-generating studies (Quanbeck et al., 2012) in which certain observations are made,

leading to new hypotheses that sprout new ideas, projects and approaches for further validation steps. This is where targeted metabolomics come into place.

In contrast to untargeted metabolomics studies, targeted metabolomics studies are focussed on a specific metabolite and/or set of metabolites that have previously been characterized and validation steps are required to characterize and quantify these features of interest (Griffiths et al., 2010; Koal & Deigner, 2010). Since the chemical properties of the features of interest are already known, sample preparation can be specifically tailored to isolate the metabolite or metabolite class of interest (Smith et al., 2006; Vettukattil, 2015), thus reducing matrix effects caused by other accompanying metabolites within the samples (Koal & Deigner, 2010).

The most common forms of metabolite validation are liquid chromatography-tandem mass spectrometry (LC-MS/MS). Coupling LC to MS/MS reduces sample complexity by firstly separating the metabolite(s) of interest in a chromatographic column and generating unique retention time(s) for the metabolite(s), followed by the generation of MS/MS spectra which can then confirm the true identity of the metabolite(s) (He & Aga, 2019; Koal & Deigner, 2010).

### 1.3.3. Sugarcane metabolomics studies

The first global sugarcane metabolomics study was conducted by three South African scientists in 2003 using gas chromatography coupled to mass spectrometry (GC-MS) in attempt to identify as many metabolites as possible, however a mere ~30 metabolites were detected, including trehalose, galactosyl, and raffinose, which are already known to be involved in the mechanisms that regulate sucrose accumulation in sugarcane (Bosch et al., 2003). Four years later, Glassop et al. (2007) conducted a metabolomics study aiming at determining which changes in the sugarcane metabolome are linked to sucrose accumulation. In this study, 55 uniquely identified metabolites were detected using GC-MS. Higher amounts of organic acids, amino acids, and tricarboxylic acid cycle intermediates were observed in the immature internodes and they decreased as the sugarcane plants matured. A metabolomics study done on the SP80-3280 variety using GC-MS reported the presence of 70 metabolites in the +1 leaf tissue including sugars, alcohols, phenols, amino acids, organic acids, inorganic and fatty acids (Ribeiro et al., 2012). Similarly, the use of GC-MS by Ferreira et al. (2018) for the metabolic profiling of sugarcane culm and bud tissues allowed for the isolation and identification of 66 metabolites with known structures. More recently, Ali et al. (2019) combined nuclear magnetic resonance (NMR), LC-MS and GC-MS to set up a metabolic profile of the juice of *S. officinarum* which led to the identification of 42 metabolites.

These metabolite numbers reported are still far below the estimated number of structurally and biologically diverse plant metabolites of between 200,000 and 1 million (Dixon & Strack, 2003; Fang et al., 2019; Rai et al., 2017). Thus, the wide chemical diversity of metabolites present in sugarcane (and plants in general) and their wide range of concentrations, highlights the need to dig deeper into the sugarcane metabolome, strengthening the justification of this PhD project presented here. It was also made evident that more research is needed to achieve a better understanding of the sugarcane metabolome and how it changes throughout development, in anatomically different tissues, and in response to environmental changes.

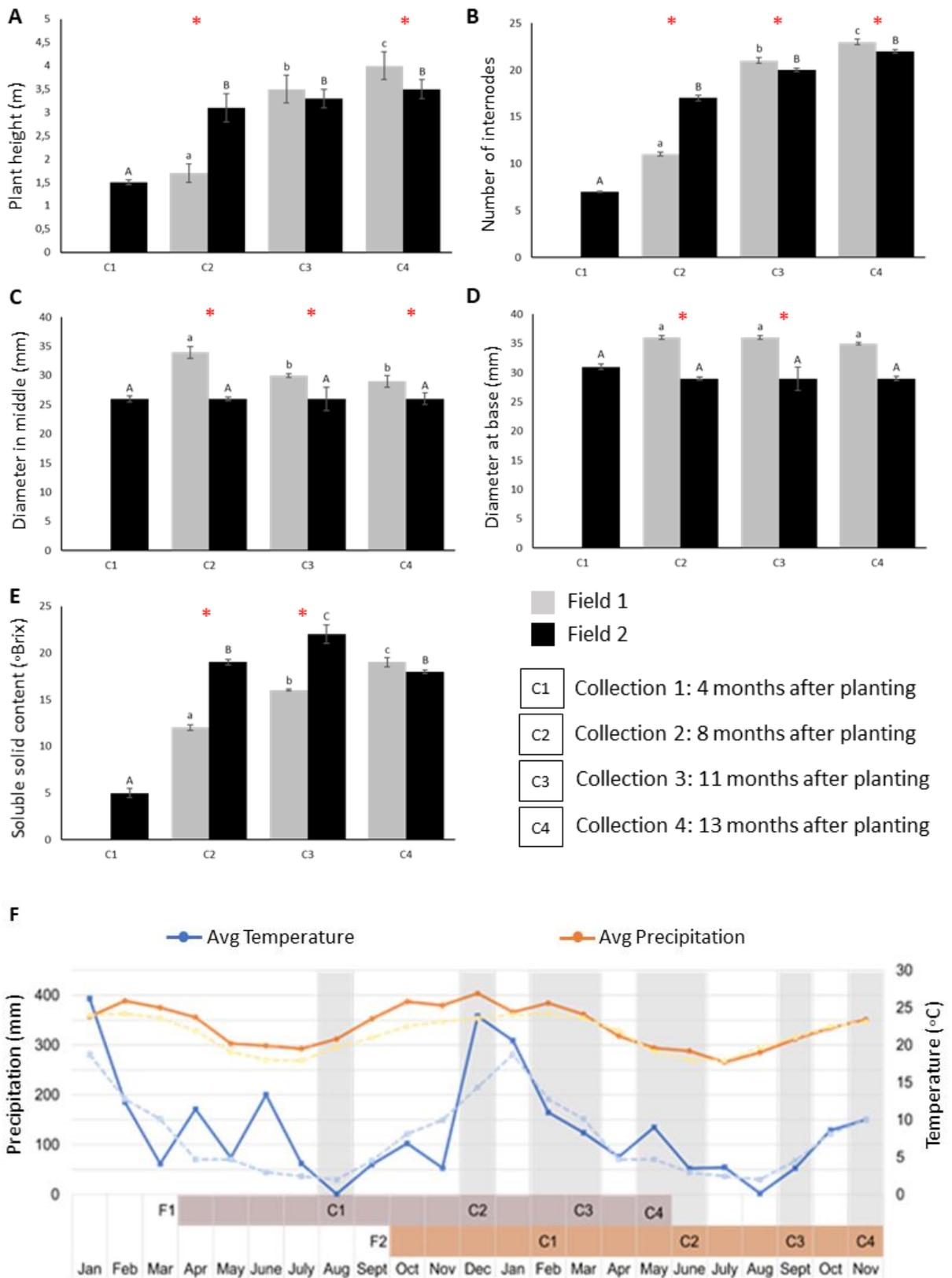
#### **1.4. Previous results obtained from studies directly associated with the PhD project presented here carried out by the Sugarcane Signal Transduction Laboratory**

The thesis presented here was a collective effort involving several groups of different expertise as part of a thematic project (FAPESP process number 2014/50921-8). The group collected samples, morphological and agrotechnological data from the commercial sugarcane variety, SP80-3280 chosen as a reference variety in several previous studies (for example, Papini-Terzi et al., 2005; Inada, 2016; Souza et al., 2019). Currently, SP80-3280 ranks among the top 20 sugarcane varieties grown in the state of São Paulo, Brazil, and is being used as a genitor in Brazilian breeding programs (Manechini et al., 2018). In addition, it is a model for large scale genomic analyses and is the cultivar with the largest collection of transcriptomics data available to date (Vettore et al., 2003). An assembled gene space of this variety has also recently been produced after roughly ten years of collaborative research led by the Sugarcane Signal Transduction Laboratory and GaTE Lab (Souza et al., 2019, Van-Sluis et al., 2019), thus possibly transforming this variety into the most complete model currently available for sugarcane studies.

The present work focused to analyse an SP80-3280 field experiment where two separate fields were analysed for 13 months (Inada, 2016; Materials and Methods - Figure 2 A). The plantings of SP80-3280 were carried out in two seasons of the year with different climatic conditions (Materials and Methods - Figure 2 B) at the Agricultural Sciences Center of the Federal University of São Carlos in Araras – SP (UFSCAR). Details regarding the plant material, experimental field conditions, sample collections and sample preparations are described in the materials and methods section and the dissertation by Davi Inada (2016).

Morphological analyses were conducted as per Consecana (2006) by other members of the group. The height of each sugarcane was measured in centimetres (cm) from the base of the culm to the apical meristem. The numbers of internodes per culm were determined by direct counting in the field. The diameter of the culms was measured in millimetres (mm) using a digital capilar from the base of the stem to the tenth internode. The Brix content of the sugarcane juice in each culm was measured in degrees Brix (°Brix) using a specific portable refractometer (N1 model, ATAGO, Japan).

The sugarcane plants field 1 (F1) experienced an initial growth period under low average temperature and precipitation, whereas the plants from field 2 (F2) experienced an initial growth period under high average temperature and precipitation (Material and Methods - Figure 1 F). These different climatic conditions led to different developmental profiles and morphologies of the plants from the two different fields. Specifically, the plants from F1 presented retarded vegetative growth and elongation during the initial four months after planting since no internodes were formed at the first collection (C1), thus no measurements of the height and the number of internodes could be taken (Figure 1 A and B).



**Figure 1.** Morphological and degrees Brix results (Inada, 2016). (A) Plant height in meters; (B) the number of internodes; (C) diameter in the middle in millimetres; (D) diameter at the base in millimetres; (E) soluble solid content in degrees Brix. (F) Climatic conditions for the field experiments. Fields 1 (F1) and 2 (F2) refer to the plants that were planted in April 2012 and October 2012, respectively. C1, C2, C3, and C4 refer to collection points 1, 2, 3 and 4 specifically 4, 8, 11 and 13 months after planting. Different letters on the graphs indicate statistically significant differences between the different collection points based on ANOVA and Fisher's LSD ( $p < 0.05$ ). The asterisks indicate statistically significant differences between the different fields based on Student's t-tests ( $p < 0.05$ ); (F) Climatic conditions which sugarcane variety SP80-3280 experienced throughout development in the two field

experiments. The average precipitation and temperature measurements during the two field experiments are plotted as continuous lines, respectively.

At the second collection (C2) of F1, that is eight months after planting, internodes were visible indicating that the vegetative growth and elongation phase had been initiated after which the plants continued to grow and elongate, even following the third (C3) and fourth (C4) collections, eleven and thirteen months after planting, respectively.

The sugarcane plants from F2 presented a rapid vegetative growth and elongation phase during the initial four months since internodes were already visible at C1 and the height of the plants could be measured at this stage (Figure 1 A and B). At C2, C3 and C4 it was observed that this growth and elongation phase had stabilized between eight and eleven months after planting since no increases in plant height were observed throughout C2 to C4.

Comparing the plants from the two different fields to each other at C2, it was evident that the sugarcane plants from F2 were significantly taller, had more internodes and presented a narrower culm diameter which persisted throughout development (Figure 1 A - D). However, at C4 the plants from F1 had surpassed the height and number of internodes of those from F2 significantly (Figure 1 A and B), suggesting that they were still in the vegetative growth and elongation phase, whereas the plants from F2 experienced the shift between this phase and the maturation phase between C2 and C3. These findings were also supported by analyzing the changes in the Brix contents over time (Figure 1 E).

The plants from F1 presented a continuous significant increase in the Brix content over time, however, it was always lower than the plants from F2 in which the Brix content had reached its maximum at C3 after which it decreased again.

In addition to the morphological and Brix measurements previously obtained by the group (Figure 1), an untargeted metabolomics study was conducted using the L1 tissue samples (Inada & Souza, 2016). Interestingly, a metabolite feature with the  $m/z$  value 341.00, which could represent either sucrose, trehalose, cellobiose, coniferin, epimelibiose, maltose, melibiose or galactinol, were found to be two times higher in L1 from F1 at C3 as compared to C2. During this time, that is between December 2012 (C2) and March 2013 (C3), there was a significant decrease in the precipitation and temperature measurements (Figure 1 F). This same metabolite feature of  $m/z$  341.00 was found to be three times higher in L1 from F2 at C3 as compared to C2. Again, during this time, that is between June (C2) and September 2013 (C3), the lowest precipitation and temperature measurements of the entire experiment were recorded.



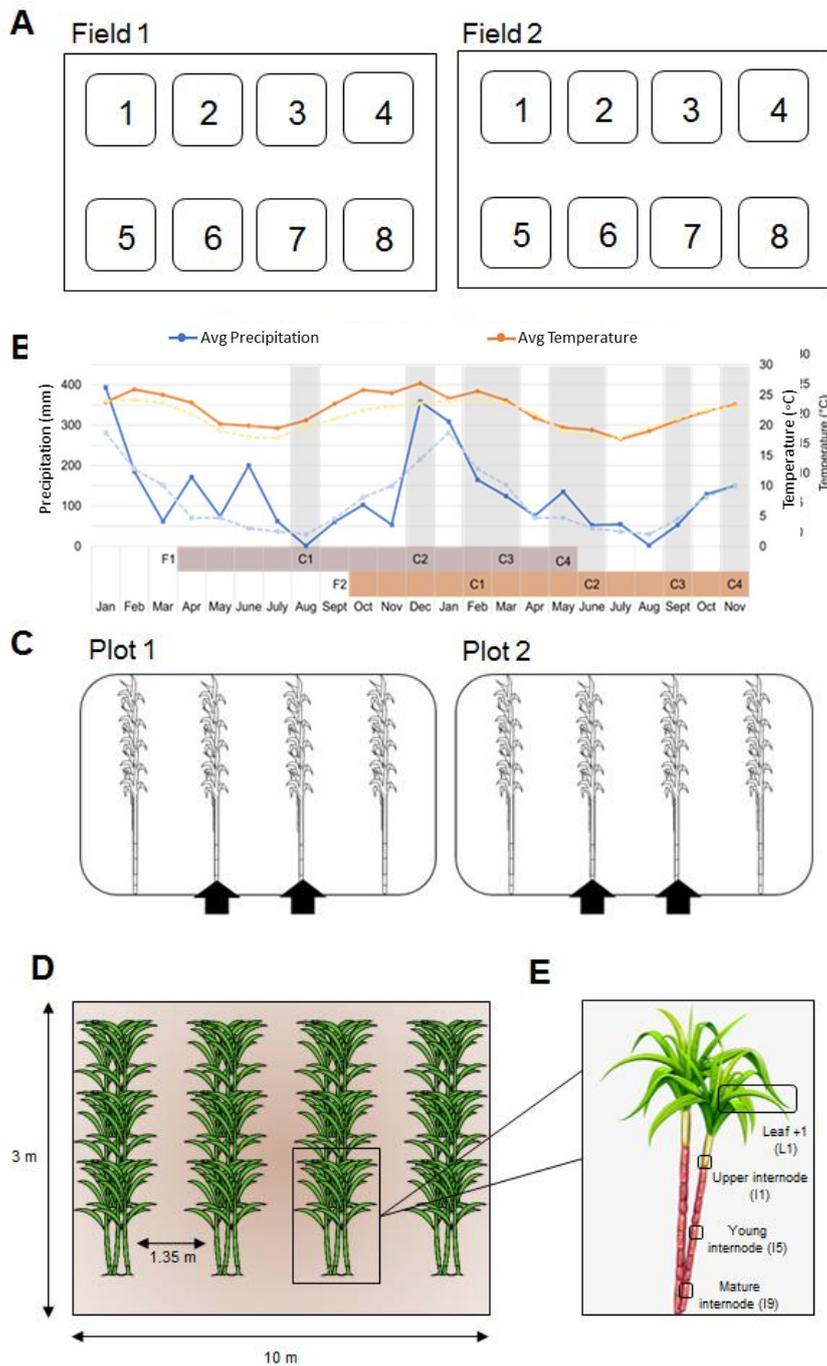
## **2. MATERIALS AND METHODS**

### **2.1. Plant material**

Different tissues, namely the leaf +1 (L1) and upper (I1), young (I5) and mature (I9) internodes from the commercial sugarcane variety SP80-3280, a variety derived from mainly interspecific crosses between the ancestral species *S. officinarum* and *S. spontaneum* (Souza et al., 2019), were used in the PhD project presented here.

### **2.2. Experimental design, field conditions and sample collection**

Two experimental fields (Figure 2 A) were prepared with eutroferric oxisol soil fertilized with 400 kg ha<sup>-1</sup> of N:P:K mixed in a ratio of 5:25:25. Each field experiment consisted of eight plots (10 m x 3 m) in which the sugarcane billets were planted in four rows with 1.35 m spaces between the furrows in each plot (Figure 2 C and D).



**Figure 2.** Schematic representation of the field experiments. (A) The two field experiments and the eight plots within each field. Fields 1 (F1) and 2 (F2) refer to the plants that were planted in April 2012 and October 2012, respectively; (B) Climatic conditions which sugarcane variety SP80-3280 experienced throughout development in the two field experiments. The average precipitation and temperature measurements during the two field experiments are plotted as continuous lines, respectively; (C) The plots from which the tissues were collected. The arrows indicate the rows from which the tissues were collected; (D) One plot indicating that the plants were planted in four rows of 10 m with 1.35 m between the furrows; (E) Tissue samples that were collected namely leaf +1 (L1), upper (I1), young (I5) and mature (I9) internode.

The first field experiment (F1) was initiated in April 2012 and extended up until May 2013, representing the conditions under which “one-and-a-half-year” sugarcane crops are cultivated. The second field experiment (F2) with the same technical specifications was initiated in October 2012 and extended up until November 2013\*, representing the conditions under which “one-year” sugarcane crops are cultivated (Figure 2 B). No irrigation was applied in this

study. However, during these periods the plants experienced different climatic conditions with regards to precipitation and temperature (Figure 2 B). Tissue samples, specifically L1, I1, I5 and I9 tissues (Figure 2 E), were collected from the two middle rows (Figure 2 C) and two plots from both fields at 4, 8, 11 and 13 months after planting (Figure 2 B) for metabolomics and transcriptomics studies. The tissue samples were cleaned with water (H<sub>2</sub>O) and 70% v/v ethanol (EtOH) prior to the submersion in liquid nitrogen for quenching purposes to bring the biological activities within the cells to a complete stop. Once the samples were frozen, they were placed in 50 ml Falcon tubes on dry ice for transportation purposes, after which they were stored at -80 °C prior to further sample processing and analyses.

\*Disclaimer: spontaneous flowering was observed in the plants from field 1 (F1) at collection 4 (C4), 13 months after planting.

### **2.3. Sample preparation**

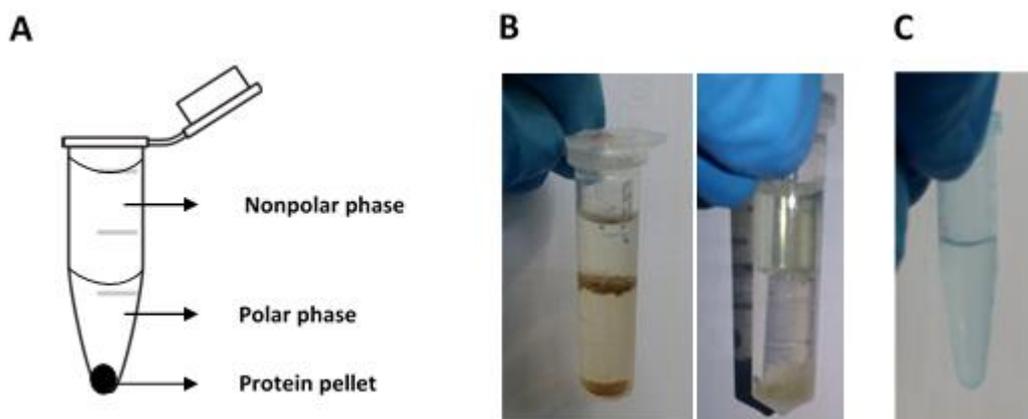
The samples were prepared by using eight replicates of each tissue (L1, I1, I5, I9) and collection point (C1, C2, C3, C4), four from one plot and four from another plot. Pools of two samples each were made to reduce within-sample variation, rendering four biological replicates. The samples were homogenized to a fine powder using a mortar and pestle for L1 and I1, and a Cuisinart Coffee Grinder (model DCG-20) for I5 and I9. Aliquots of approximately 50 mg tissue powder (exact weights and sample pools are reported in the Supplementary Table 1) were placed in 2 ml clear microtubes (Axygen Scientific) and placed in the -80 °C freezers prior to the further analyses.

## **2.4. Untargeted metabolomics analyses of four anatomically different sugarcane tissues**

### **2.4.1. Samples used and global metabolite extractions**

The same homogenized tissue powder that was made and used for the sugar analysis (Section 2.3), was used for the global metabolite extraction. The samples were removed from the -80 °C freezers and kept on dry ice directly prior to the global metabolite extractions.

An extraction method proposed by Salem et al. (2016) was used in this study. One ml of extraction solution, methyl tert-butyl ether (MTBE):methanol (MeOH) mixed in the ratio 3:1 and kept at -20 °C, was added to each tube containing the aliquots, vortexed and placed on ice on a reciprocal shaker (Labline, model 3506 000554) for 45 minutes. A volume of 650 µL of a H<sub>2</sub>O:MeOH solution mixed in the ratio 3:1 and kept at 4 °C was added to each tube, followed by the sonication at 40 kHz for 20 minutes. The tubes were vortexed again, followed by the centrifugation at 20,000 g for 5 minutes at 4 °C, using an Eppendorf centrifuge 5810 R to allow for phase separation. The top layer represented the nonpolar phase (Figure 3 A and B) and the bottom layer represented the polar phase (Figure 3 A, B and C).



**Figure 3.** Global metabolite extractions. (A) Representation of the metabolite extraction which separates the extract into three phases, the top layer represents the nonpolar phase, the bottom layer represents the polar phase and the pellet represents the biomass and protein excess; (B) Phase separations after the addition of both extraction buffers MTBE:MeOH (3:1) and H<sub>2</sub>O:MeOH (3:1); (C) The polar phase after the phase separation step.

The phases were collected separately and placed in 2 ml microtubes (Axygen Scientific) in which they were stored at -80 °C prior to the HPLC-MS analysis (Figure 3 C). The pellet of each tube was dried down in using speedvac for 1 hour at 60 °C and weighed again for normalization purposes (Supplementary Table 1).

#### 2.4.2. HPLC-MS analysis and data acquisition

In order to separate, ionize and detect the polar and semi-polar metabolites in each sample, high-performance liquid chromatography (UFLC, Shimadzu, Inc.) coupled to the high-resolution time-of-flight mass spectrometer (Bruker Daltonics, Inc.) was implemented. A pentafluorophenyl propyl ligand column (Phenomenex Luna PFP, 100 mm x 4.60 mm x 2.6 μm) was coupled to an ESI-microTOF mass spectrometer, operating in the positive and negative ionization modes. MPA and MPB used for compound separation and detection in the positive ionization mode were 0.1% formic acid (FA) in H<sub>2</sub>O with a pH of 2.65 (MPA), and MPA:MeOH:ACN mixed in the ratio 5:25:70 (MPB). On the other hand, MPA and MPB used for compound separation and detection in the negative ionization mode were 20 mM NH<sub>4</sub>Ac in water with a pH of 7.35 (MPA) and MPA:MeOH mixed in the ratio 5:95. A volume of 200 μL of each sample was transferred to a glass insert inside a glass vial prior to the analysis. A volume of 10 μL was injected into the HPLC-MS system and the flow rate was set at 500 μL min<sup>-1</sup>. Gradient runs were conducted, and the information is shown in Table 1.

**Table 1.** Gradient information for the high-performance liquid chromatography coupled mass spectrometry (HPLC-MS) runs using a pentafluorophenyl propyl ligand column (Phenomenex Luna PFP, 100 mm x 4.60 mm x 2.6  $\mu$ m).

Time (minutes)	% of pump A	% of pump B
0.01	98	2
1.00	98	2
15.00	47	53
15.50	10	90
20.00	10	90
23.00	98	2
25.00	0	0

### 2.4.3. Untargeted metabolomics data generated with HPLC-MS - data processing

The raw HPLC-MS data were extracted and converted to the .mzXML file format after which they were separated based on tissue type to prevent possible matrix effects since the four tissues are anatomically and metabolically different.

The tools and packages used here were all implemented using the R programming language. The first package used in the workflow for the processing of the raw HPLC-MS data was XCMS (various forms (**X**) of **C**hromatography **M**ass **S**pectrometry). This package was used for the first step in the pre-processing of this HPLC-MS data, such as peak filtration and identification, baseline correction, matching the peaks across the different samples, correcting of the deviation in retention times and finally filling in the missing peak information (Smith et al., 2006).

The CAMERA package (**C**ollection of **A**lgorithms for **M**etabolite **p**rofile **A**nnotation) which groups all features derived from the same compound, annotates the specific ion species and proposes accurate compound masses, was then implemented for ion, adduct and fragment annotation (Kuhl et al., 2012).

The outputs generated were then used as inputs for ProbMetab to predict the most likely compounds present in each tissue (Supplementary Tables 2 – 5), thus allowing for a more robust and accurate metabolic profile of each tissue collected. It uses a Bayesian model to provide a list of compound candidates ranked by their respective probabilities. This package was developed specifically for the annotation of LC-MS data by using a set of known chemical reactions between candidate compounds since combinations which are detected together would make more biochemical sense than others (Silva et al., 2014). This package also allowed direct links to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, mapping each metabolite to their respective pathways. The intensity of each predicted metabolite was normalized to the dry weight (DW) of the sample (Supplementary Table 1) and selected for the downstream analyses.

### 2.4.4. Normality testing, data dimensionality reduction and model validation

Raw metabolomics data require pre-processing, in this case via the use of XCMS, CAMERA and ProbMetab, in order to generate “clean” data which were the inputs for the data pre-treatment steps. Different data transformation and scaling techniques were explored in order to select the optimal data pre-treatment methods which

would allow for the untangling of the observed biological variance and to exclude any possible technical variance introduced throughout the metabolomics analyses.

As mentioned, the intensities of each feature were normalized by the DW after which the different transformation techniques, such as no transformation, log<sub>10</sub> transformation, square and cube root transformations, were tested. Shapiro Wilk's normality tests (Vinaixa et al., 2012) were conducted in the R programming language to select the transformation technique rendering the most normally distributed data. The cases in which the null hypothesis of the Shapiro Wilk's normality test, which states that the population is normally distributed (Vinaixa et al., 2012), could not be rejected, thus the cases in which the p-values were found to be above 0.05, were counted and the percentages were calculated in order to determine which transformation serves as optimal for biological interpretability. QQ-plots were also generated for each tissue to allow for visual inspection and to represent the need for data transformation.

Auto and pareto scaling methods were applied to further improve data quality after the transformation steps were implemented. Firstly, the Shapiro-Wilk's normality tests were performed on these transformed data sets after scaling, and secondly, the results were analysed via the generation of Principal Component Analysis (PCA) models using MetaboAnalyst 4.0 (Xia et al., 2018).

#### **2.4.5. Metabolomic tissue profiles – discriminant analyses**

Once the optimal metabolomics datasets (processed, normalized, transformed, and scaled metabolomics datasets) were obtained, they were used as inputs for the construction of PCA and Partial Least Squares Discriminant Analysis (PLS-DA) models in MetaboAnalyst 4.0. The models were investigated, and from the PLS-DA models, the metabolites identified with VIP scores above 1 were selected for generating heatmaps in MetaboAnalyst 4.0 (Wei et al., 2018; Xia & Wishart, 2011) to highlight the main contributors of the observed separations. This was done in order to determine the main metabolic differences between the four different tissues analysed (tissue comparisons).

The tissues were then separated to identify metabolic differences observed in each tissue between the plants that were planted in different seasons of the year (field comparisons). The same discriminant analyses were thus conducted namely PCA and PLS-DA, as described.

### **2.5. Transcriptomics analyses of four anatomically different tissues from sugarcane**

#### **2.5.1. DNA microarray experiments and data acquisition**

It should be noted that the generation of the transcriptomics data was previously conducted by the group, however, a brief description of the materials and methods is outlined here.

Microarray experiments were conducted as per Lembke et al. (2012) using the same samples that were used for the targeted and untargeted metabolomics analyses as mentioned in the previous sections. These microarray experiments were conducted using custom oligoarrays (CaneRegNet oligoarray) following the Aligent 44K x 4 format and manufacturer's protocol (Lembke et al., 2012).

The samples were prepared by extracting 200 ng of RNA from four biological replicates, two from each plot within each field. This was used to generate cDNA which was labelled with Cy3 and Cy5 dyes using Quick Amp

Two-Colour Labelling Kits (Aligent). For the remainder of the steps the Gene Expression Hybridization Kit, Gene Expression Wash Buffer Kit and RNeasy Mini Kit were used.

For each collection point, two hybridizations per tissue sample were performed. For the leaves, a pool of all leaves and for the internodes, a pool of all internodes were used as references. The oligoarrays were read using the GenePix 4000B scanner device (Molecular Devices) and the fluorescence data were processed by Feature Extraction software 9.5.3 (Aligent). The data were then processed and uploaded via the SUCEST-FUN database (<http://sucest-fun.org/wsapp/>). To generate the datasets to be used for the subsequent analyses, downloadable files were generated specifically by selecting the “search significantly expressed genes” option. Furthermore, the “show genes in at least one experiment” was chosen, which means all transcripts present in at least one sample were selected, and the log<sub>2</sub> intensity option was chosen for generating the downloadable datasets.

### **2.5.2. The usage of the processed microarray data and multi-dimensional scaling (MDS) analyses**

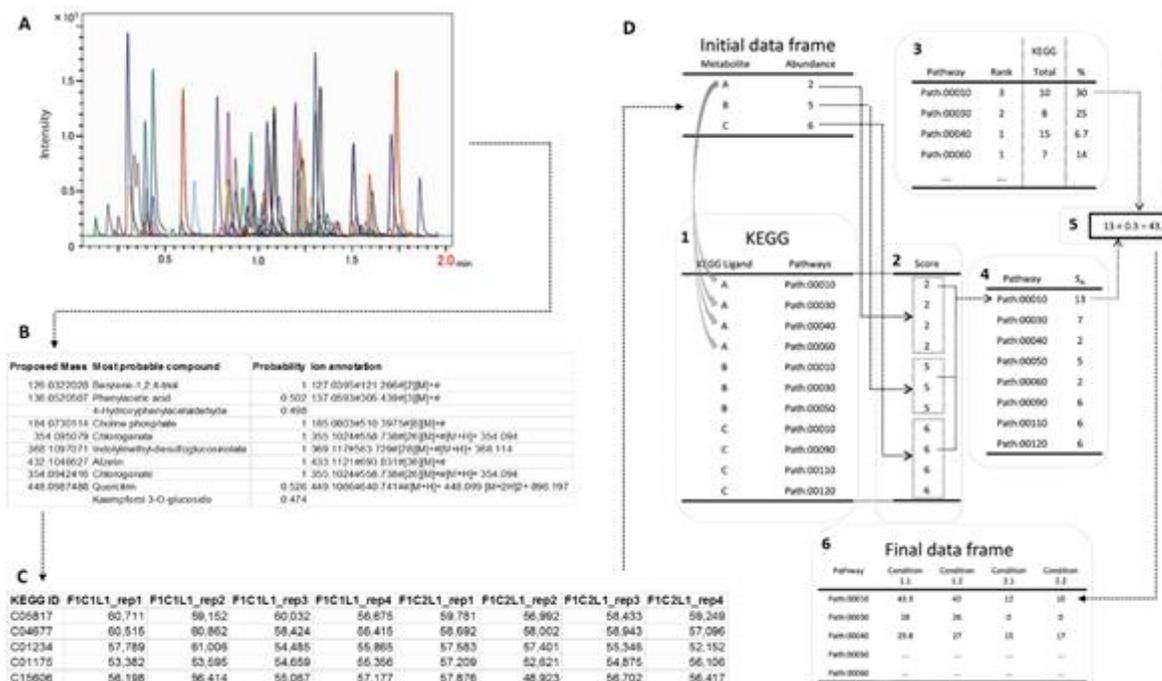
The data mining and analyses of the transcriptomics data were in fact part of the PhD project presented here. Out of the 21,902 different probes represented on the CaneRegNet oligoarray, the significantly expressed transcripts, both sense and antisense, from all replicates from all tissues, collection points and fields, were selected via the SUCEST-FUN platform to be used for MDS analyses. Heatmaps were constructed with the 500 most variable genes within the datasets. Specific gene clusters were detected and gene set enrichment analyses (GSEA) were conducted using the topGO package in the R programming language with the respective gene sets in each cluster to determine the top Gene Ontology (GO) terms related to each cluster (Alexa & Rahnenfuhrer, 2019).

## **2.6. The integration between two data modalities: metabolomics and transcriptomics data**

### **2.6.1. Pathway Activity Profiling (PAPi)**

Using the “clean” metabolomics data generated as described in section 2.5.4, an R package, Pathway Activity Profiling (PAPi), was then implemented in order to determine which metabolic pathways from the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database (Kanehisa, 2000) are represented and activated (based on the pathway activity scores) by the different metabolites detected in each tissue (Aggio et al., 2010). A description of the workflow follows.

After the raw HPLC-MS data were generated (Figure 4 A) and processed, the KEGG metabolites predicted by ProbMetab (Figure 4 B) and their normalized, transformed and scaled intensities in each sample (Figure 4 C) were subjected to PAPi (Figure 4 D) to measure which pathways are represented by the predicted KEGG metabolites in each tissue over time, and how the identified pathways change between the different conditions throughout development (across collection comparisons).



**Figure 4.** Workflow of Pathway Activity Profiling (PAPi) and its inputs adapted from Aggio et al., 2010. (A) Raw data from HPLC-MS is processed to render (B) the most probable metabolite, their annotations and in turn their (C) KEGG compound IDs and relative abundances which are (D) the inputs for PAPi. Step 1: All pathways for which each metabolite is known to play a part are collected from the KEGG database. Step 2: Each identified pathway then receives a score based on the abundance/relative abundance of the metabolite to which it is linked. Step 3: The total number of metabolites associated with each pathway is recorded and the pathways are then ranked according to the number of metabolites with which they are associated. The percentage of detected metabolic intermediates is then calculated for each listed pathway. Step 4: Finally, we sum over the scores for each pathway to obtain the total pathway score,  $S_A$ . Step 5: The  $S_A$  is normalized by dividing by the proportion of metabolites detected from its respective pathway. Step 6: The final data frame is obtained which includes all the pathways and their  $S_A$  within each condition.

## 2.6.2. Selection of the input data for the Multi-Omics Factor Analysis (MOFA)

The activated metabolic pathways identified by PAPi in section 2.6.1. were used to identify and select the transcripts (as described in section 2.5.1.) involved in each of the activated metabolic pathways. The sequences of these transcripts were obtained from the SUCREST-FUN database and were used as an input for blast2GO (Conesa et al., 2005) analyses to retrieve the enzyme codes (E.C.) and to annotate and assign each significantly expressed transcript to the KEGG database. Only the annotated transcripts which matched to at least one KEGG enzyme on at least one of the activated metabolic pathway maps were selected for downstream analyses for the integration between the metabolomics and transcriptomics data modalities using the Multi-Omics Factor Analysis (MOFA) tool (Argelaguet et al., 2018).

## 2.6.3. The implementation of the Multi-Omics Factor Analysis (MOFA)

In order to determine which and where variation is shared between the two data modalities represented in this study, namely metabolomics and transcriptomics data, MOFA (Argelaguet et al., 2018) was implemented using the output results from PAPi (data from section 2.6.1) and the significantly expressed transcripts involved in the respective

activated pathways (data from section 2.6.2) to highlight which and how the metabolic pathways are altered throughout sugarcane development in the field.

All factors, in this case, the activated metabolic pathways and the selected significantly expressed transcripts explaining less than 1% of the variation in both metabolomics and transcriptomics data were excluded from the analysis as per default settings and advice from the creator of MOFA (Argelaguet et al., 2018). The latent factors (LFs) with the most shared variation between the two data modalities were investigated by selecting the top 10 metabolic pathways and the top 300 significantly expressed transcripts shared within each LF. The 300 significantly expressed transcripts were used as an input for GSEA using the topGO package in the R programming language to get the top 10 GO terms related to the respective transcripts (Alexa & Rahnenfuhrer, 2019).

## **2.7. The selection of the targeted metabolic pathways and possible genes of interest**

Based on the results obtained from the integration of the two “omics” using MOFA, and extensive literature reviews, four pathways were selected for further consideration to dig deeper into the sugarcane metabolome and how it is altered throughout development. The pathway activities of these pathways were plotted and analysed using MetaboAnalyst 4.0 (Chong et al., 2018; Xia et al., 2009). Target genes were highlighted by pinpointing the exact metabolites within each of the selected pathways, and enzymes directly upstream or downstream of these metabolites were selected. The transcripts from section 2.6.2 which corresponded to the target genes were selected. Analysis of Variance (ANOVA) and Fisher’s Least Significant Difference (LSD) statistical tests were conducted ( $p < 0.05$ ) in the R programming language and heatmaps were generated using MetaboAnalyst 4.0 (Chong et al., 2018; Xia et al., 2009). To crosscheck the annotation of the selected target genes, the Basic Local Alignment Search Tool (BLAST) was used to perform similarity searches on the National Center for Biotechnology Information (NCBI) database.



### 3. RESULTS

#### 3.1. Untargeted metabolomics analyses of four anatomically different sugarcane tissues

##### 3.1.1. HPLC-MS data processing, annotation and the generation of “clean” metabolomics data for downstream analyses

For the untargeted metabolomics analyses, a total of 108 samples were analysed consisting of leaf +1 (L1), upper internode (I1), young internode (I5) and mature internode (I9) tissues. These samples were collected from SP80-3280 planted in two experimental fields experiencing different climatic conditions (Field 1 – F1 and Field 2 – F2) over periods of 13 months, carrying out collections at 4 months (Collection 1 – C1), 8 months (Collection 3 – C2), 11 months (Collection 3 – C3) and 13 months (Collection 4 – C4) after planting. These samples were used for HPLC-MS analyses and the generation of untargeted metabolomics data.

The first step in the pre-processing of the untargeted metabolomics data generated via HPLC-MS, was the implementation of XCMS which was able to carry out feature detection, nonlinear rt alignment and relative quantification across all samples (Smith et al., 2006). In each tissue, hundreds of metabolic features from XCMS were detected, in both positive and negative ionization modes (Table 2).

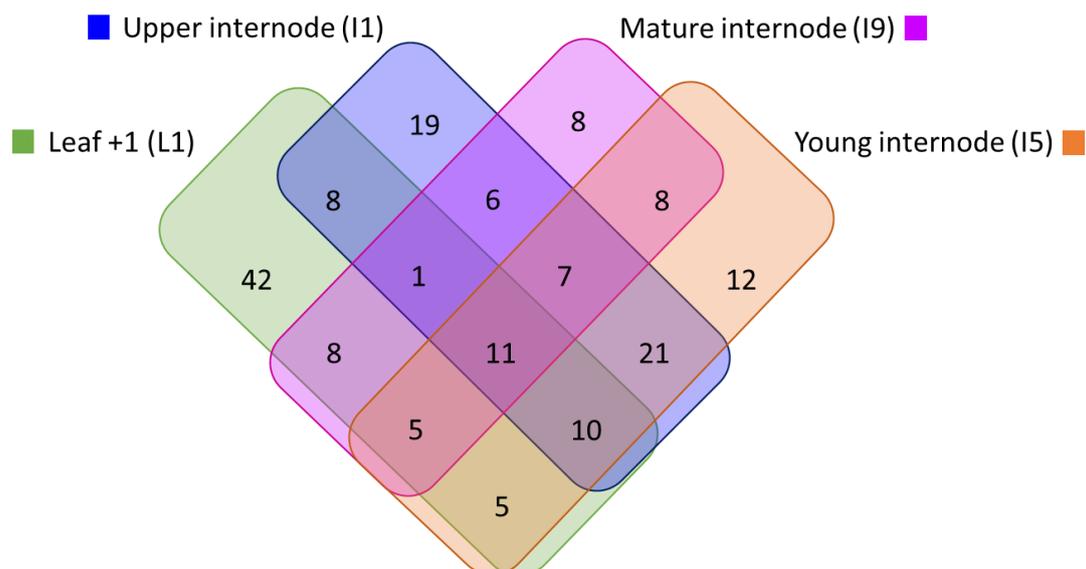
**Table 2.** Number of features detected from XCMS in both positive and negative ionization modes; the corresponding amount of XCMS features with at least 1 KEGG ID; the total possible amount of KEGG IDs detected and the unique KEGG IDs detected in each tissue namely leaf +1 (L1), upper internode (I1), young internode (I5) and mature internode (I9).

Tissue	XCMS pos	XCMS neg	XCMS corresponding to at least 1 KEGG	Possible KEGG IDs	Unique KEGG IDs
L1	965	549	103	158	90
I1	1251	508	98	176	84
I5	978	409	90	190	79
I9	683	292	72	153	54

In the four different tissues analysed, 103, 98, 90 and 72 XCMS features matched to at least one KEGG ID (Table 2). Out of the possible 158, 176, 190 and 153 metabolites suggested by ProbMetab, 90, 84, 79 and 54 metabolites with unique KEGG IDs were predicted in the respective tissues (Supplementary Tables 2 - 5).

Mostly polar and semi-polar compounds were extracted from the different sugarcane tissues since the polar phase of the extraction (Figure 3 C) were selected for HPLC-MS analysis. Metabolites present in both primary and secondary metabolism were extracted and predicted to be present in each tissue (Supplementary tables 2 - 5).

The highest number of metabolites (103 XCMS features matching to 90 unique KEGG metabolites) were detected in L1, the source, decreasing down the stem towards I9, the sink (74 XCMS features matching to 54 unique KEGG metabolites) as shown in Table 2. In L1 most of the metabolites suggested by ProbMetab were tissue-specific (42 unique KEGG metabolites) and merely 11 metabolites were shared among all tissues (Figure 5). The number of tissue-specific metabolites also decreased from the source to the sink tissue.



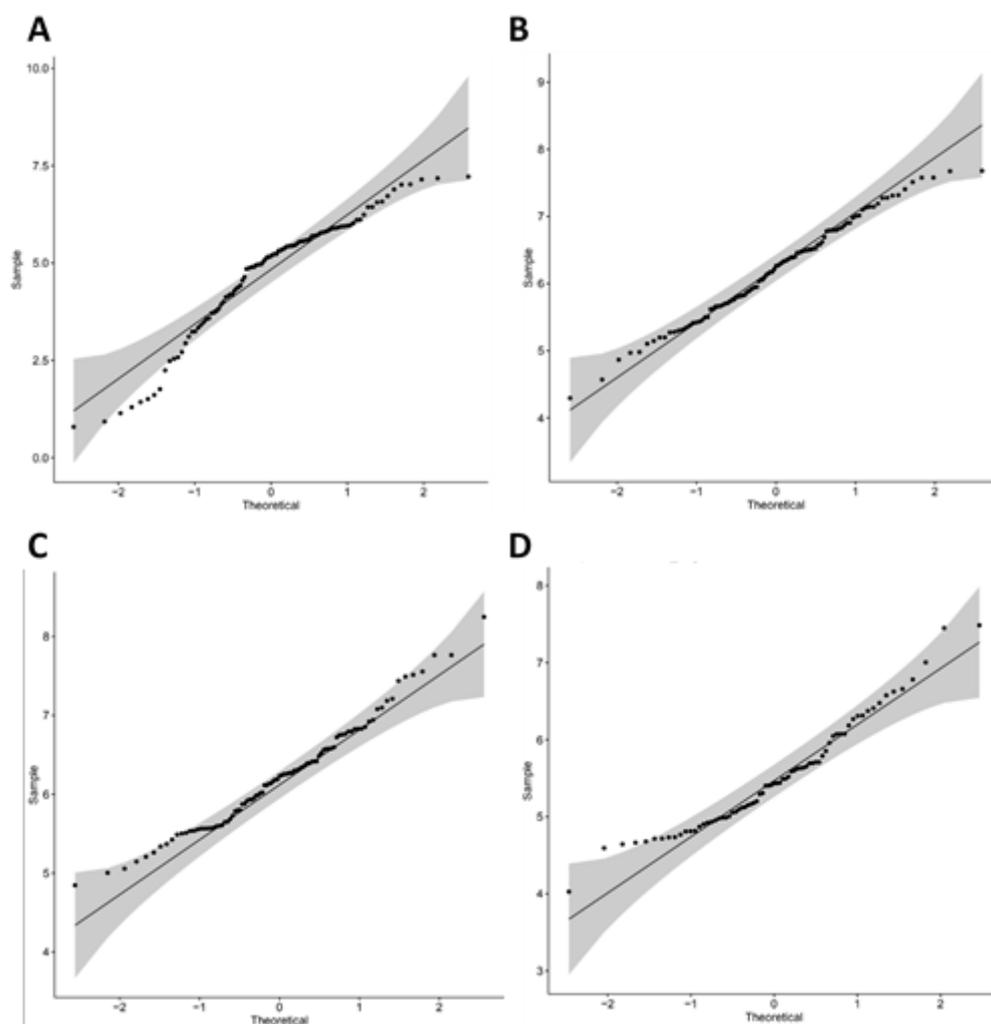
**Figure 5.** Venn diagram of metabolite numbers from the metabolite profiles generated via HPLC-MS of each tissue collected throughout sugarcane SP80-3280 development in the field.

After the normalization of the metabolomics data by dry weight (DW), different data transformations (none, log10, square root and cube root) and Shapiro Wilk's normality tests were done. Based on these tests, it was found that in all tissues the log10 transformation showed the best results in terms of transformation in being able to eliminate some of the heteroscedasticity (refers to the circumstance in which the variability of a variable is unequal across the range of values of a second variable that predicts it) in all of the groups, having 37.50%, 71.40%, 100% and 60% of all samples from each tissue presenting normal distribution after transformation (Table 3).

**Table 3.** Percentages of the differently transformed metabolomics data generated for the leaf +1 (L1), upper (I1), young (I5) and mature internodal (I9) tissue samples in which the Shapiro Wilk's test for normality were  $p > 0.05$ .

Tissue	none	log10	square root	cube root
L1	0.00%	37.50%	0.00%	12.50%
I1	0.00%	71.40%	0.00%	0.00%
I5	0.00%	100.00%	0.00%	0.00%
I9	0.00%	60.00%	0.00%	0.00%

QQ-plots (scatterplots created by plotting two sets of quantiles against one another) were also generated for each tissue to allow for visual inspection and to represent the need for data transformation prior to downstream statistical analyses of untargeted metabolomics data (Figure 6).



**Figure 6.** QQ-plots of the log10 transformed metabolomics data in each tissue. (A) Leaf +1 (L1); (B) upper (I1) internodes; (C) young (I5) internodes and (D) mature (I9) internodes.

After the implementation of auto scaling and pareto scaling on the DW normalized and log10 transformed data, Shapiro Wilk's normality tests were done again. No differences were observed between the two different scaling methods based solely on the Shapiro Wilk's normality tests (Table 4), thus the results had to be further analysed via the generation of PCA models.

**Table 4.** Percentages of the log10 transformed metabolomics data generated for the leaf +1 (L1), upper (I1), young (I5) and mature internodal (I9) tissue samples, in conjunction with auto scaled and pareto scaled metabolomics data, in which the Shapiro Wilk's test for normality were  $p > 0.05$ .

Tissue	auto scale	pareto scale
L1	37.50%	37.50%
I1	71.40%	71.40%
I5	100.00%	100.00%
I9	60.00%	60.00%

The PCA models generated for each tissue were inspected to determine which type of scaling technique would yield the most significant results with regards to the biological question at hand. In the case of L1 and I9, no scaling yielded the most significant models with the sum of the first two components (64.1% + 14.7%; 56.4% + 19.6%) resulting in explaining 78.8% and 76.0% of the total variance of the data, respectively (Table 5). However, for I1 and I5 the auto scaling method yielded the most significant models with the sum of the first two components (44.7% + 13.9%; 36.5% + 22.2%) resulting in explaining 58.6% and 58.7% of the total variance of the data, respectively (Table 5).

**Table 5.** Principal Component Analysis (PCA) results generated for the log10 transformed metabolomics data of the leaf +1 (L1), upper (I1), young (I5) and mature internodal (I9) tissue samples when no scaling, auto scaling and pareto scaling techniques were applied on the metabolomics data.

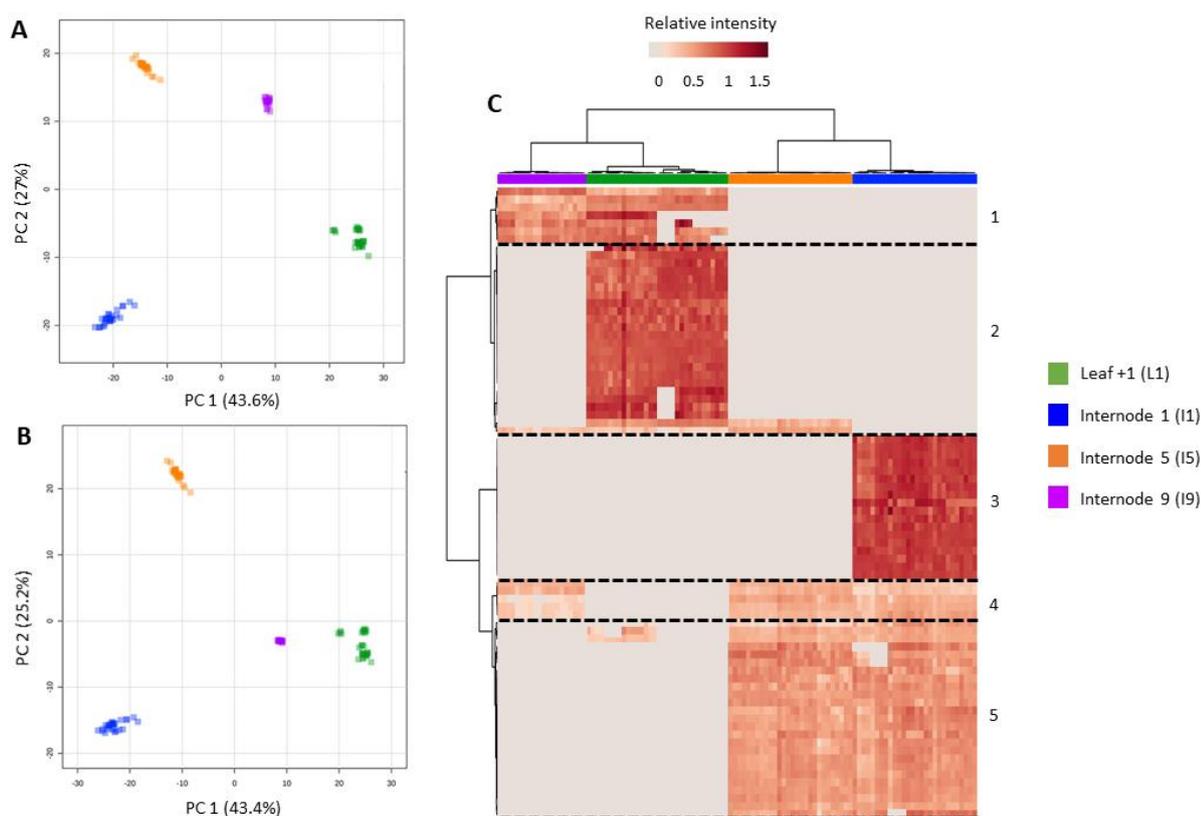
Tissue	principal component	no scale	auto scale	pareto scale
L1	PC1	60.50%	45.00%	56.30%
	PC2	14.80%	14.20%	13.90%
I1	PC1	47.50%	43.80%	43.00%
	PC2	15.80%	14.20%	13.90%
I5	PC1	37.40%	35.70%	29.00%
	PC2	24.80%	21.40%	25.00%
I9	PC1	62.30%	54.70%	58.50%
	PC2	16.50%	20.20%	18.40%

Finally, the main results demonstrated that the log10 transformation was the optimal data transformation technique for all tissues (L1, I1, I5 and I9) analysed. In addition, it was observed that no scaling technique was needed to be applied to yield optimal results for the untargeted metabolomics data, based on Shapiro Wilk's normality tests and PCAs. Using these strategies, the data was successfully prepared for data dimensionality reduction and discriminant analyses.

### 3.1.2. Data dimensionality reduction and discriminant analyses – PCA and PLS-DA models

#### 3.1.2.1. Between tissue comparisons – all tissues, fields and collection points joined for discriminant analyses

When the processed metabolomics datasets from all tissues (L1, I1, I5 and I9), collection points (C1, C2, C3 and C4) and fields (F1 and F2) were concatenated for the generation of PCA and PLS-DA models, clear separations were observed between the four tissues (Figure 7 A and B) proving that the majority of the metabolic variation can be explained regarding the differences between the different tissues (L1, I1, I5 and I9).



**Figure 7.** Discriminant models for between tissue comparisons. (A) PCA and (B) PLS-DA models generated from dry weight (DW) normalized and log10 transformed metabolomics data, combining all tissues (L1, I1, I5 and I9), collection points (C1, C2, C3 and C4) and experimental fields (F1 and F2). (C) Heatmap representation of the 79 main metabolites responsible for the separations (VIP scores  $\geq 1.0$  from PLS-DA analysis).

Following the heatmap analysis, five main clusters were observed, separating the different tissues (Figure 7 C). The first cluster consisted of 7 highly abundant tissue specific metabolites for the L1 and I9, mainly tyrosine precursors, phenylpropanoids and flavonoids (Supplementary table 6).

The second cluster consisted of 24 tissue specific metabolites for L1, mainly containing nucleic acids and nucleotides such as Uracil, 5-Thymidylic acid, Deoxyuridine monophosphate (dUMP) and Deoxyinosine monophosphate (dIMP). Additionally, a key component from the Calvin Cycle and Glycolysis, 3-Phosphoglycerate (3PG) was abundant in the L1 tissue (Supplementary table 6).

Cluster three consisted of 18 metabolites exclusively detected in the I1 tissue, mostly represented by alkaloids such as Secologanin, 4-Hydroxyphenylacetaldehyde, Hypoxanthine and 7-Methyluric acid. Purine and Zeatin precursors were also abundant in this tissue, as well as the organic osmolyte Glycine Betaine (Supplementary table 6).

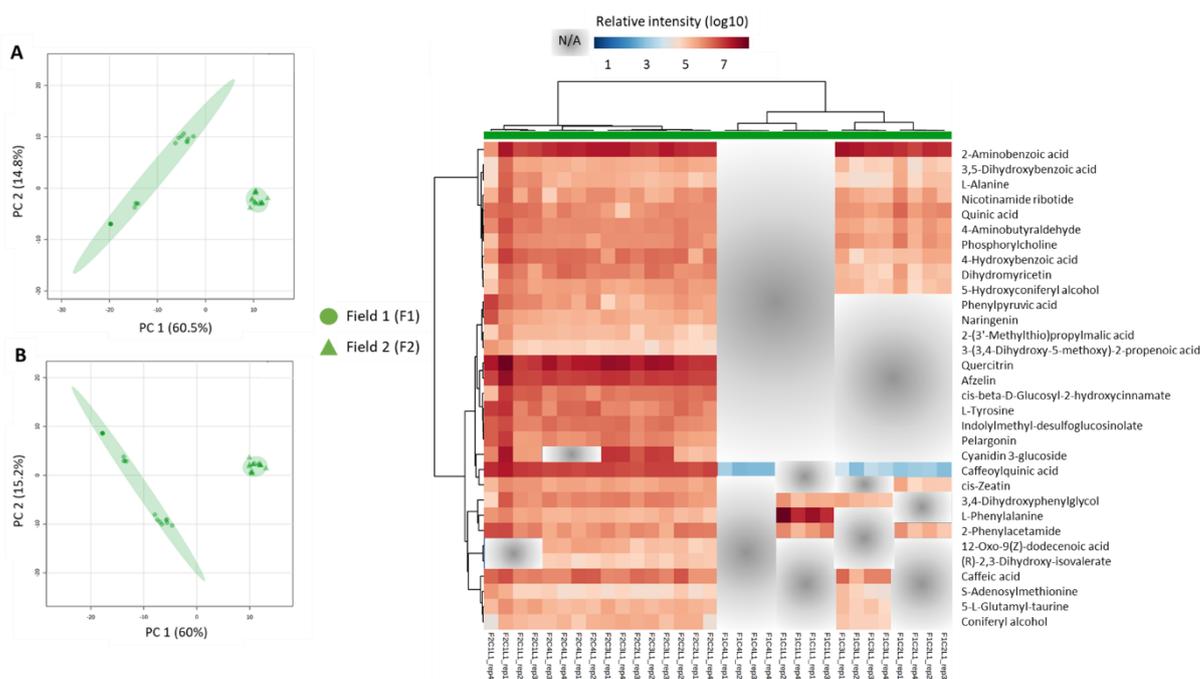
The fourth cluster was made up of 8 compounds tissue specific for the internodes (I1, I5 and I9). These compounds were mainly flavonoids and glucosinolates such as Kaempferol-3-O-glucoside, Naringenin, Cyanidin 3-glucoside, S-(4-Methylthiobutylthiohydroximoyl)-L-cysteine and Indolylmethyl-desulfoglucosinolate (Supplementary table 6).

Lastly, cluster five consisted of 22 metabolites present in high abundances in the I1 and I5 tissue. This cluster was represented by a wider variety of metabolites including amino acid and amino acid intermediates, sugars

and nucleic acids from primary metabolism, and flavonoids and phenylpropanoids from secondary metabolism such as ethylene precursors, L-proline, spermidine, *p*-coumaric acid, ferulic acid (Supplementary table 6).

### 3.1.2.2. Between field comparison – tissues separated; fields and collection points joined

When tissue specific PCA and PLS-DA models were generated for each tissue, namely the leaf +1 (L1), upper internode (I1), young internode (I5) and mature internode (I9), and when each tissue was separately analysed, they mainly reflected the differences between the two different conditions, namely fields 1 and 2 (F1 and F2) as pointed out by Figures 8-11.

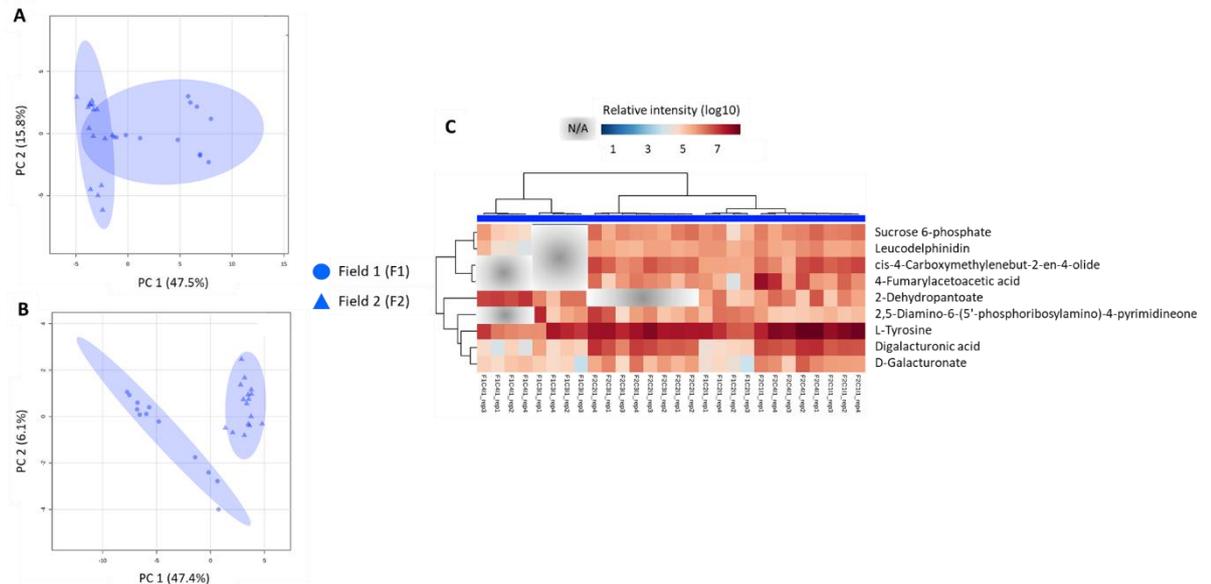


**Figure 8.** Discriminant models for between field comparisons generated from the normalized by dry weight (DW) and log10 transformed metabolomics data from the leaf +1 (L1) tissues, combining collection points (C1, C2, C3 and C4) and experimental fields (F1 and F2). (A) PCA and (B) PLS-DA models for L1; (C) Heatmap representation of the 32 main metabolites responsible for the separations (VIP scores  $\geq 1.0$  from PLS-DA analysis) and NA refers to non-detected metabolites.

For the L1 tissue, separations were observed with regards to the fields (F1 and F2) when analysing the first components (Figure 8 A and B). The metabolites with VIP scores from the PLS-DA models above 1.0 (yielding a total of 32 metabolites), were used for heatmap analysis. An overall view of the heatmap showed that most of the identified metabolites were higher in L1 from F2 as compared to F1 (Figure 8 C). These metabolites are mostly involved in phenylpropanoid biosynthesis; phenylalanine, tyrosine, and tryptophan metabolism; and flavone and flavonol biosynthesis.

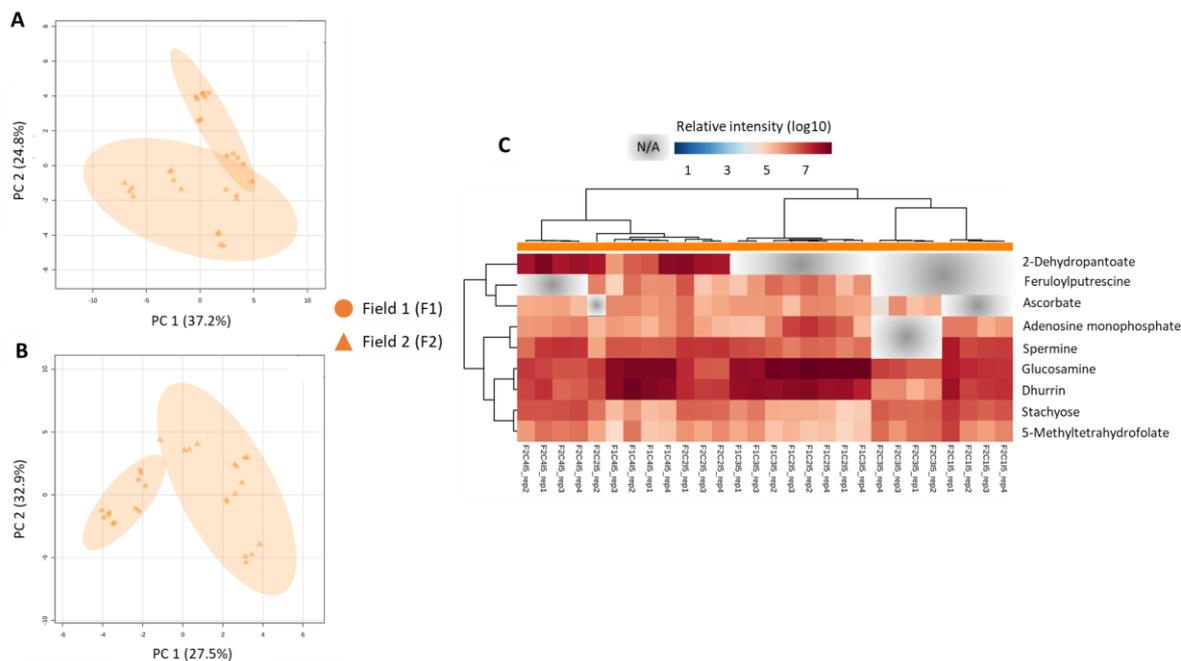
From the list of field specific metabolites in L1 from F2, that means metabolites only detected in the leaves of the plants from field 2, flavonoids such as afzelin and quercitrin were abundant (Figure 8 C). On the other hand, the metabolite cis-Zeatin was absent in all L1 samples from F1 except at collection 2 (C2) which was carried out 8 months after planting. Caffeoylquinic acid was detected in the L1 from F1 at collections 2, 3 and 4 (C2, C3 and C4), however the levels were 2-3 times lower than in the L1 from F2 (Figure 8 C; Supplementary table 2).

As shown in a cluster from the heatmap, Caffeic acid was listed and was only detected in L1 from F1 at C3 when there was a decrease in the precipitation, however still much lower as compared to the levels in the L1 from F2 (Supplementary Table 2). For the I1 tissue, the separations between the two fields were not as prominent when the PCA and PLS-DA models were inspected (Figure 9 A and B). Nine metabolites were identified based on the PCA and PLS-DA models when considering VIP scores from the PLS-DA models above 1.0. These metabolites corresponded to metabolites involved in primary metabolism, specifically in pentose and glucuronate interconversions and in tyrosine metabolism. The amino acid L-Tyrosine was highly abundant in this tissue, under all conditions. On the other hand, Digalacturonic acid D-Galacturonate were more abundant in the I1 from F2 as compare to F1 (Figure 9 C).



**Figure 9.** Discriminant models for between field comparisons generated from the normalized and log10 transformed metabolomics data from the upper internode (I1) tissues, combining collection points (C1, C2, C3 and C4) and experimental fields (F1 and F2). (A) PCA and (B) PLS-DA models for I1; (C) Heatmap representation of the 9 main metabolites responsible for the separations (VIP scores  $\geq 1.0$  from PLS-DA analysis) and NA refers to non-detected metabolites.

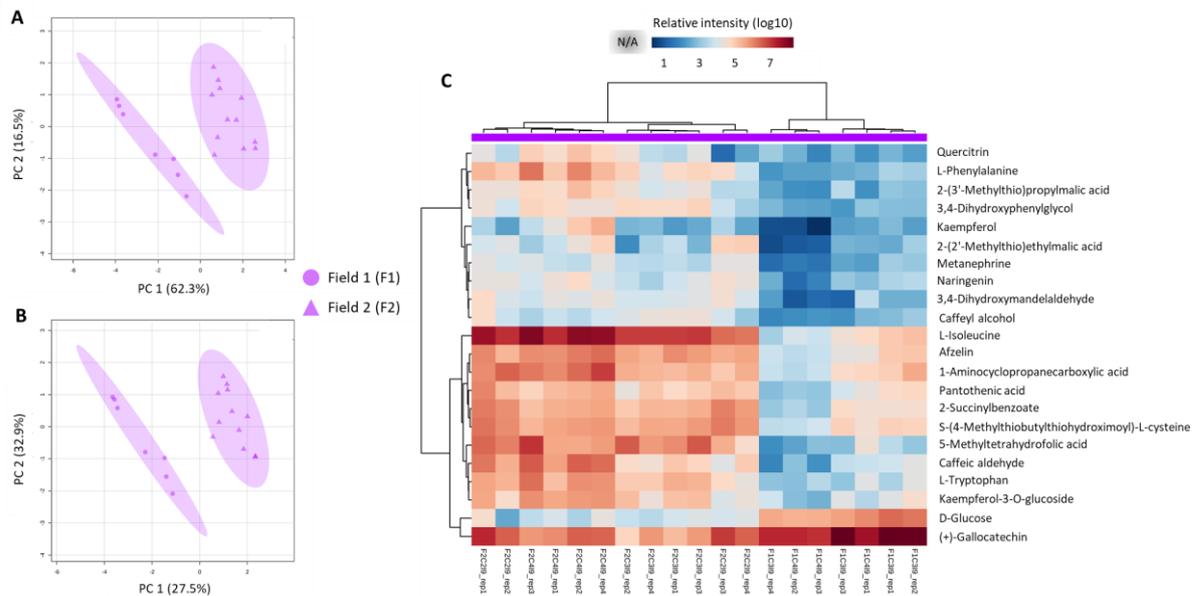
For the young internode (I5), even less of a separation could be seen between the fields (F1 and F2) as shown by the PCA and PLS-DA models in Figure 10 A and B.



**Figure 10.** Discriminant models for between field comparisons generated from the normalized and log<sub>10</sub> transformed metabolomics data from the young internode (I5) tissues, combining collection points (C1, C2, C3 and C4) and experimental fields (F1 and F2). (A) PCA and (B) PLS-DA models for I5; (C) Heatmap representation of the 9 main metabolites responsible for the separations (VIP scores  $\geq 1.0$  from PLS-DA analysis) and NA refers to non-detected metabolites.

Nine metabolites were identified with VIP scores above 1.0 from the PLS-DA model and are represented on the heatmap (Figure 10 C). As for the I1, these metabolites from I5 were again from primary metabolism such as amino acid and amino sugar metabolism, and nucleotide metabolism. Two metabolites namely Glucosamine and Dhurrin were more abundant in the I5 from F1 as compared to F2. On the other hand, Stachyose and 5-Methyltetrahydrofolate were higher in I5 from F2 as compared to F1 (Figure 10 C; Supplementary Table 4).

Lastly, in correspondence with L1, for the mature internode (I9) greater separations between the different conditions (F1 and F2) were observed (Figure 11 A and B). Regarding the metabolites with a VIP score above 1.0 from the PLS-DA model, 22 were identified. These metabolites represent pathways such as flavone and flavonol biosynthesis; phenylpropanoid biosynthesis; flavonoid biosynthesis; and phenylalanine, tyrosine and tryptophan biosynthesis. The metabolite (+)-Gallocatechin was abundant in all the samples of the I9 tissue (Figure 11 C; Supplementary table 5).

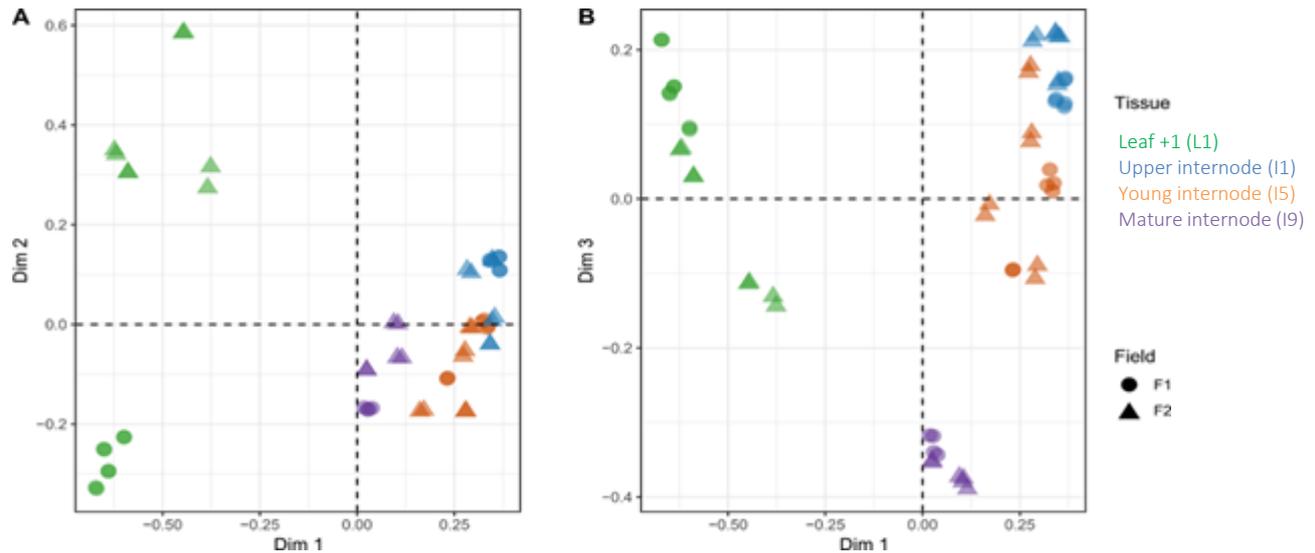


**Figure 11.** Discriminant models for between field comparisons generated from the normalized and log<sub>10</sub> transformed metabolomics data from the mature internode (I9) tissues, combining collection points (C1, C2, C3 and C4) and experimental fields (F1 and F2). (A) PCA and (B) PLS-DA models for I9; (C) Heatmap representation of the 22 main metabolites responsible for the separations (VIP scores  $\geq 1.0$  from PLS-DA analysis) and NA refers to non-detected metabolites.

Most of these metabolites were relatively higher in the I9 from F2 as compared to F1. In contrast, the D-Glucose metabolite, the only sugar out of the 22 metabolites, were higher in the I9 from F1 as compared to F2 (Figure 11 C; Supplementary table 5).

### 3.2. Transcriptomics analyses of four anatomically different tissues from SP80-3280

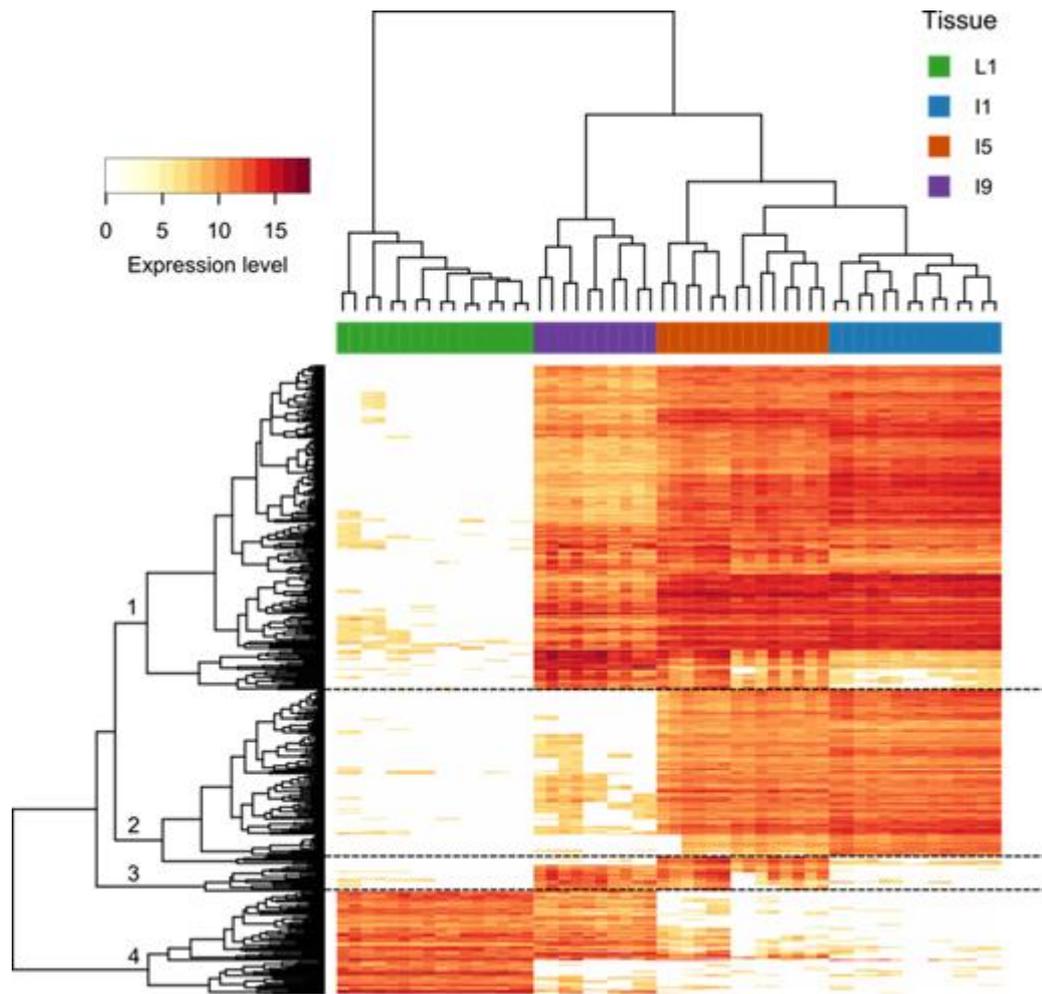
Out of the 21,902 different probes represented on the CaneRegNet oligoarray, 15,552 (13,112 sense and 2,440 antisense) were significantly expressed regarding all tissues, collection points and field conditions. Multidimensional scaling (MDS) analyses using these transcripts evidenced a distinct expression pattern between the different tissues (L1, I1, I5 and I9), represented mainly by the first dimension (Figure 12 A).



**Figure 12.** Multidimensional scaling (MDS) analyses of the transcriptomics data. (A) Dimension 1 (Dim1) vs Dimension 2 (Dim2); (B) Dim1 vs Dimension 3 (Dim3). L1, I1, I5 and I9 refer to leaf +1 and upper, young, and mature internodes. Field 1 (F1) and 2 (F2) refer to the plants that were planted in April 2012 and October 2012, respectively.

This pattern corresponded to both the PCA and PLS-DA models generated from the metabolomics data (Figure 7 A and B), and thus highlighting the fact that the most of the variation within the dataset is attributed to the metabolic differences between the distinct anatomical tissues. However, the second dimension allowed for the distinction between the different experimental conditions (F1 and F2) in the L1 and I9 samples, respectively. In addition, the third dimension contributed to the clear distinction between I9 from the remainder of the tissues (Figure 12 B).

Heatmap analysis of the 500 most variable genes within the dataset allowed for, in addition to the clear distinction between the tissues, the identification of four specific expression patterns (Figure 13) regardless of the collection point, highlighting yet again that the main differences exist due to the different anatomical tissues and their specified metabolism.



**Figure 13.** Heatmap constructed using top 500 most variable genes responsible for the variations in the different tissues (L1, I1, I5 and I9) from a transcriptomics point of view.

The first cluster included 258 genes of which the expression was mostly absent in L1. This gene set was enriched for genes with the GO terms associated with carbohydrate metabolism, specifically cell wall organization and construction (Table 6).

**Table 6.** The Gene Ontology (GO) IDs and terms represented by the genes from each of the clusters identified by the heatmap analysis.

Cluster	GO ID	Term	p-value
1	GO:0071554	cell wall organization or biogenesis	2.0E-07
	GO:0005985	sucrose metabolic process	1.5E-05
	GO:0005984	disaccharide metabolic process	2.6E-05
	GO:0009311	oligosaccharide metabolic process	3.6E-05
	GO:0044264	cellular polysaccharide metabolic process	3.9E-05
	GO:0005982	starch metabolic process	3.9E-05
	GO:0071555	cell wall organization	5.4E-05
	GO:0005976	polysaccharide metabolic process	5.7E-05
	GO:0045229	external encapsulating structure organization	8.4E-05

	GO:0044262	cellular carbohydrate metabolic process	1.5E-04
2	GO:0051276	chromosome organization	7.1E-09
	GO:0006270	DNA replication initiation	4.7E-06
	GO:0006996	organelle organization	1.2E-05
	GO:0006261	DNA-dependent DNA replication	1.3E-05
	GO:0016043	cellular component organization	1.8E-05
	GO:0006260	DNA replication	4.6E-05
	GO:0032392	DNA geometric change	7.4E-05
	GO:0032508	DNA duplex unwinding	7.4E-05
	GO:0007059	chromosome segregation	1.0E-04
	GO:0007049	cell cycle	2.7E-04
3	GO:1903338	regulation of cell wall organization	1.8E-03
	GO:2000652	regulation of secondary cell wall biogenesis	1.8E-03
	GO:0010089	xylem development	2.7E-03
	GO:0046271	phenylpropanoid catabolic process	2.7E-03
	GO:0046274	lignin catabolic process	2.7E-03
	GO:0010087	phloem or xylem histogenesis	7.1E-03
	GO:0009834	plant-type secondary cell wall biogenesis	8.0E-03
	GO:0009809	lignin biosynthetic process	1.5E-02
	GO:0009808	lignin metabolic process	1.6E-02
	GO:0000209	protein polyubiquitination	1.9E-02
4	GO:0019748	secondary metabolic process	8.7E-04
	GO:0098754	Detoxification	1.1E-03
	GO:0009404	toxin metabolic process	1.4E-03
	GO:0009407	toxin catabolic process	1.4E-03
	GO:0009636	response to toxic substance	1.6E-03
	GO:0006001	fructose catabolic process	3.6E-03
	GO:0015947	methane metabolic process	7.1E-03
	GO:0043446	cellular alkane metabolic process	7.1E-03
	GO:0043436	oxoacid metabolic process	9.5E-03
	GO:0006082	organic acid metabolic process	9.5E-03

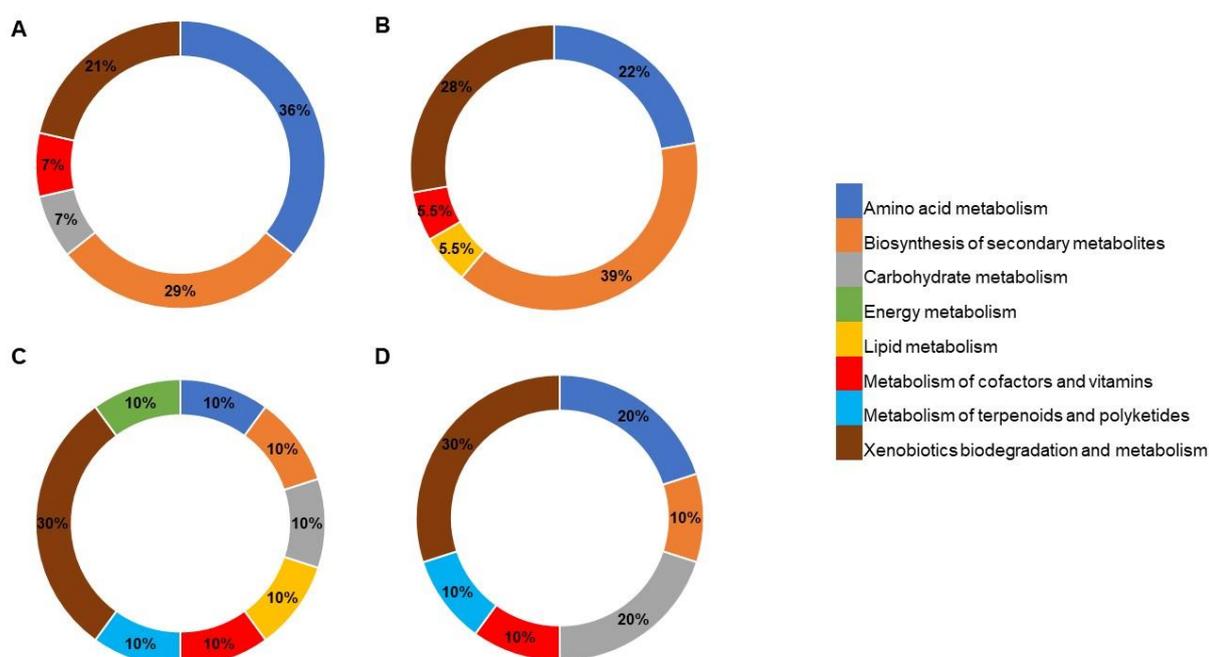
The second cluster was composed of 139 genes whose expression were higher in I1 and I5, and were enriched for GO terms related to the cell cycle, including chromosome organization and DNA replication (Table 6)

The third cluster presented 21 genes most expressed in I5 and I9 and were enriched for GO terms related to the development of conducting vessels (xylem and phloem) and lignin metabolism (Table 6).

Finally, the fourth cluster included 82 genes whose expression were higher in L1 and the top GO term pointed towards to secondary metabolism (Table 6).

### 3.3. MOFA – the tool leading the unsupervised integration between two data modalities

Based on the combination between the metabolomics and transcriptomics data via MOFA (Argelaguet et al., 2018), it was possible to point out the main metabolic pathways containing both the identified metabolites and the transcripts of the genes coding for the enzymes directly upstream or downstream of these specific metabolites. For the L1 tissue, it was found that the main metabolic pathways, based on the KEGG BRITE classification (<http://www.genome.jp/kegg/>), that were activated and altered throughout development were amino acid metabolism (36%), biosynthesis of secondary metabolites (29%), xenobiotics biodegradation and metabolism (21%), carbohydrate metabolism (7%) and metabolism of cofactors and vitamins (7%) as shown in Figure 14 A.



**Figure 14.** Doughnut graphs representing the main KEGG metabolic pathways identified to be altered throughout development in the (A) leaf +1 (L1); (B) upper (I1); (C) young (I5) and mature (I9) internodal tissues.

Most of the GO terms generated from the transcriptomics data from the MOFA model for the L1 tissue, pointed towards amino acid metabolism and thus complementing the findings reported from the metabolomics data, as shown in Table 7. Interestingly, the metabolomics data added one crucial extra layer of information since it was able to pinpoint exactly which amino acid and small molecule metabolic pathways are altered in the L1 tissue (and also in the other tissues as described below) throughout development, namely phenylalanine, tyrosine and tryptophan metabolism; and cyanoamino acid metabolism.

**Table 7.** An integrative view of the combined output results from MOFA for the leaf +1 (L1), upper (I1), young (I5) and mature internodal (I9) tissues. For the transcriptomics data modality, the GO IDs and GO terms are listed. For the metabolomics data modality, the KEGG map IDs and KEGG map names are listed.

Tissue	Transcriptomics data modality		Metabolomics data modality	
	GO ID	GO term	KEGG map ID	KEGG map name
L1	GO:0044281	small molecule metabolic process	map00940	Phenylpropanoid biosynthesis
	GO:0006520	cellular amino acid metabolic process	map00941	Flavonoid biosynthesis
	GO:0043436	oxoacid metabolic process	map00360	Phenylalanine metabolism
	GO:0006082	organic acid metabolic process	map00640	Cyanoamino acid metabolism
	GO:0019752	carboxylic acid metabolic process	map00966	Glucosinolate biosynthesis
	GO:1901566	organonitrogen compound biosynthetic process	map00030	Pentose phosphate pathway
	GO:0008652	cellular amino acid biosynthetic process	map00400	Phenylalanine, tyrosine and tryptophan metabolism
	GO:1901605	alpha-amino acid metabolic process	map00380	Tryptophan metabolism
	GO:0044283	small molecule biosynthetic process	map00362	Benzoate degradation
	GO:1901607	alpha-amino acid biosynthetic process	map00960	Tropane, piperidine and pyridine alkaloid biosynthesis
	GO:0016311	Dephosphorylation	map00350	Tyrosine metabolism
	GO:0051186	cofactor metabolic process	map00623	Toluene degradation
	GO:0017144	drug metabolic process	map00627	Aminobenzoate degradation
	GO:0055086	nucleobase-containing small molecule metabolism	map00130	Ubiquinone and other terpenoid-quinone biosynthesis
I1	GO:0010035	response to inorganic substance	map00943	Isoflavonoid biosynthesis
	GO:0010077	maintenance of inflorescence meristem identity	map00966	Glucosinolate biosynthesis
	GO:0009414	response to water deprivation	map00944	Flavone and flavonol biosynthesis
	GO:0006952	defense response	map00680	Methane metabolism

	GO:0009415	response to water	map00360	Phenylalanine metabolism
	GO:0005980	glycogen catabolic process	map00940	Phenylpropanoid biosynthesis
	GO:0046398	UDP-glucuronate metabolic process	map00998	Biosynthesis of secondary metabolites - unclassified
	GO:0050832	defense response to fungus	map00130	Ubiquinone and other terpenoid-quinone biosynthesis
	GO:0006950	response to stress	map00980	Metabolism of xenobiotics by cytochrome P450
	GO:0045944	positive regulation of transcription by RNA polymerase II	map00524	Neomycin, kanamycin and gentamicin biosynthesis
	GO:0006457	protein folding	map00950	Isoquinoline alkaloid biosynthesis
	GO:0061077	chaperone-mediated protein folding	map00350	Tyrosine metabolism
	GO:0006298	mismatch repair	map00564	Glycerophospholipid metabolism
	GO:0050896	response to stimulus	map00965	Betalain biosynthesis
	GO:0006950	DNA duplex unwinding	map00261	Monobactam biosynthesis
	GO:0032508	multi-organism process	map00410	beta-Alanine metabolism
	GO:0051704	response to fungus	map01055	Biosynthesis of vancomycin group antibiotics
	GO:0009620	DNA geometric change	map00480	Glutathione metabolism
I5	GO:0005975	carbohydrate metabolic process	map00908	Zeatin biosynthesis
	GO:0051186	cofactor metabolic process	map00954	Stilbenoid, diarylheptanoid and gingerol biosynthesis
	GO:0017144	drug metabolic process	map00564	Glycerophospholipid metabolism
	GO:0042737	drug catabolic process	map00230	Purine metabolism
	GO:0098754	Detoxification	map00640	Propanoate metabolism

	GO:0009636	response to toxic substance	map00770	Pantothenate and CoA biosynthesis
	GO:0045229	external encapsulating structure organization	map00626	Naphthalene degradation
	GO:0044281	small molecule metabolic process	map00361	Chlorocyclohexane and chlorobenzene degradation
	GO:0055086	nucleobase-containing small molecule metabolism	map00350	Tyrosine metabolism
	GO:0098869	cellular oxidant detoxification	map00643	Styrene degradation
I9	GO:0006470	protein dephosphorylation	map00860	Porphyrin and chlorophyll metabolism
	GO:0009072	aromatic amino acid family metabolic process	map00630	Glyoxylate and dicarboxylate metabolism
	GO:0016311	dephosphorylation	map00998	Biosynthesis of secondary metabolites - unclassified
	GO:0005975	carbohydrate metabolic process	map00040	Pentose and glucuronate interconversions
	GO:0009073	aromatic amino acid family biosynthetic process	map00680	Methane metabolism
	GO:0016053	organic acid biosynthetic process	map00380	Tryptophan metabolism
	GO:0046394	carboxylic acid biosynthetic process	map00523	Polyketide sugar unit biosynthesis
	GO:0044281	small molecule metabolic process	map00340	Histidine metabolism
	GO:0017144	drug metabolic process	map00261	Monobactam biosynthesis
	GO:0006520	cellular amino acid metabolic process	map00983	Drug metabolism - other enzymes

The secondary metabolic pathways that were affected were the phenylpropanoid biosynthesis; flavonoid biosynthesis; glucosinolate biosynthesis; and tropane, piperidine and pyridine alkaloid biosynthesis (Table 7). However, from a transcriptomics point of view, no exact and direct GO terms regarding secondary metabolism were identified. In both metabolomics and transcriptomics data modalities, cofactor metabolism was identified. Yet again, the metabolomics data indicated that it specifically referred to ubiquinone and other terpenoid-quinone biosynthesis under the KEGG BRITE classification “metabolism of cofactors and vitamins”. The only carbohydrate metabolism pathway that was sufficiently activated and altered to be picked up by and included in the MOFA model generated for the L1

tissue, and that also shared corresponding variance with the transcriptomics data, was the pentose phosphate pathway (PPP) as can be seen in Table 7. The presence of the benzoate degradation, aminobenzoate degradation, and toluene degradation metabolic pathways identified in the metabolomics data modality represented the non-plant pathways from the main KEGG BRITE classification “xenobiotics biodegradation and metabolism”. Interestingly, this phenomenon was also supported by the transcriptomics data in which “drug metabolic process” was one of the GO terms that was identified (Table 7).

In I1, 39% of the altered pathways represented secondary metabolite biosynthesis (Figure 14 B) such as isoflavonoid biosynthesis; glucosinolate biosynthesis; flavone and flavonol biosynthesis; and phenylpropanoid biosynthesis; isoquinoline alkaloid biosynthesis; and betalain biosynthesis (Table 7), all known to be involved in plant stress response and in the synthesis of lignin precursors. From the 18 GO terms that were highlighted in the MOFA model for I1, 9 indicated that the plants were in fact responding to internal and/or external stressors (Table 7). Amino acid metabolism represented 22% of the altered pathways from the metabolomics data modality, specifically phenylalanine metabolism; tyrosine metabolism; beta-alanine metabolism; and glutathione metabolism. However, no directly linked GO terms were identified regarding these pathways. Yet again, non-plant metabolic pathways representing 28% of the altered pathways were identified in the I1 tissue, including methane metabolism; metabolism of xenobiotics by cytochrome P450; neomycin, kanamycin and gentamicin biosynthesis; monobactam biosynthesis and biosynthesis of vancomycin group antibiotics. The presence of GO terms from the transcriptomics data modality related to fungal and inorganic substance responses, as well as multi-organism processes support the metabolomics data that indicated that non-plant organisms were in the plants’ proximity, either in their internal or external environment. The remaining metabolic pathways included lipid metabolism and metabolism of cofactors and vitamins, each representing 5.5% of the altered pathways. On the other hand for the transcriptomics data modality, GO terms related to the cell cycle were identified (Table 7).

From the activated pathways identified by the MOFA models for I5, all except for one pathway, equally represented the same percentage (10% respectively) of altered pathways when considering the total altered pathways (Figure 14 C). At the top of the list for altered pathways with regards to the metabolomics data, zeatin biosynthesis is listed (Table 7). When analysing the transcriptomics data, no direct links could be made with regards to zeatin biosynthesis, however the GO term nucleobase-containing small molecule metabolic process was identified (Table 7), which is most likely due to the identification of the purine metabolism, which is a metabolic pathway that was in fact highlighted in the metabolomics data modality by MOFA. Since adenine is a purine and a zeatin precursor, the possibility exists that that these pathways go hand-in-hand and are altered during development in the I5 tissue. The metabolic pathways supporting the GO terms carbohydrate metabolic process and cofactor metabolic process were highlighted as propanoate metabolism and pantothenate and CoA biosynthesis, respectively (Table 7). Tyrosine metabolism was one of the altered amino acid metabolic pathways. Yet again, activated pathways not naturally found in plants were highlighted, representing 30% of the total altered pathways in the I5 tissue. This finding was in fact corroborated by both the metabolomics and transcriptomics data seeing that metabolic pathways such as naphthalene degradation; chlorocyclohexane and chlorobenzene degradation; and styrene degradation were identified, and GO terms related to drug metabolism, detoxification processes and responses to toxic substances were highlighted from the transcriptomics data (Table 7).

Lastly, for I9, the amino acid metabolism and carbohydrate metabolism represented 20% of the altered pathways, respectively (Figure 14 D). The majority of the gene expression data for this tissue pointed towards alterations in aromatic amino acid metabolism, specifically tryptophan metabolism as pinpointed by the metabolomics

results (Table 7). The glyoxylate and dicarboxylate metabolism pathway was identified as being altered in I9, which provides intermediates for carbohydrate biosynthesis, most likely for the pentose and glucuronate metabolism pathway since this pathway was also identified in the metabolomics data modality. These observations can be supported by the carbohydrate metabolic process GO term identified in the transcriptomics data modality (Table 7). Non-plant pathways were detected again (Figure 14 D), in conjunction with the GO term drug metabolic process (Table 7), thus indicating the presence of microorganisms in the plants' internal or external environment, as were the cases for all the other tissues.

### 3.4. Narrowing down the metabolic pathways and the selection of target genes within these pathways

Based on the metabolomics and transcriptomics data obtained via HPLC-MS and DNA microarrays; the MOFA models that were generated for each tissue; the identification of shared variance between metabolomics and transcriptomics data modalities; and extensive literature reviews, specific pathways were chosen for further investigation, namely the pentose phosphate pathway (map00030), phenylalanine, tyrosine and tryptophan biosynthesis (map00400), phenylpropanoid biosynthesis (map00940) and flavonoid biosynthesis (map00941) (Supplementary Figures 2 - 5). Since it was only possible to differentiate between the two conditions of planting (“one-and-a-half-year” and “one-year” sugarcane) in the L1 (Figure 8 A and B) and I9 tissues (Figure 11 G and H), these tissues were sub-selected for the further exploration of metabolic pathway activity changes of the selected pathways throughout development, and in response to alterations in the external environments in each field. The PAPI tool was used for the generation of the metabolic pathway activity scores at each collection point, which indicates if a metabolic pathway is active under certain conditions and allows us to compare different metabolic pathway activities between different conditions (Aggio et al., 2010). The exact mode of action and steps of this tool is described in Figure 4.

The metabolic intermediates in these pathways and the SAS (EST contigs) directly related to the metabolites (via the specific enzymes directly upstream or downstream of these metabolites on the KEGG database) were identified. For the PPP, which contains 15 enzymes in the pathway, 83 SAS were found to be related to at least one enzyme (Table 8).

**Table 8.** The four main selected pathways from the KEGG database, the pathway map ID, the number of enzymes within the particular pathway and the number of SAS in the particular pathway.

Pathway name	Pathway ID	#Enzymes in Pathway	#SAS in Pathway
Pentose phosphate pathway	map00030	15	83
Phenylalanine, tyrosine and tryptophan biosynthesis	map00400	24	107
Phenylpropanoid biosynthesis	map00940	11	147
Flavonoid biosynthesis	map00941	7	15

For the phenylalanine, tyrosine and tryptophan biosynthesis pathway, which contains 24 enzymes in the pathway, 107 SAS were found to be related to at least one enzyme. For the phenylpropanoid biosynthesis pathway, which contains 11 enzymes in the pathway, 147 SAS were found to be related to at least one enzyme. For the flavonoid biosynthesis pathway, which contains 7 enzymes in the pathway, 15 SAS were found to be related to at least one enzyme.

Using these SAS and their locations on the selected KEGG maps with regards to the specific metabolites that were identified, 9 SAS were identified as target genes for studying sugarcane development in the field, in a source-sink context (Table 9).

**Table 9.** The nine narrowed down SAS transcripts from the four selected pathways from the KEGG database and their respective enzyme codes and enzyme names.

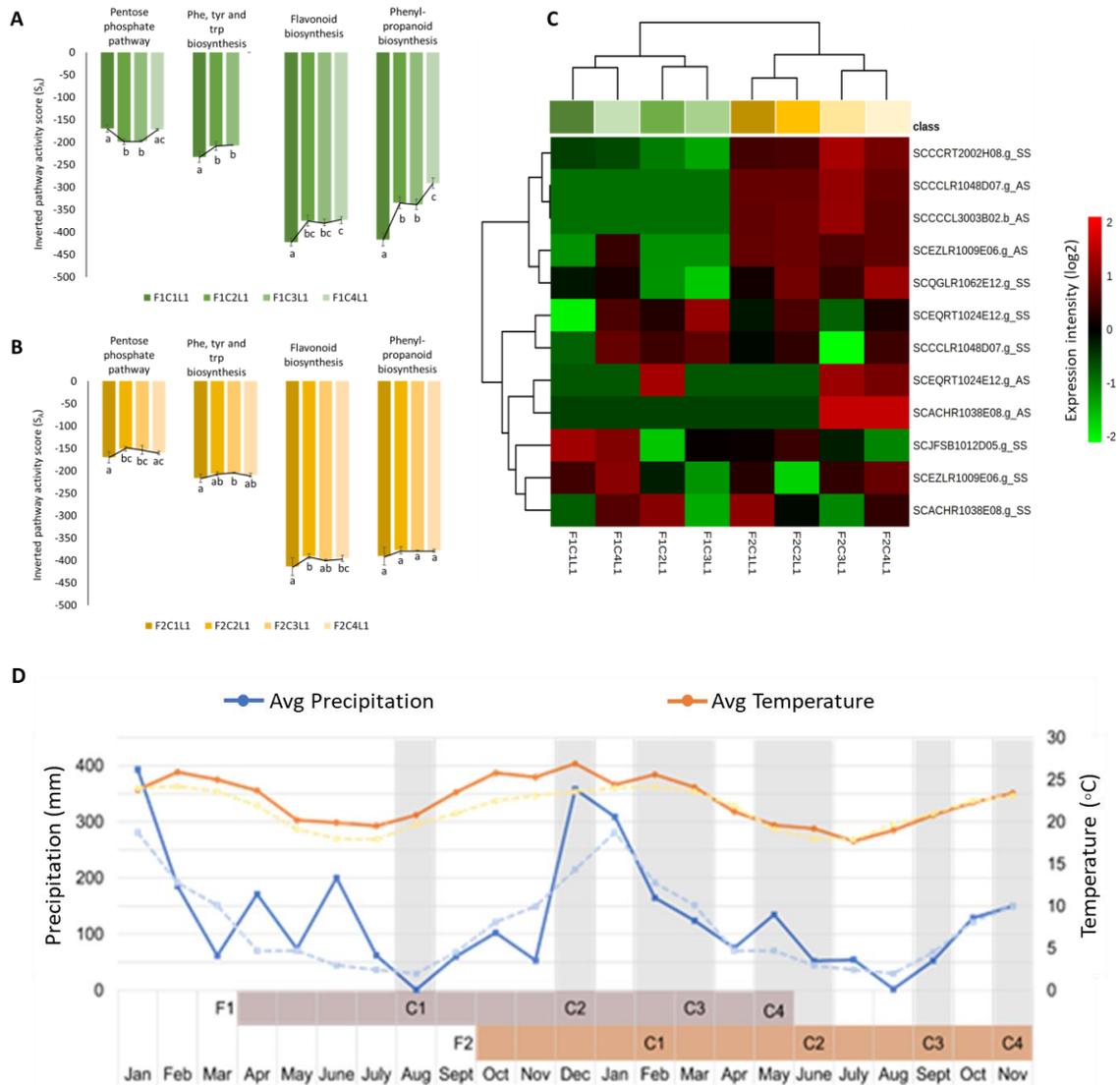
SAS name	Pathway name	Enzyme code (EC)	BLAST2GO annotation	BLASTX annotation
SCEZLR1009E06.g	Flavonoid biosynthesis	ec:1.14.11.9	flavanone 3-dioxygenase (F3H)	1-aminocyclopropane-1-carboxylate oxidase 1 (ACC) [ <i>Sorghum bicolor</i> ]
SCACHR1038E08.g	Phenylpropanoid biosynthesis	ec:1.1.1.195	cinnamyl-alcohol dehydrogenase (CAD)	probable cinnamyl alcohol dehydrogenase 6 (CAD) [ <i>Sorghum bicolor</i> ]
SCCCCL3003B02.b	Phenylpropanoid biosynthesis	ec:1.11.1.7	peroxidase (PER)	cationic peroxidase SPC4-like (PER) [ <i>Sorghum bicolor</i> ]
SCCCLR1048D07.g	Phenylpropanoid biosynthesis	ec:4.3.1.24; ec:4.3.1.25	phenylalanine ammonia-lyase (PAL); phenylalanine/tyrosine ammonia-lyase (PAL)	phenylalanine ammonia-lyase (PAL) [ <i>Saccharum</i> hybrid cultivar ROC22]
SCCART2002H08.g	Phenylpropanoid biosynthesis	ec:1.11.1.7	peroxidase (PER)	probable L-ascorbate peroxidase 3 (PER) [ <i>Sorghum bicolor</i> ]
SCEQRT1024E12.g	Phenylpropanoid biosynthesis	ec:4.3.1.24; ec:4.3.1.25	phenylalanine ammonia-lyase (PAL); phenylalanine/tyrosine ammonia-lyase (PAL)	phenylalanine ammonia-lyase (PAL) [ <i>Zea mays</i> ]
SCJFSB1012D05.g	Phenylpropanoid biosynthesis	ec:1.11.1.7	peroxidase (PER)	putative L-ascorbate peroxidase 6 (PER) [ <i>Saccharum</i> hybrid cultivar ROC22]
SCQGLR1062E12.g	Phenylalanine, tyrosine and tryptophan biosynthesis	ec:4.2.1.51; ec:4.2.1.91	prephenate dehydratase; arogenate dehydratase (ADT)	glyoxylase 1 [ <i>Saccharum</i> hybrid cultivar]
SCSGAM1094D05.g	Phenylpropanoid biosynthesis	ec:4.3.1.24; ec:4.3.1.25	phenylalanine ammonia-lyase (PAL); phenylalanine/tyrosine ammonia-lyase (PAL)	phenylalanine ammonia-lyase (PAL) [ <i>Sorghum bicolor</i> ]

From all these transcripts, both sense strands (SS) and antisense strands (AS) of transcripts were considered. All except two transcripts (SCEZLR1009E06.g and SCQGLR1062E12.g) presented corresponding results from both BLAST2GO and BLASTX annotations (Table 9). These different annotations could be attributed to the facts that

different enzymes, for example *ACC* and *F3H* share the same protein domains, and the probes from the microarrays covered the protein domains that were shared between these represented genes.

With regards to the pathway activity profiling, a score could in fact be attributed to each of the selected pathways in the L1 and I9 tissues, respectively. This score represents the level of activity where the higher the score the lower the activity (Aggio et al., 2010). Thus, to avoid confusion and to facilitate visualization, these scores were inverted as represented in Figures 17 and 18.

For the L1 from F1, which experienced a period of water limitation during the initial vegetative growth phase, focussing on when no culm had been formed at C1, the pathway activities of the phenylalanine, tyrosine and tryptophan biosynthesis; flavonoid biosynthesis; and phenylpropanoid biosynthesis were significantly lower (based on ANOVA and Fisher's LSD,  $p < 0.05$ ) as compared to when thick culm tissue had been formed at C2, and the vegetative growth and elongation phases were being carried out at C3 and C4 from F1 (Figure 15 A).



**Figure 15.** Pathway activities based on the Pathway Activity Profiling (PAPi) tool for leaf +1 (L1) from **(A)** F1 and **(B)** F2. C1, C2, C3, and C4 refer to collection points 1, 2, 3 and 4 specifically 4, 8, 11 and 13 months after planting. Different letters under the bars indicate statistically significant differences between the different collection points based on ANOVA and Fisher's LSD ( $p < 0.05$ ). **(C)** Heatmap analysis for the statistically significant expression of the selected sugarcane assembled sequences (SAS) transcripts shown as "classes". When a transcript was not expressed, it is coloured as black in the heatmap analysis. **(D)** Climatic conditions for the field experiments. Fields 1 (F1) and 2 (F2) refer to the plants that were planted in April 2012 and October 2012, respectively. C1, C2, C3, and C4 refer to collection points 1, 2, 3 and 4 specifically 4, 8, 11 and 13 months after planting.

The phenylpropanoid and flavonoid biosynthetic pathways were significantly higher at C4 as compared to C1, however, the phenylalanine, tyrosine and tryptophan biosynthesis pathway was not even detected (Figure 15 A), thus indicating that the flux through this pathway was very low or that it was suppressed with a significant drop in the recorded temperatures of approximately  $8^{\circ}\text{C}$ . In contrast, the PPP was equally and highly active at C1 and C4 (Figure 15 A). Thus, when no culms were formed, the activity of the phenylpropanoid biosynthesis; phenylalanine, tyrosine and tryptophan biosynthesis; and flavonoid biosynthesis pathways in the leaves were not affected by low precipitation and temperatures, however the PPP always responded to these changes in precipitation and temperatures in by increasing the pathway activity.

Most transcripts from these pathways, directly upstream or downstream of the identified metabolites in those pathways, presented low expression levels or were entirely absent at C1 except for 2 transcripts representing

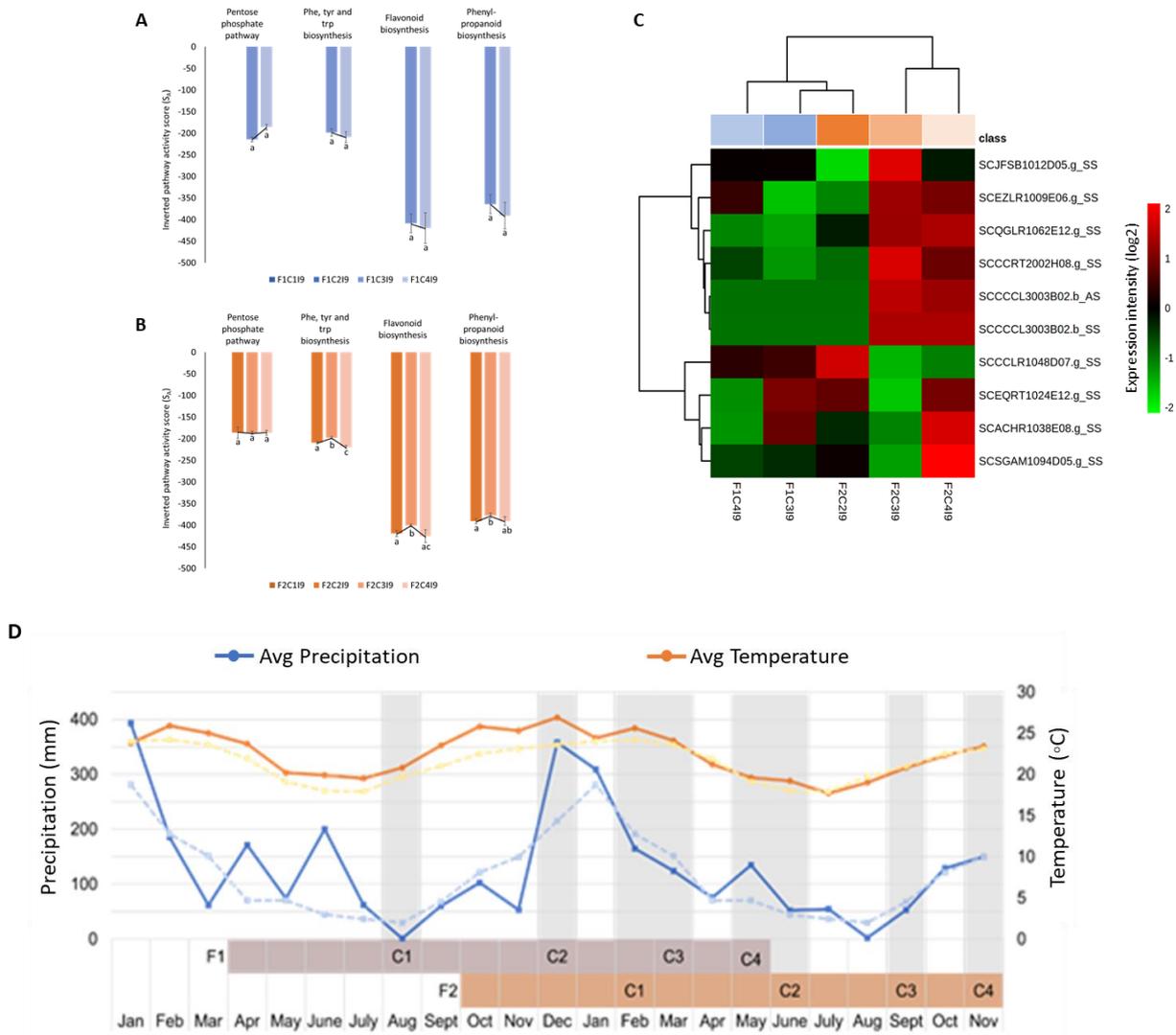
*PER* and *F3H* that presented high levels of expression (Figure 15 C). This possibly implicates these genes in the regulation of the mentioned pathways in the leaf tissue during SP80-3280 development and culm formation. The PPP and phenylpropanoid biosynthesis pathways presented increased activity as the plants continued with the vegetative growth and elongation phases (Figure 15 A). Moreover, 2 *PAL* transcripts from maize and sugarcane were highly expressed in the L1 when there was a dip in the precipitation (Figure 15 D). It is thus speculated that when water availability is low during the periods when it is crucial, that is during the initial steps of the vegetative growth phase, these pathways will be activated and the fluxes will be increased to produce all the metabolic intermediates and final products for the formation of more biomass in the form of thicker culms.

Considering the L1 from F2, the PPP and the flavonoid biosynthesis pathways significantly increased (based on ANOVA and Fisher's LSD,  $p < 0.05$ ) when the plants were experiencing rapid vegetative growth and when culm tissues were already formed (Figure 15 B). When the plants reached the maturation phase, the activities of all 4 pathways stabilized. However, significant differences were observed between C1 and C3 for the PPP and the phenylalanine, tyrosine and tryptophan biosynthesis pathways, where the pathway activities were significantly higher when lower temperature and precipitation were recorded. At C4, when flowering occurred spontaneously, the activity of the flavonoid biosynthesis pathway was significantly higher as compared to when the plants were in the vegetative growth phase at C1.

Focussing on the heatmap analyses for the leaves from F1, C1 and C4 clustered together and C2 and C3 in another group. On the other hand, for the leaves from F2, C1 and C2 clustered together and C3 and C4, respectively (Figure 15 C). The expression of most AS strands of the SAS transcripts were absent in the L1 tissue from F1, except for the AS strand of the SCEQRT1024E12.g (*PAL*) and SCEZLR1009E06.g (*F3H*). The AS of the SCEZLR1009E06.g (*F3H*) was only significantly expressed in the L1 at C4 and the SS was highly expressed at both C1 and C4 where the environmental conditions were quite similar, that is lower precipitation and temperatures. The AS of the SCEQRT1024E12.g (*PAL*) transcript was only significantly expressed at C2 when the sugarcane plants presented culm tissue for the first time. Interestingly, in the L1 from F2, the expression of all AS from the transcripts were present, except for SCEQRT1024E12.g (*PAL*) and SCACHR1038E08.g (*CAD*), which were both only significantly expressed at C3 and C4 when the plants had already reached maturity due to the drop in the precipitation and temperatures measurements.

Shifting the focus towards only the SSs, the two significantly expressed SSs of the *PAL* transcripts (SCEQRT1024E12.g and SCCCLR1048D07.g) followed the same expression pattern in the L1 of F1. The expression of both significantly increased from C1 to C2, and only the SCEQRT1024E12.g presented significant increases in its expression between C2 and C3. At C3 and C4, the expression of these transcripts was significantly higher as compared to C1 when no internodal tissues were formed. As for the L1 from F2, the expression of these *PAL* transcripts significantly increased between C2 and C3 when the maturation and the lignification of the sugarcane plants were taking place, and also between C3 and C4 when the flowering process was initiated. For the SCACHR1038E08.g (*CAD*) transcript, it significantly increased in the L1 from F1 between C1 and C2 when the sugarcane plants were starting to grow vigorously during the vegetative growth phase and when the thick culm tissue was formed. As for the L1 from F2, the expression of this transcript also increased significantly from C2 to C3 and from C3 to C4 respectively, thus implicating it in the lignification and flowering processes as well. For the SCCCRT2002H08.g (*PER*) and SCJFSB1012D05.g (*PER*) transcripts, the only notable results were obtained for the L1 tissue from F1 since the expression profiles of these transcripts presented contrasting results in the L1 from F2. The expression of these transcripts was significantly higher at C1 as compared to C3.

Regarding the pathway activity analysis of the I9 of the plants in the vegetative growth and elongation phase (C3 and C4 of F1), no significant (based on Student's t-test,  $p$ -value  $<0.05$ ) alterations were observed (Figure 16 A). On the other hand, the pathway activities of the phenylalanine, tyrosine and tryptophan biosynthesis; flavonoid biosynthesis; and phenylpropanoid biosynthesis were significantly higher in the I9 when SP80-3280 experienced water deficit (Figure 16 B).



**Figure 16.** Pathway activities based on the Pathway Activity Profiling (PAPi) tool for mature internode (I9) from (A) F1 (April 2012 plantation) and (B) F2 (October 2012 plantation). C1, C2, C3, and C4 refer to collection points 1, 2, 3 and 4 specifically 4, 8, 11 and 13 months after planting. Different letters indicate statistically significant differences between the different collection points based on ANOVA and Fisher's LSD ( $p < 0.05$ ). (C) Heatmap analysis for the statistically significant expression of the selected sugarcane assembled sequences (SAS) transcripts shown as "classes". When a transcript was not expressed, it is coloured as black in the heatmap analysis. (D) Climatic conditions for the field experiments. Fields 1 (F1) and 2 (F2) refer to the plants that were planted in April 2012 and October 2012, respectively. C1, C2, C3, and C4 refer to collection points 1, 2, 3 and 4 specifically 4, 8, 11 and 13 months after planting.

From the heatmap analysis, two clusters were observed regarding the selected transcripts (Figure 16 C). A cluster which mostly consisted of the *PER* transcripts and the only *F3H* transcript were higher in the I9 that were in the maturation phase (C3 from F2) as compared to the I9 that were still in the vegetative growth phase (C3 and C2 from F1; and C1 from F2). *CAD* and *PAL* transcripts were present in another cluster which presented higher

expression in the I9 of the plants experiencing water deficit and flowering. Interestingly, only one AS transcript was detected and it represented the *PER* gene (SCCCCL3003B02.b) and it presented the same results as those of the SSs (SCCCRT2002H08.g and SCJFSB1012D05.g) where their expressions were higher corresponding to the lignification process and lower with regards to the flowering process. From the three SAS representing the *PAL* gene, only one (SCEQRT1024E12.g) was significantly altered. In both fields, the expression of this transcript was significantly higher when culm lignification was occurring, and in F2 when flowering was occurring, as were the cases for the L1 tissues. The transcript representing *CAD* (SCACHR1038E08.g) was reduced as the lignification process was carried out in the I9 tissue and was higher when the flowering process occurred.

These results allowed us to explore important metabolic pathways involved in sugarcane development in photosynthetic, that is source, and storage, that is sink, tissues as described below. ]



## 4. DISCUSSION

### 4.1. Sugarcane phenology and morphological changes in response to precipitation and temperature alterations throughout development under field conditions

Sugarcane development is mainly being affected by four abiotic factors namely water availability (Hemrabha et al., 2004; Prabu et al., 2011), temperature variability (Ebrahim et al., 1998; Greenland, 2005; D. L. Liu et al., 1998), soil properties (Dominy et al., 2002; Johnson et al., 2016), and solar radiation (Robertson et al., 1996; Smit & Singels, 2006). In our study, abiotic stress was not the main focus, however during sugarcane development in the field, changes in the environment occur and a period of natural water deficit occurs which terminates the vegetative growth phase and initiates the maturation phase where sucrose is accumulated in the culms, a process known as “drying off” (Robertson et al., 1996; Singels et al., 2005).

The combination of the genetic makeup of the variety and its interaction with the external environment will determine the eventual developmental patterns (Greenland, 2005; Sanchurn et al., 2012; Singels & Inman-Bamber, 2011), as was shown in this experiment, and the eventual yield outcome, as previously described (de Medeiros Barbosa, 2015). The choice of the variety is crucial for obtaining raw material with appropriate maturation during the ripening season, thus, when choosing the variety, variables such as yield (e.g. tonne/hectare), industrial characteristics (e.g. Brix) and suitability for production environments, should be considered (Cardozo & Sentelhas, 2013). The variety studied here, SP80-3280, is a commercial variety with relatively high sugar and biomass yields and is well adapted to the climatic conditions of the state of São Paulo, Brazil (Riaño-Pachón & Mattiello, 2017; Souza et al., 2019).

It has been proved that the planting date and planting conditions have much greater effects on crop yield (biomass production) than on ripening (sucrose accumulation) in sugarcane crops (Alexander et al., 1973; Cardozo, 2012). In this study, alterations regarding precipitation and temperature led to different developmental patterns of SP80-3280 planted in two seasons of the year. The different developmental patterns were established during the early growth phases, around four months after planting, producing “one-and-a-half-year” and “one-year” sugarcane as described by de Medeiros Barbosa (2015). Sugarcane plants require large amounts of water, even under non-stressed conditions since they have to supply enough water to their leaves to sustain carbon fixation and overall growth (da Silva et al., 2013; Wiedenfeld, 2000). For the “one-and-a-half-year” sugarcane from the first field experiment, the observed retarded vegetative growth during the initial months after planting can be attributed to the lack of water availability since it has been proved that internode development and culm elongation are the most sensitive morphological alterations in sugarcane affected by water deficits (Carr & Knox, 2011; Lakshmanan & Robinson, 2013).

Thicker culms for the “one-and-a-half-year” sugarcane were also observed, supporting the findings by Cardozo (2012) which stated that when tillering occurs early in development under lower precipitation and temperatures, thicker culms are observed. It is known that sugarcane has the ability to tolerate early phase water deficits without significantly affecting future yields (Silva et al., 2013), as was the case with “one-and-a-half-year” sugarcane plants seeing that the vegetative growth and elongation phases were initiated as soon as the environmental conditions improved and precipitation increased. However, when water deficits are experienced during the rapid culm elongation phase, as was the case for the “one-year” sugarcane plants, the maturation phase is initiated due to the reduction in sink strength for structural growth (Singels et al., 2005), and in this case, yield loss can be substantial (Bell et al., 2018; Marli et al., 2012). These reports and our findings thus correspond and validate statements by de Medeiros Barbosa

(2015) which reported that “one-and-a-half-year” sugarcane have higher yields (> 120 t/ha) as compared to “one-year” sugarcane (< 100 t/ha) as described by de Medeiros Barbosa (2015).

A previous study which aimed at determining the effect of water deficit on sugarcane growth, showed that even under severe water limitations when culm elongation was halted, photosynthesis continued, however, at a reduced rate (Suriyan & Chalernpol, 2009), as was the case here seeing that photosynthate was still being produced in both “one-and-a-half-year” and “one-year” sugarcane, even when culm elongation was halted.

Mostly all previous studies alike the one presented here were conducted under controlled conditions and merely a few have explored field experiments, which highlight the relativity of this current study yet again. In an extensive field trial which measured the same growth parameters as here (plant height, number of internodes, culm diameter, and soluble sugar content), significant reductions were reported under stress conditions, except for culm diameter. Our results correspond to this seeing that the “one-and-a-half-year” sugarcane plants, which experienced low precipitation during the first months after planting, presented significantly larger culm diameters which persisted throughout development. This suggests that under conditions with lower amounts of available water, the carbon flow is directed towards culm thickening and lignification instead of culm elongation. This speculation is supported by other relative studies where a decrease in water availability resulted in the activation of stress response genes that overlapped with genes involved in culm lignification processes (Hu et al., 2009; Papini-Terzi et al., 2009; dos Santos et al., 2015; Suman et al., 2012).

Environmental temperatures also play an important role in sugarcane development and are considered a rate-limiting factor of culm elongation and carbon partitioning (Moore & Botha, 2013). It has been reported that when sugarcane is grown under higher temperatures, the carbon flow is directed towards fiber production instead of sucrose accumulation (Bonnett et al., 2006). Additionally, it has been shown that lignin, cellulose and hemicellulose contents increase with increasing temperatures in monocots (Moura et al., 2010). Another study indicated that a low lignin accumulation sugarcane variety presented a decrease in lignin content under lower temperatures, and *vice versa* for a high lignin accumulation sugarcane variety (dos Santos et al., 2015). This corresponds to our results seeing that SP80-3280 is not a high lignin accumulation variety.

Based on the results obtained here, it could be concluded that the planting season is one of the main factors determining the developmental profiles of SP80-3280, specifically regarding precipitation measurements. These plants can support early phase water deficit by continuing photosynthate production and carbon assimilation, however, the carbon flow will be directed towards culm thickening and lignification, instead of culm elongation and culm sugar accumulation. By planting SP80-3280 under low precipitation and temperatures, more biomass could be accumulated in the form of fiber over a longer period of time. In contrast, when sucrose accumulation is desired, rapid early-stage vegetative growth and culm elongation are necessary, followed by a period of “natural” water deficit, or “drying off” in order to activate the culm sugar accumulation phase. The shifts between the different developmental stages are dependent on signalling cascades occurring due to interactions between the plants’ interior and exterior environments.

Certain molecules and metabolic pathways, such as trehalose and trehalose metabolism, have been identified as signalling molecules and signalling pathways that can trigger carbon assimilation and can direct carbon flow (Bae et al., 2005; Delorge et al., 2014; Glassop et al., 2007; Lunn et al., 2014; Ponnu et al., 2011). Thus, it was of biological and economical importance to validate whether trehalose contents are increased under the moderate water deficit as described in the study presented here. It was also pointed out that other metabolites and metabolic pathways should be explored in order to determine how sugarcane metabolism is affected by differences in environmental conditions

in the field, specifically precipitation and temperature, which were the main objectives of the sections following this one.

#### **4.2. A metabolomics pipeline for the successful exploration of the sugarcane metabolome: from the analytical tool to the metabolite name**

In order to explore complex plant extracts, such as those from different sugarcane tissues as studied here (specifically the high-sugar content internodes), an appropriate analytical tool (Frag, 2014; Halket et al., 2005; Hall et al., 2002) and subsequent data generation and pre-treatment steps need to be put into place to yield high quality data for further downstream analyses (Nay Min Min, 2018; Salem et al., 2020). Here, successful analytical and bioinformatics pipelines were established for the analysis of the fairly unexplored sugarcane metabolome. It has been reported that the use of LC-MS is the analytical tool of choice when conducting plant metabolomics studies (De Vos et al., 2007; Grata et al., 2008; Moco et al., 2006).

Due to the complex nature of different plant metabolomes, LC-MS has been proved to be able to separate and detect key semi-polar compounds such as important secondary metabolite groups, especially in combination with the use of soft ionization techniques such as electrospray ionization (ESI) as used in the study presented here (De Vos et al., 2007). However, in order to get to the main results needed for drawing biological conclusions, that is the metabolite names and their relative intensities in each sugarcane tissue, rigorous sample pre-processing steps (Supplementary Figure 1) are required (Lämmerhofer & Weckwerth, 2013; Rudaz, 2015), which are considered the major drawback in plant metabolomics studies (Jaeger & Liseč, 2018; Salem et al., 2020; Töpfer et al., 2018).

After the collection of the raw data, that is the spectral data from the mass spectrometer in this case, a link must be made with the statistical analysis to answer the biological questions. Data pre-processing is of immense importance since these steps can significantly affect the subsequent statistical results (Ren et al., 2015). The main objective of data pre-processing is to improve the comparability between different samples. XCMS has been described as one of the desired tools for the processing of metabolomics data derived from LC-MS (Albóniga et al., 2020; Domingo-Almenara & Siuzdak, 2020; Forsberg et al., 2018; Smith et al., 2006). It generates peak tables with metabolite features and their respective mass to charge ratios ( $m/z$ ) and retention times ( $rt$ ). However, many metabolite features could represent one metabolite. To group the features arising from the same metabolite and to conduct the ion annotation steps, CAMERA has been described as an optimal choice (Kuhl et al., 2012) and was integrated here as part of the metabolomics data analysis pipeline. It is known, in order to draw conclusions regarding the biological questions at hand, that the nature of the metabolite feature needs to be elaborated prior to conducting statistical analysis, leading to the use of ProbMetab for metabolite annotation in each sample (Silva et al., 2014).

Studies have found that large and economically important groups of plant secondary metabolites such as alkaloids, phenolic acids, phenylpropanoids, flavonoids, glucosinolates, polyamines and derivatives thereof can be analyzed via the use of LC-MS (Moco et al., 2006; Rischer et al., 2006; Tolstikov et al., 2003). Thus, as mentioned in the results section, following the extraction protocol used in this study (Salem et al., 2016), semi-polar secondary metabolites were abundant in the sugarcane tissue extracts. This corresponds to and complements the above-mentioned results, thus strengthening the concreteness and robustness of the untargeted metabolomics results obtained here, specifically referring to the metabolite annotation following the pre-processing steps.

The nature of metabolomics data prevents us from conducting any statistical analysis prior to the data pre-treatment steps, which refer to the processes that metabolomics data need to undergo in order to extract the biologically relevant information from large datasets following the data pre-processing steps (van den Berg et al., 2006). This is specifically done in order to correct for aspects that hinder the interpretation of metabolomics data by highlighting the relevant biological information and improving the biological interpretation (Timmerman et al., 2015; van den Berg et al., 2006). The type of data pre-treatment method to use will depend on the nature of the data as well as the biological question at hand. In this study, the aim was specifically to determine how the metabolism of four different anatomical tissues from SP80-3280 differ from each other, and also between the two different field conditions. In order to determine which data pre-treatment method to use, it was necessary to analyse how the data react to each treatment and which effect each pre-treatment has on the data (van den Berg et al., 2006). Shapiro-Wilk's normality tests (Shapiro & Wilk, 1965) were implemented on each transformed dataset (none, log10, square root, cube root) to select for the optimal data transformation method yielding the most normally distributed datasets. This is crucial seeing that the assumption of normality is required to conduct most statistical tests and the Shapiro-Wilk's normality test performs the best and is the most powerful in determining normality in large datasets (Hanusz & Tarasińska, 2015). Based on these tests, and also the visual inspection of the graphic method to test for normality, which is the generation of Q-Q plots (Wilk & Gnanadesikan, 1968), it was observed that the log10 data transformation method improved the normality of the datasets the most. These findings correspond to previously reported studies regarding metabolomics data analysis pipelines (Jaeger & Lisec, 2018; Steinfath et al., 2008).

Scaling on the other hand, is also an important data pre-treatment method employed in metabolomics data analysis, aiming at adjusting for the huge differences in the fold changes between the different metabolites in the samples in order to make them more comparable (Gromski et al., 2015; van den Berg et al., 2006). Scaling can affect the downstream statistical analysis and final results, thus, in addition to the data transformation, different scaling techniques (none, auto, pareto) were also tested. In this case, the Shapiro-Wilk's normality tests were not able to determine which scaling technique yields the most normally distributed datasets. Consequently, PCA was conducted since it provided visual representations of how the different scaling techniques are able to highlight different aspects of the data in the scores and loadings (Jackson & Edward, 1991; Jolliffe, 2002).

As mentioned, the biological question at hand will be the deciding factor for choosing the type of data pre-treatment method. Since one of the questions in this study was to determine how the metabolism in each tissue is different between the different fields, the scaling methods that generated the most biologically sensible PCA models for each tissue, that is models that separated the two fields into more distinct clusters for each tissue, were selected.

Finally, the metabolomics data were "clean" and ready for the downstream statistical analysis using the metabolite names and their normalized, transformed and scaled intensities to answer the biological questions. It was concluded that this mode of action and these pipelines developed here, yield high quality results and can be implemented in the future for conducting sugarcane metabolomics studies, specifically working with complex sample matrices, such as those from sugarcane extracts.

### **4.3. The exploration of the sugarcane metabolome highlights the metabolic differences between four anatomical tissues and identify economically valuable compounds**

The metabolic profiles of the four different SP80-3280 tissues were mostly made up of polar and semi-polar metabolites involved in aromatic amino acid and plant secondary metabolism. This can be explained by the specific metabolite extraction protocol used (Salem et al., 2016), the fact that only the polar phase of the extractions (Figure 3 C) were selected for HPLC-MS analysis, and the parametrization of the HPLC-MS equipment. It has been described that HPLC-MS that makes use of negative ESI is able to ionize and detect common secondary plant metabolites such as phenolic compounds and cinnamic acid derivatives namely phenylpropanoids, stilbenes, anthocyanins and flavonoids (Rambla et al., 2015), as was the case in our study. The analyses of sugars and sugar phosphates have also been well established and studies have been done using the negative ionization mode (Games & Lewis, 1980; Wamelink et al., 2005). Additionally, previous findings have demonstrated that the positive ionization mode favours the ionization and detection of alkaloids (da Silva et al., 2017; Jakobová et al., 2012) and amino acids (de Puit et al., 2014; Harder et al., 2011). These reports strengthen the reason for the use of both the positive and negative ionization modes employed in the study, as well the detection of the specific metabolites in the tissues analysed here (Supplementary tables 2-5).

The highest number of metabolites were present in the leaves and decreased down the stem (Figure 5) which agree with previous metabolomics studies conducted on sugarcane tissue (Glassop et al., 2007). Interestingly, a relatively higher number of metabolites were reported in the current study in comparison to other untargeted sugarcane metabolomics studies in the literature (Bosch et al., 2003; Ferreira et al., 2018; Glassop et al., 2007), thus, contributing to knowledge gaps regarding the sugarcane metabolome and presenting an improved method for conducting sugarcane metabolomics studies.

The leaves of sugarcane, and plants in general, are the first and primary aerial tissues which will detect and respond to changes in the external environment (Li et al., 2016; Smit & Singels, 2006; Wahid & Close, 2007), explaining the presence and the large amount and chemically diverse metabolites detected in the leaf tissue as compared to the internodal tissue as reported in this study (Supplementary tables 3-5). Additionally, culm lignification down the sugarcane culm and cell wall recalcitrance to the breakdown for successful metabolite extraction are also the possible causes for this phenomenon (Bottcher et al., 2013; Hodgson-Kratky et al., 2019), corroborating the decrease of the detected metabolites down the culm. These findings can also be supported by a study on the metabolic shifts in the development of tomato (*Lycopersicon esculentum*) which reported significant decreases in metabolic activity during ripening (Osorio et al., 2011) seeing that it is also considered that sugarcane ripening occurs basipetally (Moore et al., 2013).

The combined implementation of both PCA and PLS-DA led to the identification of the metabolic differences between the four tissues studied here (Figure 7), following the metabolomics profiling of each of these tissues. PCA and PLS-DA are multivariate methods used for the reduction of complex datasets in order to highlight the variance within (Uarrota et al., 2014; Worley & Powers, 2013). They have been extensively used in metabolomics studies and have led to the unravelling and understanding of important processes regarding plant metabolism (Abdel-Farid et al., 2014; Li et al., 2013; Okada et al., 2010). These two statistical tests were used to complement each other since it has been evidenced that initial unsupervised PCA models can formulate initial biological hypotheses, followed by the use of supervised PLS-DA models to test and verify the PCA models (Worley & Powers, 2013). Permutation

analyses (Supplementary figure 6) allowed for the validation of the PLS-DA models generated for each tissues, and in return, ensured high quality results by preventing the overfitting of the models (Szymańska et al., 2012).

Based on the PCA and PLS-DA models and their shared principal components (PCs), in this case the specific metabolites, it was possible to pinpoint the main metabolites responsible for the separation of the tissues into different clusters (Figure 7). It was found that metabolites involved in the upstream processes of ethylene biosynthesis, plant-pathogen interactions and flavonoid biosynthesis were high in the upper and young internodes. Ethylene is an endogenous plant hormone (Yang & Hoffman, 1984) involved in different stages of plant development (Fong Chong et al., 2010; Pratt & Goeschl, 1969; Roberts, 2018) and it has also been described as a roleplayer in plant-pathogen interactions (Boller, 2018; Broekaert et al., 2006; Ecker & Davis, 1987), supporting our reported findings based on different developmental tissues of sugarcane. To elaborate, the abundance of metabolites involved in ethylene production in the upper and young internodes could possibly be explained by the presence of endophytic microorganisms in these environments which were pointed out by the transcriptomics results (described in section 3.2.) which reported the presence of transcripts and GO terms related to non-plant organisms (Table 7). However, the presence of ethylene metabolism transcripts and transcription factors were observed (Supplementary table 7) and previously described (Cunha et al., 2017), thus not excluding the possibility of sugarcane being able to produce these metabolites by itself. Regarding the anatomy of the sugarcane culm, it is known that it consists of recurring internodes surrounded by a leaf blade and sheath (Moore et al., 2013; Moore & Botha, 2013). As the culm elongates, these leaves senesce basipetally down the culm. Studies have shown that ethylene is involved in the abscission, or shedding, of leaves (Botton & Ruperti, 2019; Morgan, 1984), thus giving an additional possible explanation of the results obtained here.

Looking specifically at metabolites themselves and not the metabolic pathways in which they participate, L-proline, one of the metabolites identified as being responsible for distinguishing between the different sugarcane tissues (Figure 7; Supplementary table 6), caught our attention. It presented high abundances in the upper (I1) and young (I5) internodal culm tissue (Supplementary tables 3 & 4). These results correspond to a previous study reporting that this amino acid is one of the major free amino acids in sugarcane culm tissue (Tejera et al., 2006). It has also been proved that this amino acid is present in low amounts in mature sugarcane culm tissue (Guerzoni et al., 2014), as was found in our study since the abundance decreased down the culm. Proline is different from other amino acids regarding its structure, which is cyclic and thus provides conformational rigidity to protein structures (MacArthur & Thornton, 1991). The accumulation of proline in plants under stress conditions have been elaborated extensively (Kavi Kishor et al., 2015; Sharma & Verslues, 2010), however, it has also been proved that this amino acid is present in plants under normal physiological conditions (Kavi Kishor & Sreenivasulu, 2014), taking part in plant growth and development (Funck et al., 2012; Mattioli et al., 2009). Since the upper and young internodes are actively growing, and culm elongation is occurring in these regions, the higher levels of proline are required since this metabolite has been described as a key player needed for cell elongation in previous studies (Dvořáková et al., 2012).

Another metabolite, spermidine, was also higher in the upper and young internodes as compared to the leaves and mature internodes. Spermidine is a polyamine (PA) and is widely distributed among eukaryotes and prokaryotes (Chen et al., 2018; Liu et al., 2013). These metabolites can exist in their free form (PA), or covalently (CC-PA) or non-covalently (NCC-PA) conjugated as described in (Gholami et al., 2013). In higher plants, they are mostly present in their free form and it has been found that these metabolites follow tissue specific accumulation patterns, as was found in our study. Supporting our results, a study conducted by Mustafavi et al. (2018) concluded that spermidine is higher in plant culm as compared to the leaf tissue. As mentioned, the upper and young internodal tissues are actively

growing and expanding, and high PA levels have been associated with vigorous plant growth and metabolism (Zhao et al. 2004; Cai et al. 2006), possibly supporting the high spermidine levels in these specific sugarcane tissues.

Our results indicated higher levels of *p*-coumaric acid in the same tissues as spermidine and it has been shown that PAs can bind with phenolic compounds such as hydroxy cinnamic acid, coumaric acid, caffeic acid, or ferulic acid to form CC-PAs (Luo et al., 2009; Martin-Tanguy, 1997), thus possibly supporting our observations and the positive association between spermidine and *p*-coumaric acid. When PAs bind with molecules such as acidic proteins, membrane phospholipids, and nucleic acids, NCC-PAs are formed (Igarashi & Kashiwagi, 2015), which have been implicated in growth and development processes such as DNA replication, gene transcription, cell division and membrane stability (Chen et al., 2018). These processes overlap with the transcriptomics results obtained (described in section 3.2) where there were higher expressions of genes with the GO terms related to cell cycle, including chromosome organization and DNA replication (Table 7) in the upper and young internodes as reported by the transcriptomics data. Thus, there is increasing evidence that targeting the increase of PAs such as spermidine can positively affect plant growth and productivity (Chen et al., 2018). The presence of PA biosynthesis transcripts was also detected in the transcriptomics data retrieved in the present study (Supplementary table 7), thus the targeting of these genes will be possible for hypothesis testing in SP80-3280.

On the other hand, the metabolite *p*-coumaric acid by itself has also been linked to plant growth. The exogenous application of *p*-coumaric acid to chia (*Salvia hispanica*) led to increased shoot elongation and biomass accumulation, and presented positive correlations with proline contents (Nkomo et al., 2019), as was observed in our case. A metabolomics study conducted on two maize (*Zea mays*) varieties (Co125 and W401) identified *p*-coumaric acid in the cell walls of the culms and concluded that this metabolite decreased with cell age as the lignification process increased down the culm (Migné et al., 1994), supporting our results of higher contents of *p*-coumaric acid in the upper and young internodes as compared to the mature internodes. This same study also presented that ferulic acid decreases as the tissues matured, that is down the sugarcane culms, which was also the case in our study.

In contrast to these observations, metabolites involved in tyrosine metabolism and glucosinolate biosynthesis were found to be higher in the leaves (L1) and mature internodes (I9) as compared to the younger internodes (I1 and I5) studied here (Figure 7; Supplementary table 6). It has been described that tyrosine-derived metabolites, for example tocopherols (Maeda et al., 2014), plastoquinones (Nowicka & Kruk, 2010), cyanogenic glycosides (Mithöfer & Boland, 2012) and suberin (Roppolo & Geldner, 2012) play important roles in plants. They can act in plant defence, attraction of pollinators, serve in electron transport and form part of the structural support in the cell walls of plants (Schenck & Maeda, 2018). In the literature, evidence suggest that the tyrosine biosynthetic pathway differ within the plant kingdom (Yoo et al., 2013). These interspecies variations might be the gateway to enhance the production of tyrosine and tyrosine-derived metabolites via synthetic biology platforms seeing that these compounds are of great economic importance (Schenck & Maeda, 2018).

A metabolite called 2-(2'-Methylthio)ethylmalic acid from the glucosinolate biosynthetic pathway, was also highlighted in the leaves and mature internode (Figure 7; Supplementary table 6). It is considered an aliphatic glucosinolate since it is derived from methionine, as are the majority of glucosinolates in plants (Redovniković et al., 2012). It has been shown that glucosinolates have herbivore deterrent (Badenes-Perez et al., 2014; Wittstock et al., 2003), fungicidal (Fahey et al., 2001; Rongai et al., 2009), bactericidal (Borges et al., 2015), nematocidal (Avato et al., 2013; Oliveira et al., 2011), and allelopathic properties in plants (Halkier & Gershenzon, 2006; Wittstock & Halkier, 2002). In the industry, these glucosinolates from plants have also been greatly explored and exploited for their roles in cancer treatment and the regulation of blood glucose levels in humans (Dinkova-Kostova & Kostov, 2012; Johnson,

2002), as biopesticides in crop plants (Claros Cuadrado et al., 2019; Popova et al., 2017) and flavour compounds (Halkier & Gershenzon, 2006). It has been described that glucosinolate distribution in plants vary under different environmental conditions (Wittstock & Halkier, 2002) and depending on the different plant organs and tissues (Halkier & Gershenzon, 2006), as was the case here seeing they were higher in the leaves and mature internodes. The enzyme responsible for the production of the aliphatic glucosinolate identified here, methylthioalkylmalate (MAM) synthase, has been found to be highly expressed in the leaves and lower culm tissue of *Arabidopsis*, which resulted in higher contents of this metabolite in these tissues (Redovniković et al., 2012), supporting our findings. The overexpression of this gene in Chinese cabbage (*Brassica rapa*) led to increased levels of aliphatic glucosinolates with no altered or negative phenotypes (Zang et al., 2008), however not much is known about sugarcane glucosinolates, thus highlighting new possibilities for the exploration of the sugarcane metabolome and its use in the production of economically valuable products in certain tissues.

#### **4.4. Metabolite accumulation in different sugarcane tissues allow to distinguish between the different developmental profiles of “one-year” and “one-and-a-half-year” sugarcane crops**

In order to identify the exact metabolites responsible for the separations between the “one-and-a-half-year” sugarcane from field 1 (F1) and “one-year” sugarcane from field 2 (F2) in each tissue, the PCA and PLS-DA models were inspected separately, and the main metabolites responsible for the separations in each tissue (L1, I1, I5 and I9) were highlighted (Figures 8 - 11).

Looking at the leaf tissue, the majority of the identified metabolites were higher in the leaves of the “one-year” sugarcane plants from field 1, which presented earlier maturation, shorter internodes, and thinner culm diameters (Figure 1 A - E). During the vegetative growth phase, these plants experienced a period of water deficit (Figure 1 F) which led to the initiation of the maturation phase, a natural process known as “drying off”. Consequently, an unwanted shift to the reproductive stage (flowering) occurred at the end of the experiment. It could be concluded that the SP80-3280 plants from this field experienced water deficit at the moment when the presence of sufficient water was crucial, namely the vegetative growth stage. Additionally, these plants went through all the phenological changes associated with the life cycle of these crops (Moore & Botha, 2013), whereas the “one-and-a-half-year” sugarcane plants from field 2 only experienced the tillering and vegetative growth phases.

Out of the 32 metabolites identified in the leaves for being responsible for the separation between the two fields, the majority of them were phenylpropanoids and intermediates of the phenylpropanoid biosynthesis pathway (Figure 8 C). Phenylpropanoids are phenolic, plant secondary metabolites derived from the aromatic amino acids phenylalanine and tyrosine (Barros et al., 2016) which were also abundant in the leaves of the “one-year” plants (Figure 8 C), and are commonly divided into 5 groups namely flavonoids, monolignols, phenolic acids, stilbenes, and coumarins (Noel et al., 2005; Vogt, 2010). Phenylpropanoid biosynthesis commences with the deamination of phenylalanine to yield cinnamate via the *PAL* enzyme. Further hydroxylations of cinnamic acid give rise to *p*-coumaric acid and caffeic acid, one of the compounds that was also abundant in the leaves of the “one-year” sugarcane plants. Furthermore, hydroxycinnamoyl CoA thioesters are formed which are the main substrates for the production of lignin monomers and flavonoids (Hahlbrock & Grisebach, 1979; Hahlbrock & Scheel, 1989). It is known that the lack of water affects plant flavonoid biosynthesis. A study conducted on drought tolerant wheat plants proved that the genes involved in

the flavonoid biosynthetic pathway and the accumulation of flavonoids in the leaves were closely related to the observed drought tolerance (Ma et al., 2014). Increases in flavonoid contents were also observed in leaves of the Chinese skullcap plants (*Scutellaria baicalensis*) grown under low soil water contents, and additionally, the key genes involved in the flavonoid biosynthetic pathway, *PAL* and chalcone synthase (*CHS*), were upregulated (Qin et al., 2010). This also corresponds to a previous study that showed that oxidative and drought stress conditions increased the flavonoid contents in *Arabidopsis* leaves, specifically anthocyanins (Nakabayashi et al., 2014), as was the case here seeing that pelargonin, which is involved in anthocyanin biosynthesis, was abundant in the leaves of the “one-year” sugarcane plants (Figure 8 C). Anthocyanins have antioxidant properties (Earling et al., 2019; Kähkönen & Heinonen, 2003; Neill & Gould, 2003), thus the water deficit that the “one-year” sugarcane plants experienced might be the reason why pelargonin levels were high in these leaves.

Two other flavonoids, afzelin and quercitrin, both of which have been reported to be present in many plant leaves (Chen et al., 2002; de Barros et al., 2016; He et al., 2016; Lee et al., 2014; Shi et al., 2018) were only detected in the leaves of the “one-year” sugarcane from field 2 (Figure 8 C). Afzelin is a flavonol glycoside which is well known for its function as a reactive oxygen species (ROS) scavenger (Bergman et al., 2003; Mira et al., 1994; Trembl & Šmejkal, 2016). Water deficit leads to oxidative stress and the formation of ROS (Cruz de Carvalho, 2008; Gupta, 2010), possibly supporting the higher abundances of these types of metabolites in the leaves from the “one-year” sugarcane plants (Figure 8 C) since they underwent a period of water deficit when they were in the vegetative growth phase (Figure 1 F), a phase in which water availability is crucial, as already mentioned. Quercitrin on the other hand, is a flavonoid and it has been shown to inhibit carbonyl reductase (NADPH), aldehyde reductase and tyrosinase (Imamura et al., 2000; Jeong & Shim, 2004; Matsuda et al., 2002). This role of the inhibition of tyrosinase, an enzyme responsible for the breakdown of tyrosine (Bijelic et al., 2015; Kanteev et al., 2015), might explain why the tyrosine levels were high in the leaves of the “one-year” sugarcane plants. A study done to determine the metabolic response of maize leaves under water deficit presented increases in the tyrosine levels (Bowne et al., 2012). In addition to serving as an input for the biosynthesis of the above-mentioned compounds, tyrosine has also been described to be important for plant growth. Foliar application of tyrosine to beetroot (*Beta vulgaris*) led to increases in the leaf carotenoid and chlorophyll contents, as well as higher leaf carbohydrate contents (Rashad El-Sherbeny & Da Silva, 2013), which were also the case in our study where higher amounts of sucrose were present in the leaves of the “one-year” sugarcane plants, and also as previously described (Cardozo & Sentelhas, 2013).

In addition to the phenylpropanoids that were abundant in the leaves of the “one-year” sugarcane, the monolignols cis-beta-D-Glucosyl-2-hydroxycinnamate, caffeic acid and caffeoylquinic acid, were also abundant (Figure 8 C). Monolignols are the monomeric units that form the base of lignin in plants. Lignin biosynthesis has been studied in almost all model plants (Faraji et al., 2018; Kawaoka & Ebinuma, 2001; Li et al., 2010; Storme, et al., 2012) and agronomically important crops (Ambavaram et al., 2011; Dixon et al., 1996; Hisano et al., 2011; Morreel, et al., 2012), providing evidence of the control of monolignol transport and polymerization in different cell types and tissues. The phenomenon that these compounds are being produced in non-structural tissues such as leaves has been reported (Hu et al., 2010; Lanot et al., 2006; Min et al., 2014), and the expression of monolignol biosynthetic genes are not necessarily correlated with the presence of lignin (Wang et al., 2013) seeing that they can be used to produce a wide range of other phenylpropanoid derived compounds (Bassard et al., 2010; Goujon et al., 2003). A co-expression analysis conducted to study lignin biosynthesis in *Arabidopsis* and *Brachypodium* used monolignol biosynthetic (4-coumarate coenzyme A ligase - *4CL*; cinnamate 4-hydroxylase - *4H*; caffeoyl-CoA *O*-methyltransferase - *CCoAOMT*) and monolignol polymerization (Dirigent protein 6 - *DIR6*; laccase 4 - *LAC4*; laccase 17 - *LAC17*) genes as input for the gene networks.

Separate gene clusters were observed, with the monolignol biosynthetic genes presenting co-expression patterns with some genes involved in the phenylpropanoid biosynthetic pathway (phenylalanine ammonia lyase - *PAL*; caffeic acid *O*-methyltransferase - *COMT*; cinnamyl alcohol dehydrogenase - *CAD*; hydroxycinnamoyl-CoA reductase - *CCR*; *p*-coumaroyl shikimate 3'-hydroxylase - *C3H*; hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl transferase - *HCT*), and the monolignol polymerization genes presenting co-expression profiles with the cellulose synthase (*CE3*) genes instead (Ruprecht & Persson, 2012). Thus, these results suggest that genes such as *LAC4* and *LAC17*, or other monolignol polymerization genes should be targeted when lignin related studies are conducted (Wang et al., 2015; Zhao et al., 2013). In addition, transporters of monomers could also provide useful insights on how these compounds are synthesized in the leaves and transported to the tissues that provide structural support to the plants, such as the culm.

In the upper and young culm tissue, the main metabolites highlighting the relatively small differences between the two fields (F1 and F2), that is between “one-year” and “one-and-a-half-year” sugarcane, were found to be mostly involved in primary metabolism (Figures 9 and 10; Supplementary tables 4 and 5). However, two metabolites namely digalacturonic acid and d-galacturonate, which are pectic cell wall monomers (Khanh et al., 1991; Searle-van Leeuwen et al., 1996), were higher in the upper internodes of the “one-year” sugarcane plants from field 2 (Figure 9 C), which presented thicker culms and were shorter (Figure 1 A). These compounds are considered important constituents of plant biomass and their microbial degradation has been investigated for converting pectin-rich agricultural residues such as sugar beet pulp, citrus waste and apple pomace into value added products (Edwards and Doran-Peterson, 2012; Marzo et al., 2019). At this moment, these types of residues are sold as animal feed at low values (Edwards & Doran-Peterson, 2012). Since the upper and young internodes of sugarcane are not rich in sugar, the possibility exists of exploring these aerial parts for value added products when thick culms are present.

From the “one-year” sugarcane, higher amounts of an amino acid, L-glutamate, were reported. The exogenous application of L-glutamate to Arabidopsis, rice and tomato plants resulted in the inhibition of growth in the specific tissues to which it was applied to (Kan et al., 2017; Kim et al., 2010; Walch-Liu et al., 2006; Yang et al., 2017). On the other hand, the expression of a bacterial NADPH glutamate dehydrogenase (GDH) gene, which is responsible for the breakdown of glutamate, in tobacco plants led to an increase in the overall plant growth and biomass production (Ameziane et al., 2000). Thus, the high amounts of L-glutamate in the “one-year” compared to the “one-and-a-half-year” sugarcane plants could possibly explain the growth inhibition at the apical parts of the sugarcane culms seeing that the “one-year” plants were significantly shorter than the “one-and-a-half-year” plants. A carbohydrate, D-glucosamine was also implicated here as being lower in these exact tissues in the “one-year” sugarcane plants, thus corresponding to the other observation made by Ameziane et al. (2000) where they reported an increase in the total carbohydrate content when the GHD gene is highly expressed and consequently, low glutamate levels.

The cyanogenic glucoside, dhurrin, synthesized from tyrosine, was initially identified in sorghum (*Sorghum bicolor*) and it is reported that this compound is abundant in young plants and absent in mature plants (Dunstan & Henry, 1902). It was also shown that this compound is present in the aerial shoots of sorghum and decreased as the plants age (Akazawa et al., 1960). A study was done to determine if the dhurrin biosynthetic genes are present in sugarcane by data mining through the SUCEST-FUN database (<http://sucest-fun.org/wsapp/>). The authors reported sugarcane enzymes with high similarity to those of sorghum, suggesting that the this pathway might be conserved in these organisms (Junior et al., 2007). The results from our study confirmed the presence of dhurrin in sugarcane and also conforms to the reports by Dunstan and Henry (1902) and Akazawa et al. (1960) seeing that the sugarcane plants that did not reach the maturation phase presented higher amounts of dhurrin in the young internodal tissue as

compared to the plants that had gone through the maturation phase. This compound has also been reported to play a role in pathogen response, and as previously mentioned and described in section 4.4., endophytic microorganisms were in fact present in these tissues.

Mature internodal tissue from sugarcane is known for its ability to accumulate large amounts of sugar, mainly in the form of sucrose, when the plants reach the maturation phase. However the breakdown products of sucrose are also present in these tissues due to a process called futile cycling (Nguyen-Quoc & Foyer, 2001; Uys et al., 2007). When the plants are still actively growing, this process is responsible for producing the products and energy for other biosynthetic reactions, as was the case for the “one-and-a-half-year” sugarcane plants and thus the main reason why D-glucose was higher in the mature internodal tissue of these plants in comparison with the “one-year” sugarcane plants seeing that these plants had already entered the sucrose accumulation, thus the maturation, phase.

#### **4.5. The exploration of the sugarcane transcriptome proves overlaps between the metabolite profiles drawn for each anatomical tissue**

Global gene expression studies, such as the one that was conducted here, allow for the identification of groups of target genes presenting tissue specific expression (Casu et al., 2003, 2015; Grimplet et al., 2007). From a biotechnological perspective, these types of genes can possibly be considered as strong candidates for providing promoters that can be used for a more controlled expression of transgenes in crops (Hennig, 2007). In this study, a data dimensionality-reduction technique, specifically MDS, was implemented to reduce and group the variation present within the huge datasets of transcripts and their expression intensities (Hefner, 2007). The employment of enrichment techniques in “omics” studies is crucial for the understanding of complex systems. They reduce the complexity of the data at hand and improve the interpretation of the results by integrating a priori biological knowledge of the genes under study (Alhamdoosh et al., 2017; Marco-Ramell et al., 2018). The exploration of the sugarcane transcriptome via the generation of these MDS models and GSEA, led to the identification of gene clusters responsible for the anatomical differences between the four sugarcane tissues analysed here (Figure 12 and 13; Table 7). By using the combination of these two strategies, biologically relevant information could be extracted from the large lists of significantly expressed genes in each tissue, and it allowed for the comparison of the results obtained from the analysis of the sugarcane metabolome.

In this study, three different data dimensionality-reduction techniques were implemented separately in order to determine if all three would yield similar results (Nguyen & Holmes, 2019), from both a metabolomics and transcriptomics point of view, which was in fact the case (Figures 8, 11 and 12). The majority of the variation within the transcriptomics datasets could be explained by the differences between the four anatomical tissues and distinctions between fields, namely “one-year” and “one-and-a-half-year” sugarcane, could be made in the leaves and mature internodal tissue, as was the case when analysing the PCA and PLS-DA models that were drawn with the metabolomics data (Figures 8 and 11). Thus, the overlaps of these results of data that were generated separately adds to the robustness of the overall findings. With the annotation of the SAS from the SUCEST-FUN database (<http://sucest-fun.org/wsapp/>) and their links to the respective GO terms (Harris et al., 2004), it was possible to determine how the anatomical tissues differ from one another on a transcriptomics level.

Genes responsible for specialized carbohydrate metabolism such as cell wall organization and construction, and sugar metabolism were practically absent in the leaf tissue as compared to the internodal tissue where they were

highly expressed. To our knowledge, there is no study specifically highlighting the different transcriptomics profiles of four different anatomical tissues from sugarcane in field conditions, however it is well known that cell wall construction and lignification increase down the sugarcane culm (Bottcher et al., 2013; Casu et al., 2007; Engels, 1998). In contrast, genes involved in secondary metabolic processes and organic acid metabolism were highly expressed in the leaves, which also correspond to the results from metabolomics data analysis in which most of the metabolites responsible for the separation of the different anatomical tissues formed part of secondary metabolism and organic acid metabolism as well (Table 7).

In plants, the leaf tissue is one of the main elements responsible for perceiving and responding to changes in the external environments (Sultan, 2010; Vanneste & Friml, 2009). Plant secondary metabolism was evolved to allow plants to perceive and adapt to these changing environments (Kurosaki, 2012; Vom Endt et al., 2002), thus justifying the high expression of genes and the large amounts of secondary metabolites in the leaf tissue reported here. Additionally, the leaf is also responsible for supporting the growth of the apical meristem, or the upper internode in this case, which is actively growing and elongating. These tissues require building blocks such as amino and organic acids from the source tissue as reported here, thus justifying the high expression of these genes involved in organic amino acid biosynthesis.

This brings us to another cluster of transcripts that were highly expressed in the upper and young internodes (Figure 12; Table 7). These transcripts mostly formed part of the processes responsible for cell duplication and growth, and culm elongation, thus supporting the role of these tissues as actively growing tissues. Lastly, a cluster of transcripts related to the upstream and downstream processes of lignification were highly expressed in the more mature sugarcane tissues (Figure 12; Table 7), thus corresponding to previous studies conducted on the lignification of the sugarcane culm (Bottcher et al., 2013; dos Santos et al., 2015).

Since one of the main objectives of this study was to integrate the metabolomics and transcriptomics data since doing this can lead to great advances of understanding various physiological processes (Joyard & McCormick, 2010; Sheth & Thaker, 2014), focus was shifted towards using unsupervised methods for the integration of these two data modalities as discussed in section 4.6., and identifying key metabolic pathways based on the metabolites and the genes that code for them in the different tissues and how they are affected as environmental conditions change, as discussed in sections 4.7 and 4.8.

#### **4.6. The direct combination of two “omics” in an unsupervised manner pinpoints the main metabolic pathways being altered throughout sugarcane development**

The use of different “omics” technologies and their integration have taken off in recent years (Misra et al., 2018; Wani & Raza, 2019). It has led to the understanding of essential plant processes regarding development (Agrawal et al., 2015; Gapper et al., 2014; Hennig, 2007) and responses to changes in plants’ internal and/or external environments (Debnath et al., 2011; Gupta et al., 2013; Ruotolo et al., 2018). For example, metabolomics and transcriptomics analyses have been applied to study alterations in metabolic systems of field grown rice (Izawa et al., 2011) and barley (Kogel et al., 2010), and in transgenic *Arabidopsis* plants presenting increased leaf growth (Gonzalez et al., 2010). However, in these mentioned studies, the statistical analyses of the metabolomics and transcriptomics data were done separately without using a specific tool for their direct integration, as is the case for most other “multi-omic” or “double-omic” studies (Gao et al., 2014; Shin et al., 2018; Udhane et al., 2017), since this type of statistical

integration is an extremely complex task. With the improvements of high-throughput technologies and surges in the amounts of data being generated via these “omics” studies, a bottleneck arises: the way of dealing with these huge, heterogeneous datasets (Mochida & Shinozaki, 2011). In our study, it could be concluded that the key for highlighting target pathways, metabolites and genes, the use and the improvement of “omics” studies and the development of different bioinformatics strategies for the statistical analyses and integration of these types of data is essential (Langridge & Fleury, 2011; Mochida & Shinozaki, 2011), which was in fact one of the main objectives of this PhD study.

Here, a bioinformatics tool called Multi-Omics Factor Analysis (MOFA) was implemented in order to deal with this expected problem regarding the huge, heterogeneous datasets (Canzler et al., 2020; Ulfenborg, 2019). It is a tool designed for the direct integration of multiple data modalities in an unsupervised manner, followed by the discovery of the main sources of variation that are shared between the “multi-omic” datasets (Argelaguet et al., 2018), in our case between the metabolomics and transcriptomics datasets. Specific latent factors (LFs) are then identified which capture the major shared sources of variation across the different data modalities, and shared patterns can be highlighted using this tool. Here, the LFs represented metabolic pathways and their activities, and specific transcripts directly upstream or downstream linked to the metabolites within the highlighted metabolic pathways (Supplementary Figures 2 - 5).

It has been described that the metabolome is the closest link to the phenotype (Fiehn, 2002; Hong et al., 2016). In a review of plant metabolomics studies, the authors stated that metabolomics could provide insights on different phenotypes regarding morphology, tissue identity, development (phenology) and environmental response, among others (Hall et al., 2002). These were the exact topics that we were aiming to investigate during this PhD and its development over the years. This resulted in an important observation and conclusion: the metabolomics data modality was the golden link of our specific molecular analyses since it pinpointed the exact metabolic pathways and the respective metabolites in each pathway that were altered in the different tissues throughout development and in response to changes in the external environment (Table 10). Using this, it was also possible to pinpoint the exact target genes, as discussed in section 4.8.

For the different sugarcane tissues of SP8032-80 analysed here, the main metabolic systems that were altered throughout development and in response to changes in the external environments were amino acid metabolism, mostly phenylalanine, tyrosine and tryptophan metabolism; and secondary metabolism, mostly phenylpropanoid metabolism and the metabolism of its upstream precursors and/or downstream products. A study done to determine the metabolomic and transcriptomic changes *Arabidopsis* leaves undergo between different developmental stages, reported alterations in amino acid and organic acid metabolism, as well as secondary metabolism such as cell wall precursor related metabolism (Skirycz et al., 2010), thus corresponding to our study. Previous studies on the effects of abiotic stress on plants such as maize, cotton and tomato have reported alterations in amino acid metabolism (Batista-Silva et al., 2019; Hildebrandt, 2018; Ranieri et al., 1989; Showler, 2002). A study done in *Arabidopsis* to gain insights on the direction of amino acid metabolism under abiotic stress conditions, as well as the steps leading to secondary metabolite synthesis, proved that aromatic amino acids such as phenylalanine, tyrosine and tryptophan, are the main amino acids being synthesized under stress conditions (Hildebrandt, 2018). This also corresponds to bioinformatics studies which postulated that there are inductions in aromatic amino acid synthesis when plants are subjected to stress conditions (Obata & Fernie, 2012; Urano et al., 2009). These aromatic amino acids serve as precursors for secondary metabolites such as glucosinolates, alkaloids, and phenylpropanoids, which play important roles in plant development (Dare et al., 2017; Ferrandino & Lovisolo, 2014) and biotic and abiotic stress response (Mata-Pérez et al., 2016; Tzin

& Galili, 2010). Our results from MOFA pointed out these exact mentioned pathways (glucosinolate biosynthesis; tropane, piperidine and pyridine alkaloid biosynthesis; and phenylpropanoid biosynthesis). Additionally, the “ripening” developmental stage of grapevine (*Vitis vinifera* L.) has been associated with alterations in the phenylpropanoid metabolic pathway, specifically by the upregulation of genes involved in the flavonoid biosynthetic pathway due to a decrease in the water availability (Ferrandino & Lovisolo, 2014). In our study, abiotic stress was not the main focus, however during sugarcane development in the field, changes in the environment occur and a period of natural water deficit is necessary for the termination of the vegetative growth phase and the initiation of the maturation phase where sucrose is accumulated in the sugarcane culms, a process known as “drying off” (Robertson et al., 1996; Singels et al., 2005). Accordingly, sugarcane plants do in fact experience a period of some mild, but natural, form of abiotic stress when cultivated in the field, leading to the conclusion that this phenomenon could not be disregarded in our study. Additionally, GO terms pointing towards abiotic stress response were also highlighted in the MOFA models, showing that stress response genes are altered throughout development in response to changes in the external environment.

It is well known that the phenylpropanoids such as flavonoids are key players in the synthesis of cell wall precursors (Huang et al., 2010; Nanda et al., 2016) and function in plant development (Borevitz et al., 2000; Davin & Lewis, 1992; Vogt, 2010). A study done on transgenic tobacco plants by introducing a bean (*Phaseolus vulgaris*) phenylalanine ammonia-lyase (PAL) gene, the main enzyme in the phenylpropanoid pathway, showed that alterations to the phenylpropanoid pathway lead to altered phenotypes such as altered leaf shape and texture, reduced lignification in xylem, stunted growth, reduced pollen viability, and altered flower morphology and pigmentation (Elkind et al., 1990). Another study conducted in *Arabidopsis* also concluded that alterations to the phenylpropanoid and flavonoid pathways result in plant growth reduction and delayed flowering (Sharma et al., 2016). Since the main phenotypic differences of the plants from the two different fields from our study were related to plant height and culm diameter, the reported findings regarding the pathway activities conclude that the phenylpropanoid biosynthetic pathway and its derivative pathway, flavonoid biosynthesis, are crucial to plant development and adaptation to changes in the plant's internal and/or external environments, as was previously reported (Biała & Jasiński, 2018; Naoumkina et al., 2010; Yang et al., 2018).

The alteration of the pentose phosphate pathway (PPP) and pathways directly derived from it, throughout development was also an interesting observation. The PPP serves to maintain carbon homeostasis, to provide precursors for nucleotide and amino acid biosynthesis, to regulate oxidative stress (Pryke & Rees, 1977; Stincone et al., 2015) and most importantly, it is the main source of the reduced form of nicotinamide adenine dinucleotide phosphate (NADPH), an essential reductive coenzyme that contains an adenine nucleobase, thus the identified GO terms “nucleobase-containing small molecule metabolism” and “cofactor metabolic process” are most likely associated with the alterations of this metabolic pathway. The NADPH coenzyme is mainly produced and re-generated via NADP-dehydrogenases enzymes from the PPP such as glucose-6-phosphate dehydrogenase (G6PDH) and 6-phosphogluconate dehydrogenase (6PGDH) (Corpas & Barroso, 2014). It serves as a reducing agent for various biosynthetic reactions in the cell such as cell division, signal transduction, cell growth and development in eukaryotic cells (Corpas & Barroso, 2014; Pryke & Rees, 1976; Pryke & Rees, 1977). In our case, it is suspected that the metabolic energy produced by the PPP is most likely for fueling the production of lignin precursors (Pryke & Rees, 1977) seeing that the phenylalanine, tyrosine and tryptophan metabolism; phenylpropanoid biosynthesis; and flavonoid biosynthesis metabolic pathways were also highlighted by the MOFA models (as described above), which are the principal pathways leading to eventual lignin production (Bottcher et al., 2013; Vogt, 2010).

During the collection of plant tissues from field experiments, and the proper nature of plants as living organisms, endogenous and/or exogenous organisms are naturally present in and/or on the plants' internal and/or external environments (Bové & Garnier, 2002; Hayat et al., 2010; Vargas et al., 2003), hence the reported non-plant altered metabolic pathways all grouped under the term xenobiotics degradation and metabolism by the KEGG BRITE classification system (<http://www.genome.jp/kegg/>). These results from the metabolomics data modality were supported by the identification of GO terms from the transcriptomics data modality, which pointed out non-plant GO terms such as drug metabolic processes, and fungal and inorganic substance responses. These metabolic pathways and transcripts were not further investigated since neither plant-pathogen interactions nor the relationships between plants and endophytes were in the scope of this study. However, the presence of non-plant organisms could not just simply be ignored.

#### **4.7. Omics integration and pathway exploration led to the final identification and selection of target genes within the candidate pathways**

Using the results from the integration between metabolomics and transcriptomics data via MOFA (Argelaguet et al., 2018) and based on the literature, four metabolic pathways were highlighted as possible target candidates for the improvement of sugarcane cultivation in the field by pinpointing how developmental patterns are established during the early vegetative growth phase, and how these pathways change throughout development in two anatomically different tissues, the source and sink. As mentioned in section 4.1, it was concluded that the developmental profile of sugarcane is established within the initial months after planting, and long-term adjustments to plant growth and culm thickening take place. The thicker culm phenomenon of the “one-and-a-half-year” sugarcane plants was the first indication that secondary metabolism is affected when there is a shortage of water within the initial months after planting. A study done by subjecting soybean to drought stress during the early growth stage proved that a non-reversible phenotypic change occurs: the culms thicken, however at a cost (Ohashi et al., 2009). This observation corresponds to our study where “one-and-a-half-year” sugarcane presented thicker culms, however at a cost of a reduction in vegetative growth during the initial months. Numerous other studies on maize (Ohashi et al., 2009), wheat (Okuyama et al., 2005) and rice (Ferrer, 2015) also reported thicker culms when the plants were subjected to water limitations. These reports justify the selection of the four pathways (pentose phosphate pathway, phenylalanine, tyrosine and tryptophan biosynthesis, phenylpropanoid biosynthesis and flavonoid biosynthesis) for further inspection seeing that they are the main pathways implicated in culm formation, culm metabolism, and environmental response. Additionally, these four peculiar pathways were also chosen for the following reasons: the PPP is the main source of reducing power in the form of NADPH and metabolic intermediates, specifically sugar phosphates, which are the inputs for anabolic reactions in plants (Kruger & von Schaewen, 2003). In this pathway, carbon skeletons are produced for the biosynthesis of nucleotides, aromatic amino acids and importantly, phenylpropanoids, which led to the selection of two subsequent pathways namely phenylalanine, tyrosine and tryptophan biosynthesis (map00400), which are the main aromatic amino acids for the synthesis of various plant secondary metabolites (Tzin & Galili, 2010), and phenylpropanoid biosynthesis (map00940), the former flowing into the latter and thus leading to the production of lignin precursors. Intermediates of the phenylpropanoid biosynthetic pathway can be directed towards flavonoid biosynthesis (map00941), producing secondary metabolites that serve a multitude of functions such as serving an

antioxidant role (Agati et al., 2012; Baskar et al., 2018), and conferring drought tolerance in some plant species (Nakabayashi et al., 2014; Rao et al., 2020).

The results from the metabolic pathway activity analysis allowed for the generation of hypotheses, which is in fact the main aim of conducting untargeted studies such as the metabolomics and transcriptomics studies conducted in this study (Kell & Oliver, 2004; Quanbeck et al., 2012; Rattray et al., 2018). When inspecting the metabolism of the “one-and-a-half-year” sugarcane plants, the hypothesis is thus that the PPP and the phenylalanine, tyrosine and tryptophan biosynthesis pathways supply the energy, reducing equivalents and precursors for producing phenylpropanoids such as flavonoids that can counteract the effects of the water deficit experienced during the culm formation and vegetative growth phase, as well as contribute the building blocks needed for the thickening of the culms, however at the cost of retarded vegetative growth and thus establishing the developmental pattern of the “one-and-a-half-year” sugarcane plants. From the results from the “one-year” sugarcane, it was concluded that the PPP and flavonoid biosynthesis pathway activities are higher when the plants are experiencing active vegetative growth under favourable conditions. However, when water deficits are experienced during the active vegetative growth phase when most anatomical tissues are already formed, the activity of the biosynthesis of phenylalanine, tyrosine and tryptophan and the PPP are increased, and the maturation phase is initiated. Additionally, the flavonoid pathway was implicated in the initiation of the flowering process reported here as a collateral effect of the environmental conditions.

The gene expression profiles of the selected SAS were also considered in the hypothesis generating phase. To narrow targets down to the gene level, the specific transcripts were investigated. Interestingly, seven out of the nine target SAS were found to be representative of the phenylpropanoid pathway which is essentially responsible for the assimilation of approximately 30-40% of the organic carbon in the biosphere into the main component of cell walls, lignin (Zhang and Liu, 2015). Three of the nine were found to be representative of PAL. This enzyme is responsible for diverting carbon flux from primary to secondary metabolism (Fraser & Chapple, 2011; Zhang and Liu, 2015) and has been widely described as a regulator of plant development. From the results of the pathway activities obtained using PAPI, it was concluded that higher pathway activities of the phenylalanine, tyrosine and tryptophan biosynthesis; phenylpropanoid biosynthesis; and flavonoid biosynthesis pathways corresponded to the formation of thicker culm tissue. This is most likely due to the activity of the PAL gene seeing that it is the main enzyme that catalyzes the first step in the phenylpropanoid pathway, and thus serving as a link between primary and secondary metabolism (Barros et al., 2016; Souza et al., 2019). Transcripts related to this gene were expressed at high intensities when culm formation and lignification was occurring. A previous study aimed at investigating the role of PAL in *Brachypodium distachyon* cell wall metabolism, reported a 43% reduction in cell wall lignin contents as well as growth reductions (Cass et al., 2015), thus supporting our results. When all of the anatomical tissues of the sugarcane plants were already formed, the high expression of the PAL transcripts observed under water deficit could support its role in sugarcane’s response to changes in the external environments regarding precipitation and temperature. Another interesting finding was the expression of the CAD transcript, which presented similar results as that of PAL. It is an enzyme catalyzing the final step in the production of monolignols before their transport for the lignification step (Ma, 2010; Sibout et al., 2005). The expression of this transcript corresponded to the pathway activities of the previously mentioned pathways responsible for producing and lignifying the thicker sugarcane culms that are produced under water limitations during the early vegetative growth phase. Additionally, as the plants reached maturity, the expression of this transcript significantly increased, thus corresponding to previous studies reporting lignification during maturation. Transgenic studies in grasses such as switchgrass *Panicum virgatum* (Fu et al., 2011; Saathoff et al., 2011), maize (Fornalé et al., 2012) and *Festuca arundinacea* (Chen et al., 2003) aimed at the downregulation of CAD has proven not to directly affect the

lignin content, rather changing the lignin composition and making it more digestible during secondary ethanol production from sugarcane, highlighting the importance of targeting CAD in sugarcane. Both PAL and CAD were both implicated in the flowering phase (Figures 15 and 16). The overexpression of CAD in maize and *B. distachyon* led to increased stem weights and altered flowering times (Trabucco et al., 2013; Vermerris & McIntyre, 1999), highlighting a target for the possible control of unwanted flowering in sugarcane.

Furthermore, three out of the nine selected SAS represented a peroxidase enzyme (PER) which leads to the irreversible production of p-hydroxyphenyl lignin, guaiacyl lignin, 5-hydroxy-guaiacyl lignin and syringyl lignin (Fagerstedt et al., 2010; Morreel et al., 2010). The expression of the transcript related to PER was highly expressed under water limitations and when thick culm tissues were formed, and low expressions were reported in the mature internodes when flowering was initiated. A study done to correlate the sugarcane laccase gene, the main type of peroxidase gene in sugarcane studies, to the lignification process led to the generation of a (SoFLAC) co-expression network that showed that this gene was co-expressed with phenylpropanoid biosynthesis genes (Cesarino et al., 2013), supporting the present results. Additionally, lignin contents in mutant *Arabidopsis* plants could be restored by expressing the SoFLAC gene, however the compositions could not be restored (Cesarino et al., 2013), thus highlighting another possible target gene for altering sugarcane development and culm formation.

One SAS, a key enzyme of the flavonoid biosynthetic pathway, flavanone 3-dioxygenase (F3H) was one of the ones in our analyses, however little is known about the roles of the key enzymes in the flavonoid biosynthetic pathway in response to environmental alterations. The transcript representing F3H in our study was significantly higher expressed in the leaves of “one-and-a-half-year” sugarcane when the activity of the flavonoid biosynthesis pathway increased when lower precipitations were reported, thus adding to the robustness of the results obtained in this study. Interestingly, the expression of this transcript was significantly higher when there were drops in precipitation and temperature measurements, implicating this gene in water deficit response. In the literature, it has been described that expression of this gene increases under drought stress in the desert plant *Reaumuria soongorica* (Liu et al., 2013), in potatoes (André et al., 2009) and in transgenic tobacco plants (Song et al., 2016), thus supporting our results and highlighting F3H as a possible target gene for responding to initial water limitations in sugarcane leaves.

These results and descriptions of the main genes support the hypothesis that the four pathways selected for evaluation (Table 8) are the main regulators of sugarcane growth and development, specifically with regards to lignin precursor and eventual lignin production.



## 5. MAIN CONCLUSIONS

The main goal of this project was to implement a metabolomics approach in sugarcane to broaden the knowledge basis on this crop and to take a systems biology approach. We used a set of 108 samples collected from SP80-3280 plants grown in the, and we developed a pipeline for data integration between metabolomics and transcriptomics. The field experiments were conducted with two plantings, in April 2012 and October 2012, with the SP80-3280 plants grown for 13 months. Morphological and technological data evidenced that precipitation and temperature have significant influences on yield, and that the developmental profiles are established within the first four months of planting.

For the “one-and-a-half-year” sugarcane from the first field experiment, the observed retarded vegetative growth during the initial months after planting could be attributed to the lack of water availability since it has been proved that internode development and culm elongation are the most sensitive morphological alterations in sugarcane affected by water deficits. It was also concluded that SP80-3280 can support early phase water deficit by continuing photosynthate production and carbon assimilation, however the carbon flow is directed towards culm thickening instead of culm elongation or sugar accumulation. This is known as a long-term adjustment to the changes in the external environment. By planting SP80-3280 under low precipitation and temperatures, more biomass will be produced in the form of fiber over a longer period of time, thus highlighting optimal conditions for the cultivation of SP80-3280 when fiber accumulation is desired. In contrast, when sucrose accumulation is desired, rapid early-stage culm development and elongation are necessary, followed by a period of moderate water deficit in order to shift from the vegetative growth to the maturation phase, thus highlighting optimal conditions for the cultivation of SP80-3280 when sucrose accumulation is desired. These shifts between the different developmental stages are dependent on signalling cascades occurring due to interactions between the plants’ interior and exterior environments.

The untargeted metabolomics pipeline developed and described here, proved to be successful for the analysis of the complex metabolome of SP80-3280. We identified 90, 84, 79 and 54 metabolites in the leaf +1, upper internodes, young internodes and mature internodes of SP80-3280, which were assigned to 67, 70, 69 and 64 activated KEGG metabolic pathways. The high number and the chemical diversity of the metabolites detected here contributed to knowledge gaps regarding the relatively unexplored sugarcane metabolome and presented an alternative, improved approach for conducting untargeted metabolomics studies for sugarcane plants and possibly other crops. It also highlights the necessary steps, such as sample preparation, HPLC-MS analytics, metabolomics data processing and preparations, and the importance of data integration for answering biological questions.

We showed that the pipeline can be used for successful tissue profiling and highlighted economically and agronomically valuable compounds that are abundant in specific sugarcane tissue. Specifically, we detected secondary metabolism compounds such as phenolic compounds and glucosinolates, which were higher in the leaves and the mature culm tissue, mostly ones that serve roles in electron transport and structural support. We discuss compounds that have been studied for their economic value for having biopesticide, allelopathic, antioxidant, pharmaceutical and flavour enhancing properties. In the sugar and ethanol production industries, the leaves and remaining bagasse after the sucrose extraction processes are not currently being used to their full potential, thus the metabolites highlighted in these tissues here and their economic importance might aid in the decision to explore the opportunities of the improved utilization of these tissues. One possibility would be to increase production in the specific tissue by producing sugarcane under different conditions, by controlling the outcome of the developmental profiles as was seen in this study, or by engineering pathways.

Overlaps between the metabolomics and transcriptomics data were observed when they were analysed separately via data dimensionality-reduction techniques, as well as when they were directly integrated in an unsupervised manner using the MOFA tool. In our study, it could be concluded that highlighting target pathways (metabolites and genes), using a systems biology approach with data integration, and developing different bioinformatics strategies for the statistical analyses and integration of these types of data, is essential. For instance, when we combined metabolomics and transcriptomics data, we were able to highlight metabolites and key genes from the pentose phosphate pathway, the phenylalanine, tyrosine and tryptophan biosynthesis pathway, the phenylpropanoid biosynthesis pathway. It was concluded that the metabolomics data modality was essential for our molecular analyses since it could point out the exact metabolic pathways and the respective metabolites in each pathway that were altered in the different tissues throughout development and in response to changes in the external environment. Using Dhurrin as an example, when trying to determine if the genes of a specific biosynthetic pathway is present in an organism, the detection of the final products and/or specific metabolic intermediates in those pathways can serve as a type of validation of the presence of the pathway in the organism, assuming the presence of the genes.

For the different tissues of SP8032-80 analysed here, the main metabolic systems that were altered throughout development and in response to changes in the external environments were amino acid metabolism, mostly phenylalanine, tyrosine and tryptophan metabolism; and secondary metabolism, mostly phenylpropanoid metabolism and the metabolism of its upstream precursors and/or downstream products. Since the main phenotypic differences of the plants from the two different fields from our study were related to plant height and culm diameter, the reported findings regarding the pathway activities conclude that the phenylpropanoid biosynthetic pathway and its derivative pathway, flavonoid biosynthesis, are crucial to plant development and adaptation to changes in the plant's internal and/or external environments. It is suspected that the PPP and the phenylalanine, tyrosine and tryptophan biosynthesis pathway are the key pathways which will provide energy and intermediates for the proper functioning of the above-mentioned pathways. These pathways may be altered via the targeting of the candidate genes identified in this study (PER, CAD, PAL, F3H).

## REFERENCES

- Abdel-Farid, I. B., Sheded, M. G., & Mohamed, E. A. (2014). Metabolomic profiling and antioxidant activity of some Acacia species. *Saudi Journal of Biological Sciences*, *21*(5), 400–408.
- Agati, G., Azzarello, E., Pollastri, S., & Tattini, M. (2012). Flavonoids as antioxidants in plants: Location and functional significance. In *Plant Science* (Vol. 196, pp. 67–76). <https://doi.org/10.1016/j.plantsci.2012.07.014>
- Aggio, R. B. M., Ruggiero, K., & Villas-Bôas, S. G. (2010). Pathway Activity Profiling (PAPi): from the metabolite profile to the metabolic pathway activity. *Bioinformatics*, *26*(23), 2969–2976.
- Agrawal, P. K., Babu, B. K., & Saini, N. (2015). Omics of Model Plants. In D. Barh, M. S. Khan, & E. Davies (Eds.), *PlantOmics: The Omics of Plant Science* (pp. 1–32). Springer India.
- Akazawa, T., Miljanich, P., & Conn, E. E. (1960). Studies on Cyanogenic Glycoside of Sorghum Vulgare. *Plant Physiology*, *35*(4), 535–538.
- Albóniga, O. E., González, O., Alonso, R. M., Xu, Y., & Goodacre, R. (2020). Optimization of XCMS parameters for LC-MS metabolomics: an assessment of automated versus manual tuning and its effect on the final results. *Metabolomics: Official Journal of the Metabolomic Society*, *16*(1), 14.
- Alexa, A., & Rahnenfuhrer, J. (2019). topGO: enrichment analysis for gene ontology, version v. 2.3. 6. *R Package*.
- Alexander, A. G., & Others. (1973). *Sugarcane physiology, a comprehensive study of the Saccharum source-to-sink system*. Elsevier Scientific Publishing Co.
- Alexander, A. G., & Samuels, G. (1968). Controlled-Temperature Studies of Growth, Enzymology, and Sucrose Production by Two Sugarcane Varieties in Puerto Rico. *The Journal of Agriculture of the University of Puerto Rico*, *52*(3), 204–217.
- Alhamdoosh, M., Ng, M., Wilson, N. J., Sheridan, J. M., Huynh, H., Wilson, M. J., & Ritchie, M. E. (2017). Combining multiple tools outperforms individual methods in gene set enrichment analyses. *Bioinformatics*, *33*(3), 414–424.
- Allwood, J. W., Ellis, D. I., & Goodacre, R. (2008). Metabolomic technologies and their application to the study of plants and plant-host interactions. *Physiologia Plantarum*, *132*(2), 117–135.
- Alonso, A. P., Vigeolas, H., Raymond, P., Rolin, D., & Dieuaide-Noubhani, M. (2005). A new substrate cycle in plants. Evidence for a high glucose-phosphate-to-glucose turnover from in vivo steady-state and pulse-labeling experiments with [<sup>13</sup>C]glucose and [<sup>14</sup>C]glucose. *Plant Physiology*, *138*(4), 2220–2232.
- Ambavaram, M. M. R., Krishnan, A., Trijatmiko, K. R., & Pereira, A. (2011). Coordinated activation of cellulose and repression of lignin biosynthesis pathways in rice. *Plant Physiology*, *155*(2), 916–931.
- Ameziane, R., Bernhard, K., & Lightfoot, D. (2000). Expression of the bacterial *gdhA* gene encoding a NADPH glutamate dehydrogenase in tobacco affects plant growth and development. *Plant and Soil*, *221*(1), 47–57.
- André, C. M., Schafleitner, R., Legay, S., Lefèvre, I., Aliaga, C. A. A., Nomberto, G., Hoffmann, L., Hausman, J.-F., Larondelle, Y., & Evers, D. (2009). Gene expression changes related to the production of phenolic compounds in potato tubers grown under drought stress. *Phytochemistry*, *70*(9), 1107–1116.
- Arday, R. E. (2003). *Liquid Chromatography - Mass Spectrometry: An Introduction*. John Wiley & Sons.
- Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J. C., Buettner, F., Huber, W., & Stegle, O. (2018). Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Molecular Systems Biology*, *14*(6). <https://www.embopress.org/doi/abs/10.15252/msb.20178124>
- Avato, P., D'Addabbo, T., Leonetti, P., & Argentieri, M. P. (2013). Nematicidal potential of Brassicaceae. *Phytochemistry Reviews*, *12*(4), 791–802.

- Avonce, N., Mendoza-Vargas, A., Morett, E., & Iturriaga, G. (2006). Insights on the evolution of trehalose biosynthesis. *BMC Evolutionary Biology*, *6*, 109.
- Badenes-Perez, F. R., Gershenzon, J., & Heckel, D. G. (2014). Insect attraction versus plant defense: young leaves high in glucosinolates stimulate oviposition by a specialist herbivore despite poor larval survival due to high saponin content. *PLoS One*, *9*(4), e95766.
- Bae, H., Herman, E., & Sicher, R. (2005). Exogenous trehalose promotes non-structural carbohydrate accumulation and induces chemical detoxification and stress response proteins in *Arabidopsis thaliana* grown in liquid culture. In *Plant Science* (Vol. 168, Issue 5, pp. 1293–1301). <https://doi.org/10.1016/j.plantsci.2005.01.006>
- Barra, L., Pica, N., Gouffi, K., Walker, G. C., Blanco, C., & Trautwetter, A. (2003). Glucose 6-phosphate dehydrogenase is required for sucrose and trehalose to be efficient osmoprotectants in *Sinorhizobium meliloti*. *FEMS Microbiology Letters*, *229*(2), 183–188.
- Barros, J., Serrani-Yarce, J. C., Chen, F., Baxter, D., Venables, B. J., & Dixon, R. A. (2016). Role of bifunctional ammonia-lyase in grass cell wall biosynthesis. *Nature Plants*, *2*(6), 16050.
- Baskar, V., Venkatesh, R., & Ramalingam, S. (2018). Flavonoids (Antioxidants Systems) in Higher Plants and Their Response to Stresses. In *Antioxidants and Antioxidant Enzymes in Higher Plants* (pp. 253–268). [https://doi.org/10.1007/978-3-319-75088-0\\_12](https://doi.org/10.1007/978-3-319-75088-0_12)
- Bassard, J.-E., Ullmann, P., Bernier, F., & Werck-Reichhart, D. (2010). Phenolamides: bridging polyamines to the phenolic metabolism. *Phytochemistry*, *71*(16), 1808–1824.
- Batista-Silva, W., Heinemann, B., Rugen, N., Nunes-Nesi, A., Araújo, W. L., Braun, H.-P., & Hildebrandt, T. M. (2019). The role of amino acid metabolism during abiotic stress release. *Plant, Cell & Environment*, *42*(5), 1630–1644.
- Bell, J. M., Schwartz, R., McInnes, K. J., Howell, T., & Morgan, C. L. S. (2018). Deficit irrigation effects on yield and yield components of grain sorghum. In *Agricultural Water Management* (Vol. 203, pp. 289–296). <https://doi.org/10.1016/j.agwat.2018.03.002>
- Bergman, M., Perelman, A., Dubinsky, Z., & Grossman, S. (2003). Scavenging of reactive oxygen species by a novel glucurinated flavonoid antioxidant isolated and purified from spinach. *Phytochemistry*, *62*(5), 753–762.
- Bhatla, S. C. (2018). Plant Growth Regulators: An Overview. In *Plant Physiology, Development and Metabolism* (pp. 559–568). [https://doi.org/10.1007/978-981-13-2023-1\\_14](https://doi.org/10.1007/978-981-13-2023-1_14)
- Bhatla, S. C., & Lal, M. A. (2018). *Plant Physiology, Development and Metabolism*. <https://doi.org/10.1007/978-981-13-2023-1>
- Biala, W., & Jasiński, M. (2018). The Phenylpropanoid Case – It Is Transport That Matters. In *Frontiers in Plant Science* (Vol. 9). <https://doi.org/10.3389/fpls.2018.01610>
- Bijelic, A., Pretzler, M., Molitor, C., Zekiri, F., & Rompel, A. (2015). The Structure of a Plant Tyrosinase from Walnut Leaves Reveals the Importance of “Substrate-Guiding Residues” for Enzymatic Specificity. *Angewandte Chemie*, *54*(49), 14677–14680.
- BIOEN-WebApps*. (n.d.). Retrieved April 3, 2020, from <http://sucest-fun.org/wsapp/>
- Blum, A. (2018). Plant Breeding and Yield Stability. In *Plant Breeding for Stress Environments* (pp. 15–42). <https://doi.org/10.1201/9781351075718-2>
- Boller, T. (2018). Ethylene in pathogenesis and disease resistance. *The Plant Hormone Ethylene*. <https://www.taylorfrancis.com/books/e/9781351075763/chapters/10.1201/9781351075763-16>

- Bonnett, G. D., Hewitt, M. L., & Glassop, D. (2006). Effects of high temperature on the growth and composition of sugarcane internodes. In *Australian Journal of Agricultural Research* (Vol. 57, Issue 10, p. 1087). <https://doi.org/10.1071/ar06042>
- Borevitz, J. O., Xia, Y., Blount, J., Dixon, R. A., & Lamb, C. (2000). Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *The Plant Cell*, *12*(12), 2383–2394.
- Borges, A., Abreu, A. C., Ferreira, C., Saavedra, M. J., Simões, L. C., & Simões, M. (2015). Antibacterial activity and mode of action of selected glucosinolate hydrolysis products against bacterial pathogens. *Journal of Food Science and Technology*, *52*(8), 4737–4748.
- Bosch, S. (2005). *Trehalose and carbon partitioning in sugarcane* [Stellenbosch: University of Stellenbosch]. <http://scholar.sun.ac.za/handle/10019.1/1433>
- Bosch, S., Rohwer, J. M., & Botha, F. C. (2003). The sugarcane metabolome. *Proc S Afr Sug Technol Ass*, *7*, 129–133.
- Botha, F. C. (2007). Precision breeding to improve the usefulness of sugarcane. *Proc Int Soc Sugar Cane Technol*, *26*, 35–43.
- Bottcher, A., Cesarino, I., Santos, A. B. dos, Vicentini, R., Mayer, J. L. S., Vanholme, R., Morreel, K., Goeminne, G., Moura, J. C. M. S., Nobile, P. M., Carmello-Guerreiro, S. M., Anjos, I. A. dos, Creste, S., Boerjan, W., Landell, M. G. de A., & Mazzafera, P. (2013). Lignification in sugarcane: biochemical characterization, gene discovery, and expression analysis in two genotypes contrasting for lignin content. *Plant Physiology*, *163*(4), 1539–1557.
- Botton, A., & Ruperti, B. (2019). The Yes and No of the Ethylene Involvement in Abscission. *Plants*, *8*(6), 187.
- Bové, J. M., & Garnier, M. (2002). Phloem-and xylem-restricted plant pathogenic bacteria. *Plant Science: An International Journal of Experimental Plant Biology*, *163*(6), 1083–1098.
- Bowne, J. B., Erwin, T. A., Juttner, J., Schnurbusch, T., Langridge, P., Bacic, A., & Roessner, U. (2012). Drought responses of leaf tissues from wheat cultivars of differing drought tolerance at the metabolite level. *Molecular Plant*, *5*(2), 418–429.
- Broekaert, W. F., Delauré, S. L., De Bolle, M. F. C., & Cammue, B. P. A. (2006). The role of ethylene in host-pathogen interactions. *Annual Review of Phytopathology*, *44*, 393–416.
- Cai, Q., Zhang, J., Guo, C., and Al, E. (2006). Reviews of the physiological roles and molecular biology of polyamines in higher plants. *J. Fujian Educ. Coll.* *7*, 118–124. doi: 10.3969/j.issn.1673-9884.2006.10.039
- Caldana, C., Fernie, A. R., Willmitzer, L., & Steinhauser, D. (2012). Unraveling retrograde signaling pathways: finding candidate signaling molecules via metabolomics and systems biology driven approaches. In *Frontiers in Plant Science* (Vol. 3). <https://doi.org/10.3389/fpls.2012.00267>
- Canzler, S., Schor, J., Busch, W., Schubert, K., Rolle-Kampczyk, U. E., Seitz, H., Kamp, H., von Bergen, M., Buesen, R., & Hackermüller, J. (2020). Prospects and challenges of multi-omics data integration in toxicology. *Archives of Toxicology*, *94*(2), 371–388.
- Cardozo, N. P. (2012). Modeling sugarcane ripening as function of meteorological variables. *Journal of Science and Agricultural Meteorology*.
- Cardozo, N. P., & Sentelhas, P. C. (2013). Climatic effects on sugarcane ripening under the influence of cultivars and crop age. In *Scientia Agricola* (Vol. 70, Issue 6, pp. 449–456). <https://doi.org/10.1590/s0103-90162013000600011>
- Cardozo, N. P., Sentelhas, P. C., Panosso, A. R., Palhares, A. L., & Ide, B. Y. (2015). Modeling sugarcane ripening as a function of accumulated rainfall in Southern Brazil. *International Journal of Biometeorology*, *59*(12), 1913–1925.
- Carr, M. K. V., & Knox, J. W. (2011). THE water relations and irrigation requirements of sugar cane (*Saccharum officinarum*): a review. *Experimental Agriculture*, *47*(1), 1–25.

- Cass, C. L., Peraldi, A., Dowd, P. F., Mottiar, Y., Santoro, N., Karlen, S. D., Bukhman, Y. V., Foster, C. E., Thrower, N., Bruno, L. C., Moskvina, O. V., Johnson, E. T., Willhoit, M. E., Phutane, M., Ralph, J., Mansfield, S. D., Nicholson, P., & Sedbrook, J. C. (2015). Effects of PHENYLALANINE AMMONIA LYASE (PAL) knockdown on cell wall composition, biomass digestibility, and biotic and abiotic stress responses in *Brachypodium*. *Journal of Experimental Botany*, *66*(14), 4317–4335.
- Casu, R. E., Grof, C. P. L., 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., 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.
- Cesarino, I., Araújo, P., Sampaio Mayer, J. L., Vicentini, R., Berthet, S., Demedts, B., Vanholme, B., Boerjan, W., & Mazzafera, P. (2013). Expression of SoFLAC, a new laccase in sugarcane, restores lignin content but not S:G ratio of *Arabidopsis lac17* mutant. *Journal of Experimental Botany*, *64*(6), 1769–1781.
- Cheavegatti-Gianotto, A., de Abreu, H. M. C., Arruda, P., Bespalhok Filho, J. C., Burnquist, W. L., Creste, S., di Ciero, L., Ferro, J. A., de Oliveira Figueira, A. V., de Sousa Filgueiras, T., Grossi-de-Sá, M. de F., Guzzo, E. C., Hoffmann, H. P., de Andrade Landell, M. G., Macedo, N., Matsuoka, S., de Castro Reinach, F., Romano, E., da Silva, W. J., ... César Ulian, E. (2011). Sugarcane (*Saccharum X officinarum*): A Reference Study for the Regulation of Genetically Modified Cultivars in Brazil. *Tropical Plant Biology*, *4*(1), 62–89.
- Chen, D., Shao, Q., Yin, L., Younis, A., & Zheng, B. (2018). Polyamine Function in Plants: Metabolism, Regulation on Development, and Roles in Abiotic Stress Responses. *Frontiers in Plant Science*, *9*, 1945.
- Chen, L., Auh, C.-K., Dowling, P., Bell, J., Chen, F., Hopkins, A., Dixon, R. A., & Wang, Z.-Y. (2003). Improved forage digestibility of tall fescue (*Festuca arundinacea*) by transgenic down-regulation of cinnamyl alcohol dehydrogenase. *Plant Biotechnology Journal*, *1*(6), 437–449.
- Chen, M., Song, F., Guo, M., Liu, Z., & Liu, S. (2002). Analysis of flavonoid constituents from leaves of *Acanthopanax senticosus* harms by electrospray tandem mass spectrometry. *Rapid Communications in Mass Spectrometry: RCM*, *16*(4), 264–271.
- Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., Wishart, D. S., & Xia, J. (2018). MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. In *Nucleic Acids Research* (Vol. 46, Issue W1, pp. W486–W494). <https://doi.org/10.1093/nar/gky310>
- Christianson, C. C., Johnson, C. J. L., & Needham, S. R. (2013). The advantages of microflow LC–MS/MS compared with conventional HPLC–MS/MS for the analysis of methotrexate from human plasma. *Bioanalysis*, *5*(11), 1387–1396.
- Claros Cuadrado, J. L., Pinillos, E. O., Tito, R., Mirones, C. S., & Gamarra Mendoza, N. N. (2019). Insecticidal Properties of Capsaicinoids and Glucosinolates Extracted from *Capsicum chinense* and *Tropaeolum tuberosum*. *Insects*, *10*(5). <https://doi.org/10.3390/insects10050132>
- Clements, H. F. (1962). *The ripening of sugar cane*.

- Coelho, R. D., Maschio, R., Leal, D. P. V., da Silva BARBOSA, F., & Mauri, R. (2012). Water Productivity into Biomass and Energy for 24 Brazilian Sugarcane Varieties Under Drip Irrigation. In *2012 Dallas, Texas, July 29 - August 1, 2012*. <https://doi.org/10.13031/2013.42042>
- Coleman, R. E., & Others. (1969). Physiology of flowering in sugarcane. *Physiology of Flowering in Sugarcane*. <https://www.cabdirect.org/cabdirect/abstract/19710302780>
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*(18), 3674–3676.
- Corea, O. R. A., Bedgar, D. L., Davin, L. B., & Lewis, N. G. (2012). The aroenate dehydratase gene family: towards understanding differential regulation of carbon flux through phenylalanine into primary versus secondary metabolic pathways. *Phytochemistry*, *82*, 22–37.
- Corpas, F. J., & Barroso, J. B. (2014). NADPH-generating dehydrogenases: their role in the mechanism of protection against nitro-oxidative stress induced by adverse environmental conditions. *Frontiers of Environmental Science & Engineering in China*, *2*. <https://doi.org/10.3389/fenvs.2014.00055>
- Cortez, L. A. B. (2014). *Sugarcane bioethanol — R&D for Productivity and Sustainability*. <https://doi.org/10.5151/blucheroa-sugarcane>
- Costa, C. C. da, da Costa, C. C., & Guilhoto, J. (n.d.). Expected Growth of Sugarcane Industry and Impact on the Brazilian Economy: 2015 and 2020. In *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2406005>
- Cruz de Carvalho, M. H. (2008). Drought stress and reactive oxygen species: Production, scavenging and signaling. *Plant Signaling & Behavior*, *3*(3), 156–165.
- Dal-Bianco, M., Carneiro, M. S., Hotta, C. T., Chapola, R. G., Hoffmann, H. P., Garcia, A. A. F., & Souza, G. M. (2012). Sugarcane improvement: how far can we go? *Current Opinion in Biotechnology*, *23*(2), 265–270.
- Dare, A. P., Yauk, Y.-K., Tomes, S., McGhie, T. K., Rebstock, R. S., Cooney, J. M., & Atkinson, R. G. (2017). Silencing a phloretin-specific glycosyltransferase perturbs both general phenylpropanoid biosynthesis and plant development. *The Plant Journal: For Cell and Molecular Biology*, *91*(2), 237–250.
- da Silva, F. M. A., Bataglion, G. A., de Almeida, R. A., Heerdt, G., Sousa, I. L., da Silva Filho, F. A., de Alencar, D. C., Costa, E. V., de Souza, A. D. L., Pinheiro, M. L. B., Morgon, N. H., & Koolen, H. H. F. (2017). Positive electrospray ionization ion trap mass spectrometry and ab initio computational studies of the multi-pathway fragmentation of oxoaporphine alkaloids. *International Journal of Mass Spectrometry*, *418*, 30–36.
- da Silva, V. de P. R., da Silva, B. B., Albuquerque, W. G., Borges, C. J. R., de Sousa, I. F., & Neto, J. D. (2013). Crop coefficient, water requirements, yield and water use efficiency of sugarcane growth in Brazil. *Agricultural Water Management*, *128*, 102–109.
- Davies, W. J. (n.d.). Responses of Plant Growth and Functioning to Changes in Water Supply in a Changing Climate. In *Plant Growth and Climate Change* (pp. 96–117). <https://doi.org/10.1002/9780470988695.ch5>
- Davin, L. B., & Lewis, N. G. (1992). Phenylpropanoid Metabolism: Biosynthesis of Monolignols, Lignans and Neolignans, Lignins and Suberins. In *Phenolic Metabolism in Plants* (pp. 325–375). [https://doi.org/10.1007/978-1-4615-3430-3\\_11](https://doi.org/10.1007/978-1-4615-3430-3_11)
- de Barros, M., Mota da Silva, L., Boeing, T., Somensi, L. B., Cury, B. J., de Moura Burci, L., Santin, J. R., de Andrade, S. F., Monache, F. D., & Cechinel-Filho, V. (2016). Pharmacological reports about gastroprotective effects of methanolic extract from leaves of *Solidago chilensis* (Brazilian arnica) and its components quercitrin and afzelin in rodents. *Naunyn-Schmiedeberg's Archives of Pharmacology*, *389*(4), 403–417.

- Debnath, M., Pandey, M., & Bisen, P. S. (2011). An omics approach to understand the plant abiotic stress. *Omic: A Journal of Integrative Biology*, 15(11), 739–762.
- de Cana-de, C. dos P. (2006). Açúcar, Açúcar e Álcool do Estado de São Paulo. Manual de Instruções/Edição. *CONSECANA-SP, Piracicaba-SP*.
- Delorge, I., Janiak, M., Carpentier, S., & Van Dijck, P. (2014). Fine tuning of trehalose biosynthesis and hydrolysis as novel tools for the generation of abiotic stress tolerant plants. *Frontiers in Plant Science*, 5, 147.
- de Medeiros Barbosa, V. F. A. (2015). Chapter 3 - Planting. In F. Santos, A. Borém, & C. Caldas (Eds.), *Sugarcane* (pp. 35–51). Academic Press.
- de Puit, M., Ismail, M., & Xu, X. (2014). LCMS analysis of fingerprints, the amino acid profile of 20 donors. *Journal of Forensic Sciences*, 59(2), 364–370.
- De Vos, R. C. H., Moco, S., Lommen, A., Keurentjes, J. J. B., Bino, R. J., & Hall, R. D. (2007). Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nature Protocols*, 2(4), 778–791.
- Dinkova-Kostova, A. T., & Kostov, R. V. (2012). Glucosinolates and isothiocyanates in health and disease. *Trends in Molecular Medicine*, 18(6), 337–347.
- Dixon, R. A., Lamb, C. J., Masoud, S., Sewalt, V. J. H., & Paiva, N. L. (1996). Metabolic engineering: prospects for crop improvement through the genetic manipulation of phenylpropanoid biosynthesis and defense responses—a review. *Gene*, 179(1), 61–71.
- Dixon, R. A., & Strack, D. (2003). Phytochemistry meets genome analysis, and beyond..... In *Phytochemistry* (Vol. 62, Issue 6, pp. 815–816). [https://doi.org/10.1016/s0031-9422\(02\)00712-4](https://doi.org/10.1016/s0031-9422(02)00712-4)
- Domingo-Almenara, X., & Siuzdak, G. (2020). Metabolomics Data Processing Using XCMS. *Methods in Molecular Biology*, 2104, 11–24.
- Dominy, C., Haynes, R., & van Antwerpen, R. (2002). Loss of soil organic matter and related soil properties under long-term sugarcane production on two contrasting soils. *Biology and Fertility of Soils*, 36(5), 350–356.
- dos Santos, A. B., Bottcher, A., Vicentini, R., Sampaio Mayer, J. L., Kiyota, E., Landell, M. A. G., Creste, S., & Mazzafera, P. (2015). Lignin biosynthesis in sugarcane is affected by low temperature. *Environmental and Experimental Botany*, 120, 31–42.
- Dunstan, W. R., & Henry, T. A. (1902). VII. Cyanogenesis in Plants.—Part II. The Great Millet, *Sorghum vulgare*. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 199(312-320), 399–410.
- Dvoráková, L., Srba, M., Opatrny, Z., & Fischer, L. (2012). Hybrid proline-rich proteins: novel players in plant cell elongation? *Annals of Botany*, 109(2), 453–462.
- Dwivedi, R. S. (2003). SUGARCANE YIELD PLATEAUS AND POTENTIAL FOR YIELD INCREASES CONDITIONS OF INDIA. *Life Science Advances. Plant Physiology: A Journal of the Council of Scientific Research Integration*, 5, 435.
- Earling, M., Beadle, T., & Niemeyer, E. D. (2019). Açai Berry (*Euterpe oleracea*) Dietary Supplements: Variations in Anthocyanin and Flavonoid Concentrations, Phenolic Contents, and Antioxidant Properties. *Plant Foods for Human Nutrition*, 74(3), 421–429.
- Ebrahim, M. K., Zingsheim, O., El-Shourbagy, M. N., Moore, P. H., & Komor, E. (1998). Growth and sugar storage in sugarcane grown at temperatures below and above optimum. *Journal of Plant Physiology*, 153(5), 593–602.
- Ecker, J. R., & Davis, R. W. (1987). Plant defense genes are regulated by ethylene. *Proceedings of the National Academy of Sciences of the United States of America*, 84(15), 5202–5206.

- El-Azaz, J., Cánovas, F. M., Ávila, C., & de la Torre, F. (2018). The arogenate dehydratase ADT2 is essential for seed development in Arabidopsis. In *Plant and Cell Physiology*. <https://doi.org/10.1093/pcp/pcy200>
- El-Gewely, M. R. (1998). *Biotechnology Annual Review*. Elsevier.
- Elkind, Y., Edwards, R., Mavandad, M., Hedrick, S. A., Ribak, O., Dixon, R. A., & Lamb, C. J. (1990). Abnormal plant development and down-regulation of phenylpropanoid biosynthesis in transgenic tobacco containing a heterologous phenylalanine ammonia-lyase gene. *Proceedings of the National Academy of Sciences of the United States of America*, 87(22), 9057–9061.
- Engels, F. (1998). Alfalfa Stem Tissues: Cell-wall Development and Lignification. In *Annals of Botany* (Vol. 82, Issue 5, pp. 561–568). <https://doi.org/10.1006/anbo.1998.0705>
- Fagerstedt, K. V., Kukkola, E. M., Koistinen, V. V. T., Takahashi, J., & Marjamaa, K. (2010). Cell wall lignin is polymerised by class III secretable plant peroxidases in Norway spruce. *Journal of Integrative Plant Biology*, 52(2), 186–194.
- Fahey, J. W., Zalcmann, A. T., & Talalay, P. (2001). The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry*, 56(1), 5–51.
- Fang, C., Fernie, A. R., & Luo, J. (2019). Exploring the Diversity of Plant Metabolism. *Trends in Plant Science*, 24(1), 83–98.
- Farag, M. A. (2014). Comparative mass spectrometry & nuclear magnetic resonance metabolomic approaches for nutraceuticals quality control analysis: a brief review. *Recent Patents on Biotechnology*, 8(1), 17–24.
- Faraji, M., Fonseca, L. L., Escamilla-Treviño, L., Barros-Rios, J., Engle, N. L., Yang, Z. K., Tschaplinski, T. J., Dixon, R. A., & Voit, E. O. (2018). A dynamic model of lignin biosynthesis in *Brachypodium distachyon*. *Biotechnology for Biofuels*, 11, 253.
- Ferrandino, A., & Lovisolo, C. (2014). Abiotic stress effects on grapevine (*Vitis vinifera* L.): Focus on abscisic acid-mediated consequences on secondary metabolism and berry quality. In *Environmental and Experimental Botany* (Vol. 103, pp. 138–147). <https://doi.org/10.1016/j.envexpbot.2013.10.012>
- Ferreira, D. A., Martins, M. C. M., Cheavegatti-Gianotto, A., Carneiro, M. S., Amadeu, R. R., Aricetti, J. A., Wolf, L. D., Hoffmann, H. P., de Abreu, L. G. F., & Caldana, C. (2018). Metabolite Profiles of Sugarcane Culm Reveal the Relationship Among Metabolism and Axillary Bud Outgrowth in Genetically Related Sugarcane Commercial Cultivars. *Frontiers in Plant Science*, 9, 857.
- Ferrer, M. C. (jul2015). *Morpho-agronomic characterization and evaluation for drought tolerance of 50 selected Philippine traditional rice (Oryza sativa L.) varieties*. <http://agris.fao.org/agris-search/search.do?recordID=PH2018000238>
- Fiehn, O. (2002). Metabolomics — the link between genotypes and phenotypes. In *Functional Genomics* (pp. 155–171). [https://doi.org/10.1007/978-94-010-0448-0\\_11](https://doi.org/10.1007/978-94-010-0448-0_11)
- Fiehn, O. (2016). Metabolomics by Gas Chromatography-Mass Spectrometry: Combined Targeted and Untargeted Profiling. *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et AL]*, 114, 30.4.1–30.4.32.
- Figueira, T. R. e. S., Okura, V., Rodrigues da Silva, F., Jose da Silva, M., Kudrna, D., Ammiraju, J. S. S., Talag, J., Wing, R., & Arruda, P. (2012). A BAC library of the SP80-3280 sugarcane variety (*saccharum* sp.) and its inferred microsynteny with the sorghum genome. *BMC Research Notes*, 5(1), 185.
- Figuerola, C. M., & Lunn, J. E. (2016). A Tale of Two Sugars: Trehalose 6-Phosphate and Sucrose. *Plant Physiology*, 172(1), 7–27.

- Fong Chong, B., Mills, E., Bonnett, G. D., & Gnanasambandam, A. (2010). Early Exposure to Ethylene Modifies Shoot Development and Increases Sucrose Accumulation Rate in Sugarcane. *Journal of Plant Growth Regulation*, 29(2), 149–163.
- Fornalé, S., Capellades, M., Encina, A., Wang, K., Irar, S., Lapierre, C., Ruel, K., Joseleau, J.-P., Berenguer, J., Puigdomènech, P., Rigau, J., & Caparrós-Ruiz, D. (2012). Altered Lignin Biosynthesis Improves Cellulosic Bioethanol Production in Transgenic Maize Plants Down-Regulated for Cinnamyl Alcohol Dehydrogenase. In *Molecular Plant* (Vol. 5, Issue 4, pp. 817–830). <https://doi.org/10.1093/mp/ssr097>
- Forsberg, E. M., Huan, T., Rinehart, D., Benton, H. P., Warth, B., Hilmers, B., & Siuzdak, G. (2018). Data processing, multi-omic pathway mapping, and metabolite activity analysis using XCMS Online. *Nature Protocols*, 13(4), 633–651.
- Fosket, D. E. (1994). Biotic Factors Regulate Some Aspects of Plant Development. In *Plant Growth and Development* (pp. 517–557). <https://doi.org/10.1016/b978-0-12-262430-8.50014-0>
- Fraser, C. M., & Chapple, C. (2011). The phenylpropanoid pathway in Arabidopsis. *The Arabidopsis Book / American Society of Plant Biologists*, 9, e0152.
- Fu, C., Xiao, X., Xi, Y., Ge, Y., Chen, F., Bouton, J., Dixon, R. A., & Wang, Z.-Y. (2011). Downregulation of Cinnamyl Alcohol Dehydrogenase (CAD) Leads to Improved Saccharification Efficiency in Switchgrass. In *BioEnergy Research* (Vol. 4, Issue 3, pp. 153–164). <https://doi.org/10.1007/s12155-010-9109-z>
- Funck, D., Winter, G., Baumgarten, L., & Forlani, G. (2012). Requirement of proline synthesis during Arabidopsis reproductive development. *BMC Plant Biology*, 12, 191.
- Games, D. E., & Lewis, E. (1980). Combined liquid chromatography mass spectrometry of glycosides, glucuronides, sugars and nucleosides. In *Biological Mass Spectrometry* (Vol. 7, Issue 10, pp. 433–436). <https://doi.org/10.1002/bms.1200071006>
- Gao, W., Sun, H.-X., Xiao, H., Cui, G., Hillwig, M. L., Jackson, A., Wang, X., Shen, Y., Zhao, N., Zhang, L., Wang, X.-J., Peters, R. J., & Huang, L. (2014). Combining metabolomics and transcriptomics to characterize tanshinone biosynthesis in *Salvia miltiorrhiza*. *BMC Genomics*, 15, 73.
- Gapper, N. E., Giovannoni, J. J., & Watkins, C. B. (2014). Understanding development and ripening of fruit crops in an “omics” era. *Horticulture Research*, 1, 14034.
- García-Villalba, R., Tomás-Barberán, F. A., Faça-Berthon, P., Roller, M., Zafrilla, P., Issaly, N., & García-Conesa, M.-T. (2015). Targeted and Untargeted Metabolomics to Explore the Bioavailability of the Secoiridoids from a Seed/Fruit Extract (*Fraxinus angustifolia* Vahl) in Human Healthy Volunteers: A Preliminary Study. *Molecules*, 20(12), 22202–22219.
- Gatehouse, A. M. R. (2004). Engineering of Crops for Improved Agronomic Traits. In *Handbook of Plant Biotechnology*. <https://doi.org/10.1002/0470869143.kc018>
- Getnet, B. (2017). *Review on in vitro propagation of sugarcane to advance the value of tissue culture*. <https://pdfs.semanticscholar.org/221a/6f7fc1b6fe5f6ecdb0f990719a11c3d103ca.pdf>
- Gholami, M., Fakhari, A. R., & Ghanati, F. (2013). Selective regulation of nicotine and polyamines biosynthesis in tobacco cells by enantiomers of ornithine. *Chirality*, 25(1), 22–27.
- Giannakidis, G., Karlsson, K., Labriet, M., & Gallachóir, B. Ó. (2018). *Limiting Global Warming to Well Below 2 °C: Energy System Modelling and Policy Development*. Springer.
- Glassop, D., Roessner, U., Bacic, A., & Bonnett, G. D. (2007). Changes in the sugarcane metabolome with stem development. Are they related to sucrose accumulation? *Plant & Cell Physiology*, 48(4), 573–584.

- Glasziou, K. T., Bull, T. A., Hatch, M. D., & Whiteman, P. C. (1965). Physiology of Sugar-Cane VII. Effects of Temperature, Photoperiod Duration, and Diurnal and Seasonal Temperature Changes on Growth and Ripening. *Australian Journal of Biological Sciences*, *18*(1), 53–66.
- Goldemberg, J. (2007). Ethanol for a sustainable energy future. *Science*, *315*(5813), 808–810.
- Goldemberg, J., Coelho, S., Nastari, P., & Guardabassi, P. (2014). Production and Supply Logistics of Sugarcane as an Energy Feedstock. In *Sustainable Bioenergy Production* (pp. 213–222). <https://doi.org/10.1201/b16764-12>
- Gómez, L. D., Gilday, A., Feil, R., Lunn, J. E., & Graham, I. A. (2010). AtTPS1-mediated trehalose 6-phosphate synthesis is essential for embryogenic and vegetative growth and responsiveness to ABA in germinating seeds and stomatal guard cells. *The Plant Journal: For Cell and Molecular Biology*, *64*(1), 1–13.
- Gonzalez, N., De Bodt, S., Sulpice, R., Jikumaru, Y., Chae, E., Dhondt, S., Van Daele, T., De Milde, L., Weigel, D., Kamiya, Y., Stitt, M., Beemster, G. T. S., & Inzé, D. (2010). Increased leaf size: different means to an end. *Plant Physiology*, *153*(3), 1261–1279.
- Goujon, T., Sibout, R., Pollet, B., Maba, B., Nussaume, L., Bechtold, N., Lu, F., Ralph, J., Mila, I., Barrière, Y., Lapierre, C., & Jouanin, L. (2003). A new *Arabidopsis thaliana* mutant deficient in the expression of O-methyltransferase impacts lignins and sinapoyl esters. *Plant Molecular Biology*, *51*(6), 973–989.
- Grata, E., Boccard, J., Guillaume, D., Glauser, G., Carrupt, P.-A., Farmer, E. E., Wolfender, J.-L., & Rudaz, S. (2008). UPLC–TOF-MS for plant metabolomics: A sequential approach for wound marker analysis in *Arabidopsis thaliana*. In *Journal of Chromatography B* (Vol. 871, Issue 2, pp. 261–270). <https://doi.org/10.1016/j.jchromb.2008.04.021>
- Greenland, D. (2005). Climate Variability and Sugarcane Yield in Louisiana. *Journal of Applied Meteorology*, *44*(11), 1655–1666.
- Griffiths, W. J., Koal, T., Wang, Y., Kohl, M., Enot, D. P., & Deigner, H.-P. (2010). Targeted metabolomics for biomarker discovery. *Angewandte Chemie*, *49*(32), 5426–5445.
- Grimplet, J., Deluc, L. G., Tillett, R. L., Wheatley, M. D., Schlauch, K. A., Cramer, G. R., & Cushman, J. C. (2007). Tissue-specific mRNA expression profiling in grape berry tissues. *BMC Genomics*, *8*, 187.
- Grivet, L., & Arruda, P. (2002). Sugarcane genomics: depicting the complex genome of an important tropical crop. *Current Opinion in Plant Biology*, *5*(2), 122–127.
- Gromski, P. S., Xu, Y., Hollywood, K. A., Turner, M. L., & Goodacre, R. (2015). The influence of scaling metabolomics data on model classification accuracy. *Metabolomics: Official Journal of the Metabolomic Society*, *11*(3), 684–695.
- Guertzoni, J. T. S., Belintani, N. G., Moreira, R. M. P., Hoshino, A. A., Domingues, D. S., Filho, J. C. B., & Vieira, L. G. E. (2014). Stress-induced  $\Delta$ 1-pyrroline-5-carboxylate synthetase (P5CS) gene confers tolerance to salt stress in transgenic sugarcane. *Acta Physiologiae Plantarum / Polish Academy of Sciences, Committee of Plant Physiology Genetics and Breeding*, *36*(9), 2309–2319.
- Gupta, B., Sengupta, A., Saha, J., & Gupta, K. (2013). Plant abiotic stress: “Omics” approach. *J. Plant Biochem. Physiol*, *1*(3). [https://www.researchgate.net/profile/Bhaskar\\_Gupta5/publication/261174969\\_Plant\\_Abiotic\\_Stress\\_OMICS\\_Approach/links/0a85e5376628754c30000000/Plant-Abiotic-Stress-OMICS-Approach.pdf](https://www.researchgate.net/profile/Bhaskar_Gupta5/publication/261174969_Plant_Abiotic_Stress_OMICS_Approach/links/0a85e5376628754c30000000/Plant-Abiotic-Stress-OMICS-Approach.pdf)
- Gupta, D., & Singh, A. (2013). Metabolomics and its Role in Study of Plant Abiotic Stress Responses. In *Molecular Approaches in Plant Abiotic Stress* (pp. 130–140). <https://doi.org/10.1201/b15538-9>
- Gupta, R. B., & Demirbas, A. (n.d.). Air Pollution and Global Warming from the Use of Fossil Fuels. In *Gasoline, Diesel, and Ethanol Biofuels from Grasses and Plants* (pp. 25–40). <https://doi.org/10.1017/cbo9780511779152.003>

- Gupta, S. D. (2010). Sites of generation and physicochemical basis of formation of reactive oxygen species in plant cell. *Reactive Oxygen Species and Antioxidants in*. <https://content.taylorfrancis.com/books/e/download?dac=C2010-0-39943-X&isbn=9781439854082&doi=10.1201/9781439854082-5&format=pdf>
- Gussin, A. E. S. (1972). Does trehalose occur in angiospermae? *Phytochemistry*, *11*(5), 1827–1828.
- Hahlbrock, K., & Grisebach, H. (1979). Enzymic Controls in the Biosynthesis of Lignin and Flavonoids. *Annual Review of Plant Physiology*, *30*(1), 105–130.
- Hahlbrock, K., & Scheel, D. (1989). Physiology and Molecular Biology of Phenylpropanoid Metabolism. *Annual Review of Plant Physiology and Plant Molecular Biology*, *40*(1), 347–369.
- Halket, J. M., Waterman, D., Przyborowska, A. M., Patel, R. K. P., Fraser, P. D., & Bramley, P. M. (2005). Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *Journal of Experimental Botany*, *56*(410), 219–243.
- Halkier, B. A., & Gershenzon, J. (2006). Biology and biochemistry of glucosinolates. *Annual Review of Plant Biology*, *57*, 303–333.
- Hall, R., Beale, M., Fiehn, O., Hardy, N., Sumner, L., & Bino, R. (2002). Plant metabolomics: the missing link in functional genomics strategies. *The Plant Cell*, *14*(7), 1437–1440.
- Hanusz, Z., & Tarasińska, J. (2015). Normalization of the Kolmogorov–Smirnov and Shapiro–Wilk tests of normality. In *Biometrical Letters* (Vol. 52, Issue 2, pp. 85–93). <https://doi.org/10.1515/bile-2015-0008>
- Harder, U., Koletzko, B., & Peissner, W. (2011). Quantification of 22 plasma amino acids combining derivatization and ion-pair LC–MS/MS. In *Journal of Chromatography B* (Vol. 879, Issues 7-8, pp. 495–504). <https://doi.org/10.1016/j.jchromb.2011.01.010>
- Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., Richter, J., Rubin, G. M., Blake, J. A., Bult, C., Dolan, M., Drabkin, H., Eppig, J. T., Hill, D. P., Ni, L., ... Gene Ontology Consortium. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, *32*(Database issue), D258–D261.
- Hayat, R., Ali, S., Amara, U., Khalid, R., & Ahmed, I. (2010). Soil beneficial bacteria and their role in plant growth promotion: a review. *Annals of Microbiology*, *60*(4), 579–598.
- He, F., Li, D., Wang, D., & Deng, M. (2016). Extraction and Purification of Quercitrin, Hyperoside, Rutin, and Afzelin from *Zanthoxylum Bungeanum* Maxim Leaves Using an Aqueous Two-Phase System. *Journal of Food Science*, *81*(7), C1593–C1602.
- Hefner, R. (2007). Warren S. Torgerson, Theory and methods of scaling. New York: John Wiley and Sons, Inc., 1958. Pp. 460. *Systems Research: The Official Journal of the International Federation for Systems Research*, *4*(3), 245–247.
- Hemaprabha, G., Nagarajan, R., & Alarmelu, S. (2004). Response of sugarcane genotypes to water deficit stress. *Sugar Tech*, *6*(3), 165–168.
- Hennig, L. (2007). Patterns of beauty--omics meets plant development. *Trends in Plant Science*, *12*(7), 287–293.
- He, P., & Aga, D. S. (2019). Comparison of GC-MS/MS and LC-MS/MS for the analysis of hormones and pesticides in surface waters: advantages and pitfalls. In *Analytical Methods* (Vol. 11, Issue 11, pp. 1436–1448). <https://doi.org/10.1039/c8ay02774a>
- Herrmann, A., & Schauer, N. (2013). Metabolomics-Assisted Plant Breeding. In *The Handbook of Plant Metabolomics* (pp. 245–254). <https://doi.org/10.1002/9783527669882.ch13>
- Hertel, Rosch, T. W., & D., S. (2010). *Climate change, agriculture and poverty*. The World Bank.

- Hildebrandt, T. M. (2018). Synthesis versus degradation: directions of amino acid metabolism during Arabidopsis abiotic stress response. *Plant Molecular Biology*, *98*(1-2), 121–135.
- Hisano, H., Nandakumar, R., & Wang, Z.-Y. (2011). Genetic Modification of Lignin Biosynthesis for Improved Biofuel Production. In D. Tomes, P. Lakshmanan, & D. Songstad (Eds.), *Biofuels: Global Impact on Renewable Energy, Production Agriculture, and Technological Advancements* (pp. 223–235). Springer New York.
- Hoang, N. V., Furtado, A., Mason, P. J., Marquardt, A., Kasirajan, L., Thirugnanasambandam, P. P., Botha, F. C., & Henry, R. J. (2017). 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*, *18*(1), 395.
- Hodgson-Kratky, K., Papa, G., Rodriguez, A., Stavila, V., Simmons, B., Botha, F., Furtado, A., & Henry, R. (2019). Relationship between sugarcane culm and leaf biomass composition and saccharification efficiency. *Biotechnology for Biofuels*, *12*, 247.
- Hong, J., Yang, L., Zhang, D., & Shi, J. (2016). Plant Metabolomics: An Indispensable System Biology Tool for Plant Science. *International Journal of Molecular Sciences*, *17*(6). <https://doi.org/10.3390/ijms17060767>
- Huang, J., Gu, M., Lai, Z., Fan, B., Shi, K., Zhou, Y.-H., Yu, J.-Q., & Chen, Z. (2010). Functional analysis of the Arabidopsis PAL gene family in plant growth, development, and response to environmental stress. *Plant Physiology*, *153*(4), 1526–1538.
- Hüsemann, W., Fischer, K., Mittelbach, I., Hübner, S., Richter, G., & Barz, W. (1989). Photoautotrophic Plant Cell Cultures for Studies on Primary and Secondary Metabolism. In *Primary and Secondary Metabolism of Plant Cell Cultures II* (pp. 35–46). [https://doi.org/10.1007/978-3-642-74551-5\\_4](https://doi.org/10.1007/978-3-642-74551-5_4)
- Hu, Y., Li, W.-C., Xu, Y.-Q., Li, G.-J., Liao, Y., & Fu, F.-L. (2009). Differential expression of candidate genes for lignin biosynthesis under drought stress in maize leaves. *Journal of Applied Genetics*, *50*(3), 213–223.
- Hu, Z., Sykes, R., Davis, M. F., Charles Brummer, E., & Ragauskas, A. J. (2010). Chemical profiles of switchgrass. *Bioresource Technology*, *101*(9), 3253–3257.
- Igarashi, K., & Kashiwagi, K. (2015). Modulation of protein synthesis by polyamines. *IUBMB Life*, *67*(3), 160–169.
- Imamura, Y., Migita, T., Uriu, Y., Otagiri, M., & Okawara, T. (2000). Inhibitory effects of flavonoids on rabbit heart carbonyl reductase. *Journal of Biochemistry*, *127*(4), 653–658.
- Inada, D. T., & Souza, G. M. (2016). *Análise do metaboloma e transcriptoma da cana-de-açúcar ao longo do ciclo de maturação da planta*. <https://bdpi.usp.br/item/002832939>
- Inman-Bamber, N. G., Bonnett, G. D., Spillman, M. F., Hewitt, M. H., & Glassop, D. (2010). Sucrose accumulation in sugarcane is influenced by temperature and genotype through the carbon source - sink balance. In *Crop and Pasture Science* (Vol. 61, Issue 2, p. 111). <https://doi.org/10.1071/cp09262>
- Inman-Bamber, N. G., Bonnett, G. D., Spillman, M. F., Hewitt, M. L., & Xu, J. (2009). Source - sink differences in genotypes and water regimes influencing sucrose accumulation in sugarcane stalks. In *Crop and Pasture Science* (Vol. 60, Issue 4, p. 316). <https://doi.org/10.1071/cp08272>
- Izawa, T., Mihara, M., Suzuki, Y., Gupta, M., Itoh, H., Nagano, A. J., Motoyama, R., Sawada, Y., Yano, M., Hirai, M. Y., Makino, A., & Nagamura, Y. (2011). Os-GIGANTEA confers robust diurnal rhythms on the global transcriptome of rice in the field. *The Plant Cell*, *23*(5), 1741–1755.
- Jackson, J. E., & Edward, A. (1991). User's guide to principal components. John Wiley Sons. *Inc. , New York*, 40.
- Jaeger, C., & Lise, J. (2018). Statistical and Multivariate Analysis of MS-Based Plant Metabolomics Data. In *Methods in Molecular Biology* (pp. 285–296). [https://doi.org/10.1007/978-1-4939-7819-9\\_20](https://doi.org/10.1007/978-1-4939-7819-9_20)

- Jakabová, S., Vincze, L., Farkas, Á., Kilar, F., Boros, B., & Felinger, A. (2012). Determination of tropane alkaloids atropine and scopolamine by liquid chromatography–mass spectrometry in plant organs of *Datura* species. *Journal of Chromatography. A*, *1232*, 295–301.
- Jeong, C. H., & Shim, K. H. (2004). Tyrosinase inhibitor isolated from the leaves of *Zanthoxylum piperitum*. *Bioscience, Biotechnology, and Biochemistry*, *68*(9), 1984–1987.
- Johnson, I. T. (2002). Glucosinolates: bioavailability and importance to health. *International Journal for Vitamin and Nutrition Research. Internationale Zeitschrift Fur Vitamin- Und Ernahrungsforschung. Journal International de Vitaminologie et de Nutrition*, *72*(1), 26–31.
- Johnson, R. M., Grisham, M. P., Warnke, K. Z., & Maggio, J. R. (2016). Relationship of Soil Properties and Sugarcane Yields to Red Stripe in Louisiana. *Phytopathology*, *106*(7), 737–744.
- Jolliffe, I. T. (2002). Springer series in statistics. *Principal Component Analysis*, *29*.
- Jorge, T. F., & António, C. (2018). Plant Metabolomics in a Changing World: Metabolite Responses to Abiotic Stress Combinations. In *Plant, Abiotic Stress and Responses to Climate Change*. <https://doi.org/10.5772/intechopen.71769>
- Jossier, M., Bouly, J.-P., Meimoun, P., Arjmand, A., Lessard, P., Hawley, S., Grahame Hardie, D., & Thomas, M. (2009). SnRK1 (SNF1-related kinase 1) has a central role in sugar and ABA signalling in *Arabidopsis thaliana*. *The Plant Journal: For Cell and Molecular Biology*, *59*(2), 316–328.
- Joyard, J., & McCormick, S. (2010). Plant Systems Biology. *Plant Physiology*, *152*(2), 401–401.
- Jr, V. E. D. E. R., Nogueira, F. T. S., & Mazzafera, P. (2007). Sugarcane dhurrin: biosynthetic pathway regulation and evolution. *Proc. Int. Soc. Sugar Cane Technol.*, *26*. <https://pdfs.semanticscholar.org/17e8/9f4f535a452bac63be3b6bbc98cefcac12c2.pdf>
- Kähkönen, M. P., & Heinonen, M. (2003). Antioxidant activity of anthocyanins and their aglycons. *Journal of Agricultural and Food Chemistry*, *51*(3), 628–633.
- Kan, C.-C., Chung, T.-Y., Wu, H.-Y., Juo, Y.-A., & Hsieh, M.-H. (2017). Exogenous glutamate rapidly induces the expression of genes involved in metabolism and defense responses in rice roots. *BMC Genomics*, *18*(1), 186.
- Kanehisa, M. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. In *Nucleic Acids Research* (Vol. 28, Issue 1, pp. 27–30). <https://doi.org/10.1093/nar/28.1.27>
- Kanteev, M., Goldfeder, M., & Fishman, A. (2015). Structure-function correlations in tyrosinases. *Protein Science: A Publication of the Protein Society*, *24*(9), 1360–1369.
- Kavi Kishor, P. B., Hima Kumari, P., Sunita, M. S. L., & Sreenivasulu, N. (2015). Role of proline in cell wall synthesis and plant development and its implications in plant ontogeny. *Frontiers in Plant Science*, *6*, 544.
- Kavi Kishor, P. B., & Sreenivasulu, N. (2014). Is proline accumulation per se correlated with stress tolerance or is proline homeostasis a more critical issue? *Plant, Cell & Environment*, *37*(2), 300–311.
- Kawaoka, A., & Ebinuma, H. (2001). Transcriptional control of lignin biosynthesis by tobacco LIM protein. *Phytochemistry*, *57*(7), 1149–1157.
- Kell, D. B., & Oliver, S. G. (2004). Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. In *BioEssays* (Vol. 26, Issue 1, pp. 99–105). <https://doi.org/10.1002/bies.10385>
- Kim, T. H., Kim, E. C., Kim, S. W., Lee, H. S., & Choi, D.-W. (2010). Exogenous Glutamate Inhibits the Root Growth and Increases the Glutamine Content in *Arabidopsis thaliana*. *Journal of Plant Biology = Singmul Hakboe Chi*, *53*(1), 45–51.

- Kleessen, S., Irgang, S., Klie, S., Giavalisco, P., & Nikoloski, Z. (2015). Integration of transcriptomics and metabolomics data specifies the metabolic response of *Chlamydomonas* to rapamycin treatment. *The Plant Journal: For Cell and Molecular Biology*, *81*(5), 822–835.
- Kliebenstein, D. J. (2007). Metabolomics and Plant Quantitative Trait Locus Analysis – The Optimum Genetical Genomics Platform? In *Concepts in Plant Metabolomics* (pp. 29–44). [https://doi.org/10.1007/978-1-4020-5608-6\\_3](https://doi.org/10.1007/978-1-4020-5608-6_3)
- Koal, T., & Deigner, H.-P. (2010). Challenges in mass spectrometry based targeted metabolomics. *Current Molecular Medicine*, *10*(2), 216–226.
- Kogel, K.-H., Voll, L. M., Schäfer, P., Jansen, C., Wu, Y., Langen, G., Imani, J., Hofmann, J., Schmiedl, A., Sonnewald, S., von Wettstein, D., Cook, R. J., & Sonnewald, U. (2010). Transcriptome and metabolome profiling of field-grown transgenic barley lack induced differences but show cultivar-specific variances. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(14), 6198–6203.
- Kosar, F., Akram, N. A., Sadiq, M., Al-Qurainy, F., & Ashraf, M. (2019). Trehalose: A Key Organic Osmolyte Effectively Involved in Plant Abiotic Stress Tolerance. *Journal of Plant Growth Regulation*, *38*(2), 606–618.
- Kretschmer, P. M., Bannister, A. M., MacManus-Spencer, L. A., Paulick, M. G., & Others. (2016). A liquid chromatography–tandem mass spectrometry assay for the detection and quantification of trehalose in biological samples. *Journal of Chromatography B*, *1033*, 9–16.
- Kruger, N. J., & von Schaewen, A. (2003). The oxidative pentose phosphate pathway: structure and organisation. *Current Opinion in Plant Biology*, *6*(3), 236–246.
- Krzyzaniak, Y., Negrel, J., Lemaitre-Guillier, C., Clément, G., Mouille, G., Klinguer, A., Trouvelot, S., Héloir, M.-C., & Adrian, M. (2018). Combined enzymatic and metabolic analysis of grapevine cell responses to elicitors. *Plant Physiology and Biochemistry: PPB / Societe Francaise de Physiologie Vegetale*, *123*, 141–148.
- Kuhl, C., Tautenhahn, R., Böttcher, C., Larson, T. R., & Neumann, S. (2012). CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Analytical Chemistry*, *84*(1), 283–289.
- Kurosaki, F. (2012). Induction and activation of plant secondary metabolism by external stimuli. *Drug Discovery Research in Pharmacognosy. Vallisuta O. , and Olimat SM (Eds. ), Wwnv. Intechopen. Com*, 209–226.
- Lakshmanan, P., & Robinson, N. (2013). Stress Physiology: Abiotic Stresses. In P. H. Moore & F. C. Botha (Eds.), *Sugarcane: Physiology, Biochemistry, and Functional Biology* (Vol. 46, pp. 411–434). John Wiley & Sons Ltd.
- Lämmerhofer, M., & Weckwerth, W. (2013). *Metabolomics in Practice: Successful Strategies to Generate and Analyse Metabolic Data*. John Wiley & Sons.
- Langridge, P., & Fleury, D. (2011). Making the most of “omics” for crop breeding. *Trends in Biotechnology*, *29*(1), 33–40.
- Lanot, A., Hodge, D., Jackson, R. G., George, G. L., Elias, L., Lim, E.-K., Vaistij, F. E., & Bowles, D. J. (2006). The glucosyltransferase UGT72E2 is responsible for monolignol 4-O-glucoside production in *Arabidopsis thaliana*. *The Plant Journal: For Cell and Molecular Biology*, *48*(2), 286–295.
- Lee, S. Y., So, Y.-J., Shin, M. S., Cho, J. Y., & Lee, J. (2014). Antibacterial effects of afzelin isolated from *Cornus macrophylla* on *Pseudomonas aeruginosa*, a leading cause of illness in immunocompromised individuals. *Molecules*, *19*(3), 3173–3180.
- Legendre, B. L. (1975). Ripening of Sugarcane: Effects of Sunlight, Temperature, and Rainfall 1. *Crop Science*, *15*(3), 349–352.

- Li, B., Tang, J., Yang, Q., Cui, X., Li, S., Chen, S., Cao, Q., Xue, W., Chen, N., & Zhu, F. (2016). Performance Evaluation and Online Realization of Data-driven Normalization Methods Used in LC/MS based Untargeted Metabolomics Analysis. *Scientific Reports*, *6*, 38881.
- Li, C., Nong, Q., Solanki, M. K., Liang, Q., Xie, J., Liu, X., Li, Y., Wang, W., Yang, L., & 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.
- Lima, F. de A. e., de Abreu e Lima, F., Leifels, L., & Nikoloski, Z. (2018). Regression-Based Modeling of Complex Plant Traits Based on Metabolomics Data. In *Methods in Molecular Biology* (pp. 321–327). [https://doi.org/10.1007/978-1-4939-7819-9\\_23](https://doi.org/10.1007/978-1-4939-7819-9_23)
- Liu, D. L., Kingston, G., & Bull, T. A. (1998). A new technique for determining the thermal parameters of phenological development in sugarcane, including suboptimum and supra-optimum temperature regimes. *Agricultural and Forest Meteorology*, *90*(1), 119–139.
- Liu, M., Li, X., Liu, Y., & Cao, B. (2013). Regulation of flavanone 3-hydroxylase gene involved in the flavonoid biosynthesis pathway in response to UV-B radiation and drought stress in the desert plant, *Reaumuria soongorica*. *Plant Physiology and Biochemistry: PPB / Societe Francaise de Physiologie Vegetale*, *73*, 161–167.
- Liu, Y., Gu, D., Wu, W., Wen, X., & Liao, Y. (2013). The relationship between polyamines and hormones in the regulation of wheat grain filling. *PLoS One*, *8*(10), e78196.
- Li, X., Bonawitz, N. D., Weng, J.-K., & Chapple, C. (2010). The growth reduction associated with repressed lignin biosynthesis in *Arabidopsis thaliana* is independent of flavonoids. *The Plant Cell*, *22*(5), 1620–1632.
- Li, Z.-Y., Zhi, H.-J., Zhang, F.-S., Sun, H.-F., Zhang, L.-Z., Jia, J.-P., Xing, J., & Qin, X.-M. (2013). Metabolomic profiling of the antitussive and expectorant plant *Tussilago farfara* L. by nuclear magnetic resonance spectroscopy and multivariate data analysis. *Journal of Pharmaceutical and Biomedical Analysis*, *75*, 158–164.
- Luckner, M. (2013). *Secondary Metabolism in Microorganisms, Plants and Animals*. Springer Science & Business Media.
- Lukas, E. N. (2015). Green Economy for Sustainable Development and Poverty Eradication. In *Mediterranean Journal of Social Sciences*. <https://doi.org/10.5901/mjss.2015.v6n6s5p434>
- Lunn, J. E., Delorge, I., Figueroa, C. M., Van Dijck, P., & Stitt, M. (2014). Trehalose metabolism in plants. *The Plant Journal: For Cell and Molecular Biology*, *79*(4), 544–567.
- Lunn, J. E., Feil, R., Hendriks, J. H. M., Gibon, Y., Morcuende, R., Osuna, D., Scheible, W.-R., Carillo, P., Hajirezaei, M.-R., & Stitt, M. (2006). Sugar-induced increases in trehalose 6-phosphate are correlated with redox activation of ADPglucose pyrophosphorylase and higher rates of starch synthesis in *Arabidopsis thaliana*. *Biochemical Journal*, *397*(1), 139–148.
- Luo, J., Fuell, C., Parr, A., Hill, L., Bailey, P., Elliott, K., Fairhurst, S. A., Martin, C., & Michael, A. J. (2009). A novel polyamine acyltransferase responsible for the accumulation of spermidine conjugates in *Arabidopsis* seed. *The Plant Cell*, *21*(1), 318–333.
- MacArthur, M. W., & Thornton, J. M. (1991). Influence of proline residues on protein conformation. *Journal of Molecular Biology*, *218*(2), 397–412.
- Ma, D., Sun, D., Wang, C., Li, Y., & Guo, T. (2014). Expression of flavonoid biosynthesis genes and accumulation of flavonoid in wheat leaves in response to drought stress. *Plant Physiology and Biochemistry: PPB / Societe Francaise de Physiologie Vegetale*, *80*, 60–66.
- Maeda, H., Song, W., Sage, T., & Dellapenna, D. (2014). Role of callose synthases in transfer cell wall development in tocopherol deficient *Arabidopsis* mutants. *Frontiers in Plant Science*, *5*, 46.

- Manechini, J. R. V., da Costa, J. B., Pereira, B. T., Carlini-Garcia, L. A., Xavier, M. A., Landell, M. G. de A., & Pinto, L. R. (2018). Unraveling the genetic structure of Brazilian commercial sugarcane cultivars through microsatellite markers. *PLoS One*, *13*(4), e0195623.
- Ma, Q.-H. (2010). Functional analysis of a cinnamyl alcohol dehydrogenase involved in lignin biosynthesis in wheat. *Journal of Experimental Botany*, *61*(10), 2735–2744.
- Marco-Ramell, A., Palau-Rodriguez, M., Alay, A., Tulipani, S., Urpi-Sarda, M., Sanchez-Pla, A., & Andres-Lacueva, C. (2018). Evaluation and comparison of bioinformatic tools for the enrichment analysis of metabolomics data. *BMC Bioinformatics*, *19*(1), 1.
- Margarido, F. B., & Santos, F. (2015). Chapter 1 - Agricultural Planning. In F. Santos, A. Borém, & C. Caldas (Eds.), *Sugarcane* (pp. 1–11). Academic Press.
- Marli, S., Aparecida, F., da Graa, J. P., de Matos Pereira, L., & Vergnia, M. (2012). Sugarcane Responses at Water Deficit Conditions. In *Water Stress*. <https://doi.org/10.5772/30986>
- Martínez-Barajas, E., Delatte, T., & Schlupepmann, H. (2011). Wheat grain development is characterized by remarkable trehalose 6-phosphate accumulation pregrain filling: tissue distribution and relationship to SNF1-related .... *Plant*. <http://www.plantphysiol.org/content/156/1/373.short>
- Martin-Tanguy, J. (1997). Conjugated polyamines and reproductive development: Biochemical, molecular and physiological approaches. *Physiologia Plantarum*, *100*(3), 675–688.
- Mata-Pérez, C., Begara-Morales, J. C., Chaki, M., Sánchez-Calvo, B., Valderrama, R., Padilla, M. N., Corpas, F. J., & Barroso, J. B. (2016). Protein Tyrosine Nitration during Development and Abiotic Stress Response in Plants. *Frontiers in Plant Science*, *7*, 1699.
- Matilla, M. A. (2018). Metabolic Responses of Plants Upon Different Plant–Pathogen Interactions. In *Plant Metabolites and Regulation Under Environmental Stress* (pp. 195–214). <https://doi.org/10.1016/b978-0-12-812689-9.00010-8>
- Matsuda, H., Morikawa, T., Toguchida, I., & Yoshikawa, M. (2002). Structural requirements of flavonoids and related compounds for aldose reductase inhibitory activity. *Chemical & Pharmaceutical Bulletin*, *50*(6), 788–795.
- Matsuoka, S., Kennedy, A. J., Santos, E. G. D. dos, Tomazela, A. L., & Rubio, L. C. S. (2014). Energy Cane: Its Concept, Development, Characteristics, and Prospects. *Advances in Botany*, 2014. <https://doi.org/10.1155/2014/597275>
- Mattioli, R., Costantino, P., & Trovato, M. (2009). Proline accumulation in plants: not only stress. *Plant Signaling & Behavior*, *4*(11), 1016–1018.
- McKay, B., Sauer, S., Richardson, B., & Herre, R. (2016). The political economy of sugarcane flexing: initial insights from Brazil, Southern Africa and Cambodia. *The Journal of Peasant Studies*, *43*(1), 195–223.
- Mekonnen, Z. (2019). Observed and Projected Reciprocal Effects of Agriculture and Climate Change: Implications on Ecosystems and Human Livelihoods. In *Climate Change and Global Warming*. <https://doi.org/10.5772/intechopen.79118>
- Melloni, M. L. G., Melloni, M. N. G., Scarpari, M. S., Garcia, J. C., Landell, M. G. A., Pinto, L. R., & Others. (2015). Flowering of sugarcane genotypes under different artificial photoperiod conditions. *American Journal of Plant Sciences*, *6*(03), 456.
- Mendelsohn, R., & Dinar, A. (2009). Literature Review of Economic Impacts of Climate Change on Agriculture. In *Climate Change and Agriculture*. Edward Elgar Publishing.

- Migné, C., Prensier, G., & Grenet, E. (1994). Immunocytochemical localisation of arabinoxylans in the cell walls of maize stem and their fate after incubation in the rumen. In *Annales de Zootechnie* (Vol. 43, Issue 3, pp. 271–271). <https://doi.org/10.1051/animres:19940338>
- Min, D., Chang, H., Jameel, H., Lucia, L., Wang, Z., & Jin, Y. (2014). The structure of lignin of corn stover and its changes induced by mild sodium hydroxide treatment. *BioResources*. <http://ojs.cnr.ncsu.edu/index.php/BioRes/article/view/5167>
- Mira, L., Silva, M., & Manso, C. F. (1994). Scavenging of reactive oxygen species by silibinin dihemisuccinate. *Biochemical Pharmacology*, 48(4), 753–759.
- Misra, B. B., Langefeld, C. D., Olivier, M., & Cox, L. A. (2018). Integrated Omics: Tools, Advances, and Future Approaches. *Journal of Molecular Endocrinology*. <https://doi.org/10.1530/JME-18-0055>
- Mitchell, M. C., Pritchard, J., Okada, S., Zhang, J., Venables, I., Vanhercke, T., & Ral, J.-P. (2020). Increasing growth and yield by altering carbon metabolism in a transgenic leaf oil crop. *Plant Biotechnology Journal*. <https://doi.org/10.1111/pbi.13363>
- Mithöfer, A., & Boland, W. (2012). Plant defense against herbivores: chemical aspects. *Annual Review of Plant Biology*, 63, 431–450.
- Mochida, K., & Shinozaki, K. (2011). Advances in omics and bioinformatics tools for systems analyses of plant functions. *Plant & Cell Physiology*, 52(12), 2017–2038.
- Moco, S., Bino, R. J., Vorst, O., Verhoeven, H. A., de Groot, J., van Beek, T. A., Vervoort, J., & de Vos, C. H. R. (2006). A liquid chromatography-mass spectrometry-based metabolome database for tomato. *Plant Physiology*, 141(4), 1205–1218.
- Moore, P. H. (1995). Temporal and Spatial Regulation of Sucrose Accumulation in the Sugarcane Stem. *Functional Plant Biology: FPB*, 22(4), 661–679.
- Moore, P. H., & Botha, F. C. (2013). *Sugarcane: Physiology, Biochemistry and Functional Biology*. John Wiley & Sons.
- Moore, P. H., Paterson, A. H., & Tew, T. (2013). Sugarcane: The Crop, the Plant, and Domestication. In *Sugarcane: Physiology, Biochemistry, and Functional Biology* (pp. 1–17). <https://doi.org/10.1002/9781118771280.ch1>
- Morgan, P. W. (1984). Is ethylene the Natural Regulator of Abscission? In *Ethylene* (pp. 231–240). [https://doi.org/10.1007/978-94-009-6178-4\\_35](https://doi.org/10.1007/978-94-009-6178-4_35)
- Morreel, K., Kim, H., Lu, F., Dima, O., Akiyama, T., Vanholme, R., Niculaes, C., Goeminne, G., Inzé, D., Messens, E., Ralph, J., & Boerjan, W. (2010). Mass Spectrometry-Based Fragmentation as an Identification Tool in Lignomics. In *Analytical Chemistry* (Vol. 82, Issue 19, pp. 8095–8105). <https://doi.org/10.1021/ac100968g>
- Moura, J. C. M. S., Bonine, C. A. V., de Oliveira Fernandes Viana, J., Dornelas, M. C., & Mazzafera, P. (2010). Abiotic and biotic stresses and changes in the lignin content and composition in plants. *Journal of Integrative Plant Biology*, 52(4), 360–376.
- Müller, J., Boller, T., & Wiemken, A. (1995). Trehalose and trehalase in plants: recent developments. *Plant Science: An International Journal of Experimental Plant Biology*, 112(1), 1–9.
- Müller, J., Boller, T., & Wiemken, A. (1998). Trehalose affects sucrose synthase and invertase activities in soybean (*Glycine max* [L.] Merr.) roots. *Journal of Plant Physiology*, 153(1), 255–257.
- Mustafavi, S. H., Naghdi Badi, H., Şekara, A., Mehrafarin, A., Janda, T., Ghorbanpour, M., & Rafiee, H. (2018). Polyamines and their possible mechanisms involved in plant physiological processes and elicitation of secondary metabolites. *Acta Physiologiae Plantarum / Polish Academy of Sciences, Committee of Plant Physiology Genetics and Breeding*, 40(6), 102.

- Nakabayashi, R., Yonekura-Sakakibara, K., Urano, K., Suzuki, M., Yamada, Y., Nishizawa, T., Matsuda, F., Kojima, M., Sakakibara, H., Shinozaki, K., Michael, A. J., Tohge, T., Yamazaki, M., & Saito, K. (2014). Enhancement of oxidative and drought tolerance in Arabidopsis by overaccumulation of antioxidant flavonoids. *The Plant Journal: For Cell and Molecular Biology*, 77(3), 367–379.
- Nanda, S., Mohanty, J. N., Mishra, R., & Joshi, R. K. (2016). Metabolic Engineering of PhenylPropanoids in Plants. In *Transgenesis and Secondary Metabolism* (pp. 1–26). [https://doi.org/10.1007/978-3-319-27490-4\\_30-1](https://doi.org/10.1007/978-3-319-27490-4_30-1)
- Naoumkina, M. A., Zhao, Q., Gallego-Giraldo, L., Dai, X., Zhao, P. X., & Dixon, R. A. (2010). Genome-wide analysis of phenylpropanoid defence pathways. *Molecular Plant Pathology*, 11(6), 829–846.
- Nay Min Min. (2018). *Advances in Data Analysis in Metabolomics: Towards a Dynamic View of Responses in Plant Cell Cultures*.
- Neill, S. O., & Gould, K. S. (2003). Anthocyanins in leaves: light attenuators or antioxidants? *Functional Plant Biology: FPB*, 30(8), 865–873.
- Nguyen, L. H., & Holmes, S. (2019). Ten quick tips for effective dimensionality reduction. *PLoS Computational Biology*, 15(6), e1006907.
- Nguyen-Quoc, B., & Foyer, C. H. (2001). A role for “futile cycles” involving invertase and sucrose synthase in sucrose metabolism of tomato fruit. *Journal of Experimental Botany*, 52(358), 881–889.
- Nishiyama, M. Y., Vicente, F., Sato, P. M., Ferreira, S. S., Feltus, F. A., & Souza, G. M. (2013). Transcriptome Analysis in the Saccharinae. In A. H. Paterson (Ed.), *Genomics of the Saccharinae* (pp. 121–139). Springer New York.
- Nkomo, M., Gokul, A., Keyster, M., & Klein, A. (2019). Exogenous p-Coumaric Acid Improves *Salvia hispanica* L. Seedling Shoot Growth. *Plants*, 8(12). <https://doi.org/10.3390/plants8120546>
- Noel, J. P., Austin, M. B., & Bomati, E. K. (2005). Structure-function relationships in plant phenylpropanoid biosynthesis. *Current Opinion in Plant Biology*, 8(3), 249–253.
- [No title]. (n.d). Retrieved May 1, 2020, from [https://www.researchgate.net/profile/David\\_Fell/publication/317369476\\_Understanding\\_the\\_Control\\_of\\_Metabolism/links/5936ce64a6fdcca65863eb46/Understanding-the-Control-of-Metabolism.pdf](https://www.researchgate.net/profile/David_Fell/publication/317369476_Understanding_the_Control_of_Metabolism/links/5936ce64a6fdcca65863eb46/Understanding-the-Control-of-Metabolism.pdf)
- Nowicka, B., & Kruk, J. (2010). Occurrence, biosynthesis and function of isoprenoid quinones. *Biochimica et Biophysica Acta*, 1797(9), 1587–1605.
- Obata, T., & Fernie, A. R. (2012). The use of metabolomics to dissect plant responses to abiotic stresses. *Cellular and Molecular Life Sciences: CMLS*, 69(19), 3225–3243.
- Ohashi, Y., Nakayama, N., Saneoka, H., Mohapatra, P. K., & Fujita, K. (2009). Differences in the responses of stem diameter and pod thickness to drought stress during the grain filling stage in soybean plants. In *Acta Physiologiae Plantarum* (Vol. 31, Issue 2, pp. 271–277). <https://doi.org/10.1007/s11738-008-0229-4>
- Okada, T., Afendi, F. M., Altaf-Ul-Amin, M., Takahashi, H., Nakamura, K., & Kanaya, S. (2010). Metabolomics of medicinal plants: the importance of multivariate analysis of analytical chemistry data. *Current Computer-Aided Drug Design*, 6(3), 179–196.
- Okuyama, L. A., Federizzi, L. C., & Barbosa Neto, J. F. (2005). Plant traits to complement selection based on yield components in wheat. *Ciencia Rural*, 35(5), 1010–1018.
- Oliveira, R. D. L., Dhingra, O. D., Lima, A. O., Jham, G. N., Berhow, M. A., Holloway, R. K., & Vaughn, S. F. (2011). Glucosinolate content and nematocidal activity of Brazilian wild mustard tissues against *Meloidogyne incognita* in tomato. *Plant and Soil*, 341(1), 155–164.

- Osorio, S., Alba, R., Damasceno, C. M. B., Lopez-Casado, G., Lohse, M., Zanor, M. I., Tohge, T., Usadel, B., Rose, J. K. C., Fei, Z., Giovannoni, J. J., & Fernie, A. R. (2011). Systems Biology of Tomato Fruit Development: Combined Transcript, Protein, and Metabolite Analysis of Tomato Transcription Factor (*nor*, *rin*) and Ethylene Receptor (*Nr*) Mutants Reveals Novel Regulatory Interactions. In *Plant Physiology* (Vol. 157, Issue 1, pp. 405–425). <https://doi.org/10.1104/pp.111.175463>
- Papini-Terzi, F. S., Rocha, F. R., Vêncio, R. Z. N., Felix, J. M., Branco, D. S., Waclawovsky, A. J., Del Bem, L. E. V., Lembke, C. G., Costa, M. D. L., Nishiyama, M. Y., Jr, Vicentini, R., Vincentz, M. G. A., Ulian, E. C., Menossi, M., & Souza, G. M. (2009). Sugarcane genes associated with sucrose content. *BMC Genomics*, *10*, 120.
- Paul, M. J., & Pellny, T. K. (2003). Carbon metabolite feedback regulation of leaf photosynthesis and development. *Journal of Experimental Botany*, *54*(382), 539–547.
- Ponnu, J., Wahl, V., & Schmid, M. (2011). Trehalose-6-phosphate: connecting plant metabolism and development. *Frontiers in Plant Science*, *2*, 70.
- Popova, I. E., Dubie, J. S., & Morra, M. J. (2017). Optimization of hydrolysis conditions for release of biopesticides from glucosinolates in *Brassica juncea* and *Sinapis alba* seed meal extracts. *Industrial Crops and Products*, *97*, 354–359.
- Prabu, G., Kawar, P. G., Pagariya, M. C., & Prasad, D. T. (2011). Identification of Water Deficit Stress Upregulated Genes in Sugarcane. *Plant Molecular Biology Reporter / ISPMB*, *29*(2), 291–304.
- Pratt, H. K., & Goeschl, J. D. (1969). Physiological Roles of Ethylene in Plants. *Annual Review of Plant Physiology*, *20*(1), 541–584.
- Pryke, J. A., & ap Rees, T. (1976). Activity of the pentose phosphate pathway during lignification. In *Planta* (Vol. 132, Issue 3, pp. 279–284). <https://doi.org/10.1007/bf00399727>
- Pryke, J. A., & Rees, T. ap. (1977). The pentose phosphate pathway as a source of NADPH for lignin synthesis. In *Phytochemistry* (Vol. 16, Issue 5, pp. 557–560). [https://doi.org/10.1016/0031-9422\(77\)80014-9](https://doi.org/10.1016/0031-9422(77)80014-9)
- Qin, S. S., Chen, S. Q., & Huang, L. Q. (2010). Effect of water stress on relationship of endogenous phytohormone and active compound content in roots of *Scutellaria baicalensis* Georgi. *Chin J Exp Tradition Med Form*, *16*, 99–101.
- Quanbeck, S. M. M., Brachova, L., & Campbell, A. A. (2012). Metabolomics as a hypothesis-generating functional genomics tool for the annotation of *Arabidopsis thaliana* genes of “unknown function.” *Frontiers in Plant Science*. <https://www.frontiersin.org/articles/10.3389/fpls.2012.00015>
- Rai, A., Saito, K., & Yamazaki, M. (2017). Integrated omics analysis of specialized metabolism in medicinal plants. *The Plant Journal: For Cell and Molecular Biology*, *90*(4), 764–787.
- Rambla, J. L., López-Gresa, M. P., Bellés, J. M., & Granell, A. (2015). Metabolomic profiling of plant tissues. *Methods in Molecular Biology*, *1284*, 221–235.
- Ranieri, A., Bernardi, R., Lanese, P., & Soldatini, G. F. (1989). Changes in free amino acid content and protein pattern of maize seedlings under water stress. In *Environmental and Experimental Botany* (Vol. 29, Issue 3, pp. 351–357). [https://doi.org/10.1016/0098-8472\(89\)90009-9](https://doi.org/10.1016/0098-8472(89)90009-9)
- Rao, M. J., Xu, Y., Tang, X., Huang, Y., Liu, J., Deng, X., & Xu, Q. (2020). CsCYT75B1, a Citrus CYTOCHROME P450 Gene, Is Involved in Accumulation of Antioxidant Flavonoids and Induces Drought Tolerance in Transgenic. *Antioxidants (Basel, Switzerland)*, *9*(2). <https://doi.org/10.3390/antiox9020161>
- Rashad El-Sherbeny, M., & Da Silva, J. A. T. (2013). Foliar treatment with proline and tyrosine affect the growth and yield of beetroot and some pigments in beetroot leaves. *Journal of Horticultural Research*, *21*(2), 95–99.

- Ratray, N. J. W., Deziel, N. C., Wallach, J. D., Khan, S. A., Vasiliou, V., Ioannidis, J. P. A., & Johnson, C. H. (2018). Beyond genomics: understanding exposotypes through metabolomics. *Human Genomics*, *12*(1), 4.
- Redovniković, I. R., Textor, S., Lisnić, B., & Gershenzon, J. (2012). Expression pattern of the glucosinolate side chain biosynthetic genes MAM1 and MAM3 of *Arabidopsis thaliana* in different organs and developmental stages. *Plant Physiology and Biochemistry: PPB / Societe Francaise de Physiologie Vegetale*, *53*, 77–83.
- Reignault, P. H., Cogan, A., Muchembled, J., Lounes-Hadj Sahraoui, A., Durand, R., & Sancholle, M. (2002). Trehalose induces resistance to powdery mildew in wheat. *The New Phytologist*, *149*(3), 519–529.
- Ren, S., Hinzman, A. A., Kang, E. L., Szczesniak, R. D., & Lu, L. J. (2015). Computational and statistical analysis of metabolomics data. In *Metabolomics* (Vol. 11, Issue 6, pp. 1492–1513). <https://doi.org/10.1007/s11306-015-0823-6>
- Riaño-Pachón, D. M., & Mattiello, L. (2017). Draft genome sequencing of the sugarcane hybrid SP80-3280. *F1000Research*, *6*, 861.
- Rischer, H., Oresic, M., Seppänen-Laakso, T., Katajamaa, M., Lammertyn, F., Ardiles-Diaz, W., Van Montagu, M. C. E., Inzé, D., Oksman-Caldentey, K.-M., & Goossens, A. (2006). Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in *Catharanthus roseus* cells. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(14), 5614–5619.
- Roberts, J. A. (Ed.). (2018). Perception of Ethylene by Plants - Ethylene Receptors. In *Annual Plant Reviews online* (Vol. 54, pp. 117–145). John Wiley & Sons, Ltd.
- Robertson, M. J., Wood, A. W., & Muchow, R. C. (1996). Growth of sugarcane under high input conditions in tropical Australia. I. Radiation use, biomass accumulation and partitioning. *Field Crops Research*, *48*(1), 11–25.
- Rochfort, S. (2005). Metabolomics reviewed: a new “omics” platform technology for systems biology and implications for natural products research. *Journal of Natural Products*, *68*(12), 1813–1820.
- Roldan, M. V. G., Engel, B., de Vos, R. C. H., Vereijken, P., Astola, L., Groenenboom, M., van de Geest, H., Bovy, A., Molenaar, J., van Eeuwijk, F., & Hall, R. D. (2014). Metabolomics reveals organ-specific metabolic rearrangements during early tomato seedling development. In *Metabolomics* (Vol. 10, Issue 5, pp. 958–974). <https://doi.org/10.1007/s11306-014-0625-2>
- Rongai, D., Cerato, C., & Lazzeri, L. (2009). A natural fungicide for the control of *Erysiphe betae* and *Erysiphe cichoracearum*. *European Journal of Plant Pathology / European Foundation for Plant Pathology*, *124*(4), 613–619.
- Roppolo, D., & Geldner, N. (2012). Membrane and walls: who is master, who is servant? *Current Opinion in Plant Biology*, *15*(6), 608–617.
- Rudaz, S. (2015). Identification and data-processing methods in metabolomics. In *Identification and Data Processing Methods in Metabolomics* (pp. 2–5). <https://doi.org/10.4155/fseb2013.14.270>
- Ruotolo, R., Maestri, E., Pagano, L., Marmiroli, M., White, J. C., & Marmiroli, N. (2018). Plant Response to Metal-Containing Engineered Nanomaterials: An Omics-Based Perspective. *Environmental Science & Technology*, *52*(5), 2451–2467.
- Ruprecht, C., & Persson, S. (2012). Co-expression of cell-wall related genes: new tools and insights. *Frontiers in Plant Science*, *3*, 83.
- Saathoff, A. J., Sarath, G., Chow, E. K., Dien, B. S., & Tobias, C. M. (2011). Downregulation of Cinnamyl-Alcohol Dehydrogenase in Switchgrass by RNA Silencing Results in Enhanced Glucose Release after Cellulase Treatment. In *PLoS ONE* (Vol. 6, Issue 1, p. e16416). <https://doi.org/10.1371/journal.pone.0016416>

- Salem, M. A., Jüppner, J., Bajdzienko, K., & Giavalisco, P. (2016). Protocol: a fast, comprehensive and reproducible one-step extraction method for the rapid preparation of polar and semi-polar metabolites, lipids, proteins, starch and cell wall polymers from a single sample. In *Plant Methods* (Vol. 12, Issue 1). <https://doi.org/10.1186/s13007-016-0146-2>
- Salem, M. A., Perez de Souza, L., Serag, A., Fernie, A. R., Farag, M. A., Ezzat, S. M., & Alseekh, S. (2020). Metabolomics in the Context of Plant Natural Products Research: From Sample Preparation to Metabolite Analysis. *Metabolites*, *10*(1). <https://doi.org/10.3390/metabo10010037>
- Santchurn, D., Ramdoyal, K., Badaloo, M. G. H., & Labuschagne, M. (2012). From sugar industry to cane industry: investigations on multivariate data analysis techniques in the identification of different high biomass sugarcane varieties. *Euphytica/ Netherlands Journal of Plant Breeding*, *185*(3), 543–558.
- Santos, A. B. dos, Bottcher, A., Kiyota, E., Mayer, J. L. S., Vicentini, R., Brito, M. dos S., Creste, S., Landell, M. G. A., & Mazzafera, P. (2015). Water Stress Alters Lignin Content and Related Gene Expression in Two Sugarcane Genotypes. *Journal of Agricultural and Food Chemistry*, *63*(19), 4708–4720.
- Schenck, C. A., & Maeda, H. A. (2018). Tyrosine biosynthesis, metabolism, and catabolism in plants. *Phytochemistry*, *149*, 82–102.
- Schrimpe-Rutledge, A. C., Codreanu, S. G., Sherrod, S. D., & McLean, J. A. (2016). Untargeted Metabolomics Strategies—Challenges and Emerging Directions. *Journal of the American Society for Mass Spectrometry*, *27*(12), 1897–1905.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). In *Biometrika* (Vol. 52, Issues 3-4, pp. 591–611). <https://doi.org/10.1093/biomet/52.3-4.591>
- Sharma, D., Tiwari, M., Pandey, A., Bhatia, C., Sharma, A., & Trivedi, P. K. (2016). MicroRNA858 Is a Potential Regulator of Phenylpropanoid Pathway and Plant Development. *Plant Physiology*, *171*(2), 944–959.
- Sharma, S., & Verslues, P. E. (2010). Mechanisms independent of abscisic acid (ABA) or proline feedback have a predominant role in transcriptional regulation of proline metabolism during low water potential and stress recovery. In *Plant, Cell & Environment* (Vol. 33, Issue 11, pp. 1838–1851). <https://doi.org/10.1111/j.1365-3040.2010.02188.x>
- Sheth, B. P., & Thaker, V. S. (2014). Plant systems biology: insights, advances and challenges. In *Planta* (Vol. 240, Issue 1, pp. 33–54). <https://doi.org/10.1007/s00425-014-2059-5>
- Shi, M., He, N., Li, W., Li, C., & Kang, W. (2018). Simultaneous determination of myricetrin, quercitrin and afzelin in leaves of *Cercis chinensis* by a fast and effective method of ionic liquid microextraction coupled with HPLC. In *Chemistry Central Journal* (Vol. 12, Issue 1). <https://doi.org/10.1186/s13065-018-0391-8>
- Shin, T. H., Lee, D. Y., Lee, H.-S., Park, H. J., Jin, M. S., Paik, M.-J., Manavalan, B., Mo, J.-S., & Lee, G. (2018). Integration of metabolomics and transcriptomics in nanotoxicity studies. *BMB Reports*, *51*(1), 14–20.
- Showler, A. T. (2002). Effects of water deficit stress, shade, weed competition, and kaolin particle film on selected foliar free amino acid accumulations in cotton, *Gossypium hirsutum* (L.). *Journal of Chemical Ecology*, *28*(3), 631–651.
- Shrivastava, A. K., Solomon, S., Rai, R. K., Singh, P., Chandra, A., Jain, R., & Shukla, S. P. (2015). Physiological Interventions for Enhancing Sugarcane and Sugar Productivity. *Sugar Tech*, *17*(3), 215–226.
- Sibout, R., Eudes, A., Mouille, G., Pollet, B., Lapierre, C., Jouanin, L., & Séguin, A. (2005). CINNAMYL ALCOHOL DEHYDROGENASE-C and -D are the primary genes involved in lignin biosynthesis in the floral stem of *Arabidopsis*. *The Plant Cell*, *17*(7), 2059–2076.

- Silva, M. de A., Jifon, J. L., Santos, C. M. dos, Jadoski, C. J., & Silva, J. A. G. da. (2013). Photosynthetic capacity and water use efficiency in sugarcane genotypes subject to water deficit during early growth phase. *Brazilian Archives of Biology and Technology*, 56(5), 735–748.
- Silva, R. R., Jourdan, F., Salvanha, D. M., Letisse, F., Jamin, E. L., Guidetti-Gonzalez, S., Labate, C. A., & Vêncio, R. Z. N. (2014). ProbMetab: an R package for Bayesian probabilistic annotation of LC-MS-based metabolomics. *Bioinformatics*, 30(9), 1336–1337.
- Simões, M. dos S., Rocha, J. V., & Lamparelli, R. A. C. (2005). Growth indices and productivity in sugarcane. *Scientia Agricola*, 62(1), 23–30.
- Singels, A., Donaldson, R. A., & Smit, M. A. (2005). Improving biomass production and partitioning in sugarcane: theory and practice. *Field Crops Research*, 92(2), 291–303.
- Singels, A., & Inman-Bamber, N. G. (2011). Modelling genetic and environmental control of biomass partitioning at plant and phytomer level of sugarcane grown in controlled environments. In *Crop and Pasture Science* (Vol. 62, Issue 1, p. 66). <https://doi.org/10.1071/cp10182>
- Skirycz, A., De Bodt, S., Obata, T., De Clercq, I., Claeys, H., De Rycke, R., Andriankaja, M., Van Aken, O., Van Breusegem, F., Fernie, A. R., & Inzé, D. (2010). Developmental stage specificity and the role of mitochondrial metabolism in the response of Arabidopsis leaves to prolonged mild osmotic stress. *Plant Physiology*, 152(1), 226–244.
- Smith, C. A., Want, E. J., O’Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry*, 78(3), 779–787.
- Smit, M. A., & Singels, A. (2006). The response of sugarcane canopy development to water stress. *Field Crops Research*, 98(2), 91–97.
- Song, X., Diao, J., Ji, J., Wang, G., Guan, C., Jin, C., & Wang, Y. (2016). Molecular cloning and identification of a flavanone 3-hydroxylase gene from *Lycium chinense*, and its overexpression enhances drought stress in tobacco. *Plant Physiology and Biochemistry: PPB / Societe Francaise de Physiologie Vegetale*, 98, 89–100.
- Souza, G. M., Berges, H., Bocs, S., Casu, R., D’Hont, A., Ferreira, J. E., Henry, R., Ming, R., Potier, B., Van Sluys, M.-A., Vincentz, M., & Paterson, A. H. (2011). The Sugarcane Genome Challenge: Strategies for Sequencing a Highly Complex Genome. *Tropical Plant Biology*, 4(3), 145–156.
- Souza, G. M., Van Sluys, M.-A., Lembke, C. G., Lee, H., Margarido, G. R. A., Hotta, C. T., Gaiarsa, J. W., Diniz, A. L., Oliveira, M. de M., Ferreira, S. de S., Nishiyama, M. Y., Ten-Caten, F., Ragagnin, G. T., Andrade, P. de M., de Souza, R. F., Nicastro, G. G., Pandya, R., Kim, C., Guo, H., ... Heckerman, D. (2019). Assembly of the 373k gene space of the polyploid sugarcane genome reveals reservoirs of functional diversity in the world’s leading biomass crop. *GigaScience*, 8(12). <https://doi.org/10.1093/gigascience/giz129>
- Steinfath, M., Groth, D., Lisek, J., & Selbig, J. (2008). Metabolite profile analysis: from raw data to regression and classification. *Physiologia Plantarum*, 132(2), 150–161.
- Stincone, A., Prigione, A., Cramer, T., Wamelink, M. M. C., Campbell, K., Cheung, E., Olin-Sandoval, V., Grüning, N.-M., Krüger, A., Alam, M. T., Keller, M. A., Breitenbach, M., Brindle, K. M., Rabinowitz, J. D., & Ralser, M. (2015). The return of metabolism: biochemistry and physiology of the pentose phosphate pathway. In *Biological Reviews* (Vol. 90, Issue 3, pp. 927–963). <https://doi.org/10.1111/brv.12140>
- Sultan, S. E. (2010). Plant developmental responses to the environment: eco-devo insights. *Current Opinion in Plant Biology*, 13(1), 96–101.

- Suman, A., Ali, K., Arro, J., Parco, A. S., Kimbeng, C. A., & Baisakh, N. (2012). Molecular Diversity Among Members of the Saccharum Complex Assessed Using TRAP Markers Based on Lignin-Related Genes. *Bioenergy Research*, 5(1), 197–205.
- Suriyan, C.-U., & Chalermopol, K. (2009). Proline Accumulation, Photosynthetic Abilities and Growth Characters of Sugarcane (*Saccharum officinarum* L.) Plantlets in Response to Iso-Osmotic Salt and Water-Deficit Stress. In *Agricultural Sciences in China* (Vol. 8, Issue 1, pp. 51–58). [https://doi.org/10.1016/s1671-2927\(09\)60008-0](https://doi.org/10.1016/s1671-2927(09)60008-0)
- Szymańska, E., Saccenti, E., Smilde, A. K., & Westerhuis, J. A. (2012). Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics: Official Journal of the Metabolomic Society*, 8(Suppl 1), 3–16.
- Tejera, N., Ortega, E., Rodes, R., & Lluch, C. (2006). Nitrogen compounds in the apoplastic sap of sugarcane stem: some implications in the association with endophytes. *Journal of Plant Physiology*, 163(1), 80–85.
- Tiawari, D. K., Pandey, P., Singh, R. K., Singh, S. P., & Singh, S. B. (2009). Sugar Productivity Assessment of Newly Developed Promising Genotypes of Sugarcane. *Journal of Bio-Science*, 17, 41–44.
- Timmerman, M. E., Hoefsloot, H. C. J., Smilde, A. K., & Ceulemans, E. (2015). Scaling in ANOVA-simultaneous component analysis. *Metabolomics: Official Journal of the Metabolomic Society*, 11(5), 1265–1276.
- Tolstikov, V. V., Lommen, A., Nakanishi, K., Tanaka, N., & Fiehn, O. (2003). Monolithic silica-based capillary reversed-phase liquid chromatography/electrospray mass spectrometry for plant metabolomics. *Analytical Chemistry*, 75(23), 6737–6740.
- Töpfer, N., Seaver, S. M. D., & Aharoni, A. (2018). Integration of Plant Metabolomics Data with Metabolic Networks: Progresses and Challenges. In *Methods in Molecular Biology* (pp. 297–310). [https://doi.org/10.1007/978-1-4939-7819-9\\_21](https://doi.org/10.1007/978-1-4939-7819-9_21)
- Trabucco, G. M., Matos, D. A., Lee, S. J., Saathoff, A. J., Priest, H. D., Mockler, T. C., Sarath, G., & Hazen, S. P. (2013). Functional characterization of cinnamyl alcohol dehydrogenase and caffeic acid O-methyltransferase in *Brachypodium distachyon*. *BMC Biotechnology*, 13, 61.
- Treml, J., & Šmejkal, K. (2016). Flavonoids as potent scavengers of hydroxyl radicals. *Comprehensive Reviews in Food Science and Food Safety*, 15(4), 720–738.
- Tzin, V., & Galili, G. (2010). New insights into the shikimate and aromatic amino acids biosynthesis pathways in plants. *Molecular Plant*, 3(6), 956–972.
- Uarrota, V. G., Moresco, R., Coelho, B., Nunes, E. da C., Peruch, L. A. M., Neubert, E. de O., Rocha, M., & Maraschin, M. (2014). Metabolomics combined with chemometric tools (PCA, HCA, PLS-DA and SVM) for screening cassava (*Manihot esculenta* Crantz) roots during postharvest physiological deterioration. *Food Chemistry*, 161, 67–78.
- Udhane, S. S., Legeza, B., Marti, N., Hertig, D., Diserens, G., Nuoffer, J.-M., Vermathen, P., & Flück, C. E. (2017). Combined transcriptome and metabolome analyses of metformin effects reveal novel links between metabolic networks in steroidogenic systems. *Scientific Reports*, 7(1), 8652.
- Ulfenborg, B. (2019). Vertical and horizontal integration of multi-omics data with miodin. *BMC Bioinformatics*, 20(1), 649.
- Urano, K., Maruyama, K., Ogata, Y., Morishita, Y., Takeda, M., Sakurai, N., Suzuki, H., Saito, K., Shibata, D., Kobayashi, M., Yamaguchi-Shinozaki, K., & Shinozaki, K. (2009). Characterization of the ABA-regulated global responses to dehydration in *Arabidopsis* by metabolomics. *The Plant Journal: For Cell and Molecular Biology*, 57(6), 1065–1078.

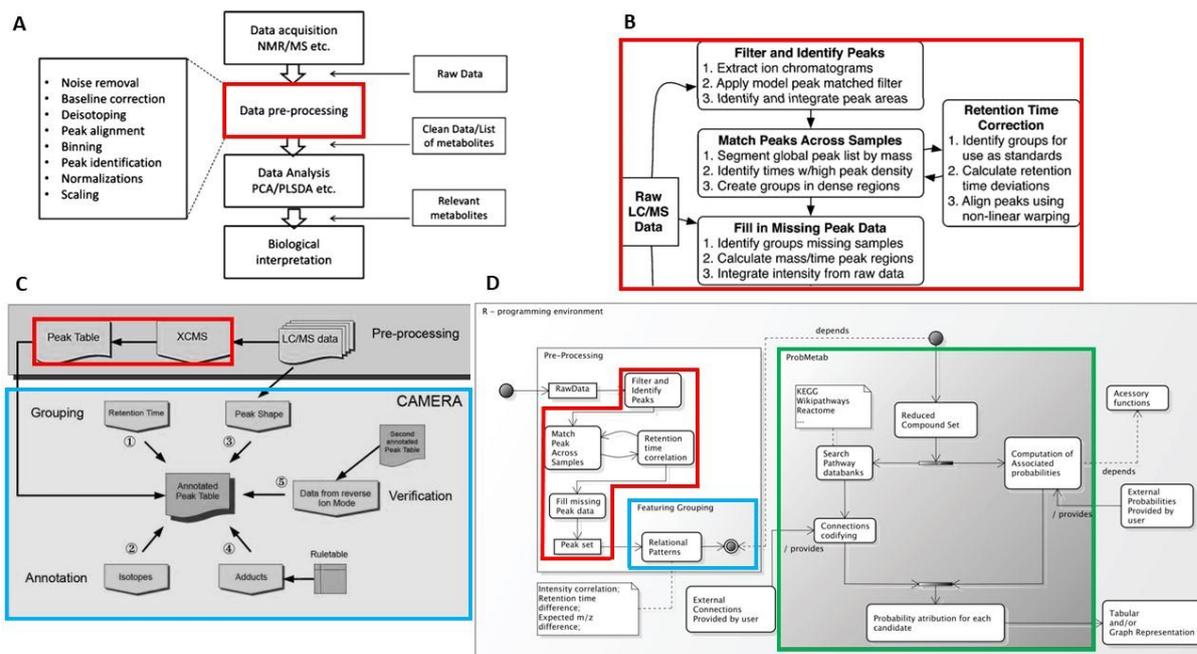
- Uys, L., Botha, F. C., Hofmeyr, J.-H. S., & Rohwer, J. M. (2007). Kinetic model of sucrose accumulation in maturing sugarcane culm tissue. *Phytochemistry*, *68*(16-18), 2375–2392.
- van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, *7*, 142.
- Vandesteene, L., Ramon, M., Le Roy, K., Van Dijck, P., & Rolland, F. (2010). A single active trehalose-6-P synthase (TPS) and a family of putative regulatory TPS-like proteins in Arabidopsis. *Molecular Plant*, *3*(2), 406–419.
- Vanholme, R., Morreel, K., Darrach, C., Oyarce, P., Grabber, J. H., Ralph, J., & Boerjan, W. (2012). Metabolic engineering of novel lignin in biomass crops. *The New Phytologist*, *196*(4), 978–1000.
- Vanholme, R., Storme, V., Vanholme, B., Sundin, L., Christensen, J. H., Goeminne, G., Halpin, C., Rohde, A., Morreel, K., & Boerjan, W. (2012). A systems biology view of responses to lignin biosynthesis perturbations in Arabidopsis. *The Plant Cell*, *24*(9), 3506–3529.
- Van Houtte, H., Vandesteene, L., López-Galvis, L., Lemmens, L., Kissel, E., Carpentier, S., Feil, R., Avonce, N., Beeckman, T., Lunn, J. E., & Van Dijck, P. (2013). Overexpression of the trehalase gene AtTRE1 leads to increased drought stress tolerance in Arabidopsis and is involved in abscisic acid-induced stomatal closure. *Plant Physiology*, *161*(3), 1158–1171.
- Vanneste, S., & Friml, J. (2009). Auxin: a trigger for change in plant development. *Cell*, *136*(6), 1005–1016.
- Vargas, C., Muniz de Padua, V. L., de Matos Nogueira, E., Vinagre, F., Masuda, H. P., Rodrigues da Silva, F., Baldani, J. I., Gomes Ferreira, P. C., & Hemerly, A. S. (2003). Signaling pathways mediating the association between sugarcane and endophytic diazotrophic bacteria: a genomic approach. *Symbiosis*. <https://dalspace.library.dal.ca/bitstream/handle/10222/78005/VOLUME%2035-NUMBERS%201-3-2003-PAGE%20159.pdf?sequence=1>
- Vermerris, W., & McIntyre, L. M. (1999). Time to flowering in brown midrib mutants of maize: an alternative approach to the analysis of developmental traits. *Heredity*, *83* (Pt 2), 171–178.
- Vettukattil, R. (2015). Preprocessing of raw metabolomic data. *Methods in Molecular Biology*, *1277*, 123–136.
- Vogt, T. (2010). Phenylpropanoid biosynthesis. *Molecular Plant*, *3*(1), 2–20.
- Vom Endt, D., Kijne, J. W., & Memelink, J. (2002). Transcription factors controlling plant secondary metabolism: what regulates the regulators? *Phytochemistry*, *61*(2), 107–114.
- Wagner de Oliveira, M., Cláudio Inácio da Silveira, L., Bosco de Oliveira, A., Henrique Pereira Barbosa, M., Gomes Pereira, M., & Bezerra Albino Oliveira, T. (2019). Sugarcane Production Systems in Small Rural Properties. In *Multifunctionality and Impacts of Organic Agriculture [Working Title]*. IntechOpen.
- Wahid, A., & Close, T. J. (2007). Expression of dehydrins under heat stress and their relationship with water relations of sugarcane leaves. *Biologia Plantarum*, *51*(1), 104–109.
- Walch-Liu, P., Liu, L.-H., Remans, T., Tester, M., & Forde, B. G. (2006). Evidence that L-glutamate can act as an exogenous signal to modulate root growth and branching in Arabidopsis thaliana. *Plant & Cell Physiology*, *47*(8), 1045–1057.
- Wallace, D. H. (2011). Physiological Genetics of Plant Maturity, Adaptation, and Yield. In *Plant Breeding Reviews* (pp. 21–167). <https://doi.org/10.1002/9781118061008.ch2>

- Wamelink, M. M. C., Struys, E. A., Huck, J. H. J., Roos, B., van der Knaap, M. S., Jakobs, C., & Verhoeven, N. M. (2005). Quantification of sugar phosphate intermediates of the pentose phosphate pathway by LC–MS/MS: application to two new inherited defects of metabolism. In *Journal of Chromatography B* (Vol. 823, Issue 1, pp. 18–25). <https://doi.org/10.1016/j.jchromb.2005.01.001>
- Wang, J., Feng, J., Jia, W., Chang, S., Li, S., & Li, Y. (2015). Lignin engineering through laccase modification: a promising field for energy plant improvement. *Biotechnology for Biofuels*, *8*, 145.
- Wang, Y., Chantreau, M., Sibout, R., & Hawkins, S. (2013). Plant cell wall lignification and monolignol metabolism. *Frontiers in Plant Science*, *4*, 220.
- Wani, N., & Raza, K. (2019). Integrative approaches to reconstruct regulatory networks from multi-omics data: A review of state-of-the-art methods. *Computational Biology and Chemistry*, *83*, 107120.
- Wiedenfeld, R. P. (2000). Water stress during different sugarcane growth periods on yield and response to N fertilization. *Agricultural Water Management*, *43*(2), 173–182.
- Wilk, M. B., & Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika*, *55*(1), 1–17.
- Wittstock, U., & Halkier, B. A. (2002). Glucosinolate research in the Arabidopsis era. *Trends in Plant Science*, *7*(6), 263–270.
- Wittstock, U., Kliebenstein, D. J., Lambrix, V., Reichelt, M., & Gershenzon, J. (2003). Glucosinolate hydrolysis and its impact on generalist and specialist insect herbivores. *Recent Advances in Phytochemistry*, *37*, 101–126.
- Worley, B., & Powers, R. (2013). Multivariate Analysis in Metabolomics. *Current Metabolomics*, *1*(1), 92–107.
- Wu, L., & Birch, R. G. (2010). Physiological basis for enhanced sucrose accumulation in an engineered sugarcane cell line. *Functional Plant Biology: FPB*, *37*(12), 1161–1174.
- Xia, J., Psychogios, N., Young, N., & Wishart, D. S. (2009). MetaboAnalyst: a web server for metabolomic data analysis and interpretation. In *Nucleic Acids Research* (Vol. 37, Issue Web Server, pp. W652–W660). <https://doi.org/10.1093/nar/gkp356>
- Yang, J., Sun, C., Fu, D., & Yu, T. (2017). Test for l-glutamate inhibition of growth of *Alternaria alternata* by inducing resistance in tomato fruit. *Food Chemistry*, *230*, 145–153.
- Yang, L., Wen, K.-S., Ruan, X., Zhao, Y.-X., Wei, F., & Wang, Q. (2018). Response of Plant Secondary Metabolites to Environmental Factors. *Molecules*, *23*(4). <https://doi.org/10.3390/molecules23040762>
- Yang, S. F., & Hoffman, N. E. (1984). Ethylene Biosynthesis and its Regulation in Higher Plants. *Annual Review of Plant Physiology*, *35*(1), 155–189.
- Yoo, H., Widhalm, J. R., Qian, Y., Maeda, H., Cooper, B. R., Jannasch, A. S., Gonda, I., Lewinsohn, E., Rhodes, D., & Dudareva, N. (2013). An alternative pathway contributes to phenylalanine biosynthesis in plants via a cytosolic tyrosine:phenylpyruvate aminotransferase. *Nature Communications*, *4*, 2833.
- Zang, Y.-X., Kim, J.-H., Park, Y.-D., Kim, D.-H., & Hong, S.-B. (2008). Metabolic engineering of aliphatic glucosinolates in Chinese cabbage plants expressing Arabidopsis MAM1, CYP79F1, and CYP83A1. *BMB Reports*, *41*(6), 472–478.
- Zhang, Y., Primavesi, L. F., Jhurreea, D., Andralojc, P. J., Mitchell, R. A. C., Powers, S. J., Schluempmann, H., Delatte, T., Wingler, A., & Paul, M. J. (2009). Inhibition of SNF1-related protein kinase1 activity and regulation of metabolic pathways by trehalose-6-phosphate. *Plant Physiology*, *149*(4), 1860–1871.
- Zhao, Q., Nakashima, J., Chen, F., Yin, Y., Fu, C., Yun, J., Shao, H., Wang, X., Wang, Z.-Y., & Dixon, R. A. (2013). Laccase is necessary and nonredundant with peroxidase for lignin polymerization during vascular development in Arabidopsis. *The Plant Cell*, *25*(10), 3976–3987.

- Zhou, M. (2017). Sugarcane (*S. officinarum* x *S. spontaneum*). In *Genetic Improvement of Tropical Crops* (pp. 291–308).  
[https://doi.org/10.1007/978-3-319-59819-2\\_9](https://doi.org/10.1007/978-3-319-59819-2_9)
- Zhao, W., Sun, G., and Li, S. (2004). Polyamines and plant stress resistance. *J. Southern Agric.* 35, 443–447. doi:  
10.3969/j.issn.2095-1191.2004.06.003

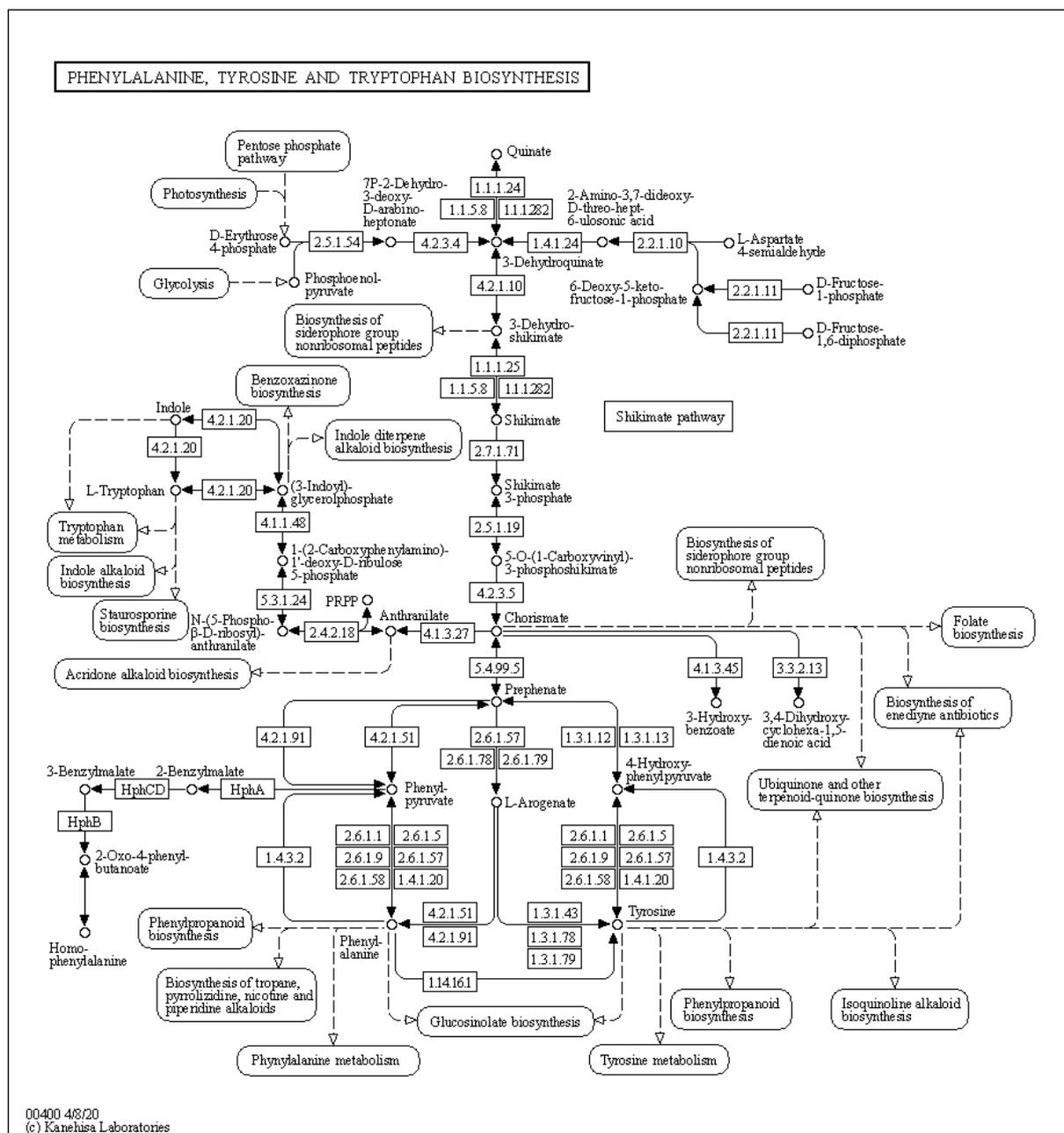


## Appendices – Supplementary figures

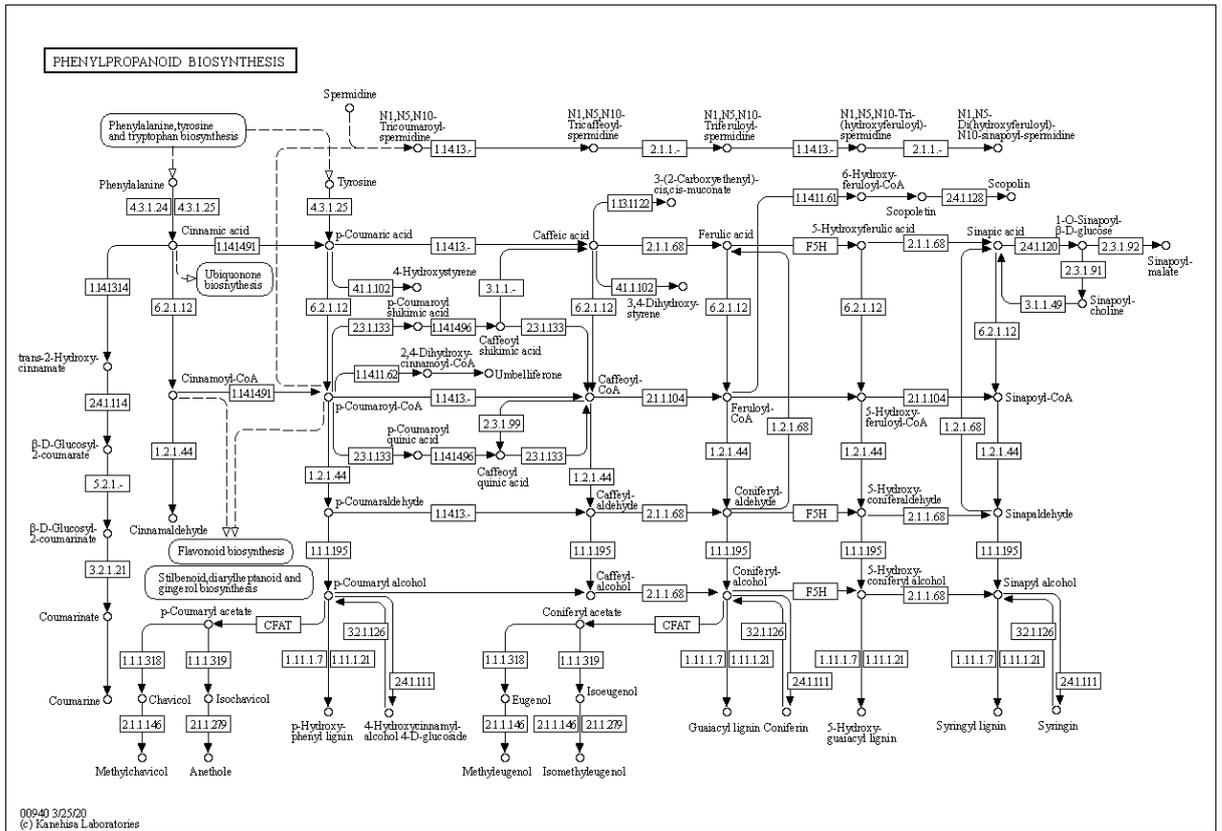


**Supplementary figure 1.** The pre-processing steps are as follows: (i) outlier screening is carried out to eliminate the LC-MS runs that deviate from the majority of the samples. Once the final sample set is determined, (ii) a filtering step is conducted that suppresses the noise while preserving the peaks in the data. (iii) A baseline correction algorithm is then selected to subtract the baseline from the raw signals. (iv) Peak detection is carried out so that each ion in each sample is represented as a peak. Since chromatographic column degradation and changes in the pH and temperatures of the environments and/or mobile phases can lead to rt drifts across samples, (v) peak matching and retention time alignment is an essential step that needs to be carried out to compare metabolomics data across samples. Following the identification of ions originating from the same metabolite, (vi) they are grouped together in the ion annotation step, and (vii) metabolite identification and annotation can then be done. The relative abundance of each metabolite needs to go through (viii) a normalization step to reduce the systematic variation of LC-MS data. The final step in the preparation of the metabolomics data for subsequent statistical analysis and biological interpretation is a step where (ix) data transformation and scaling techniques are implemented to produce more normally-distributed datasets with reduced heteroscedasticity.

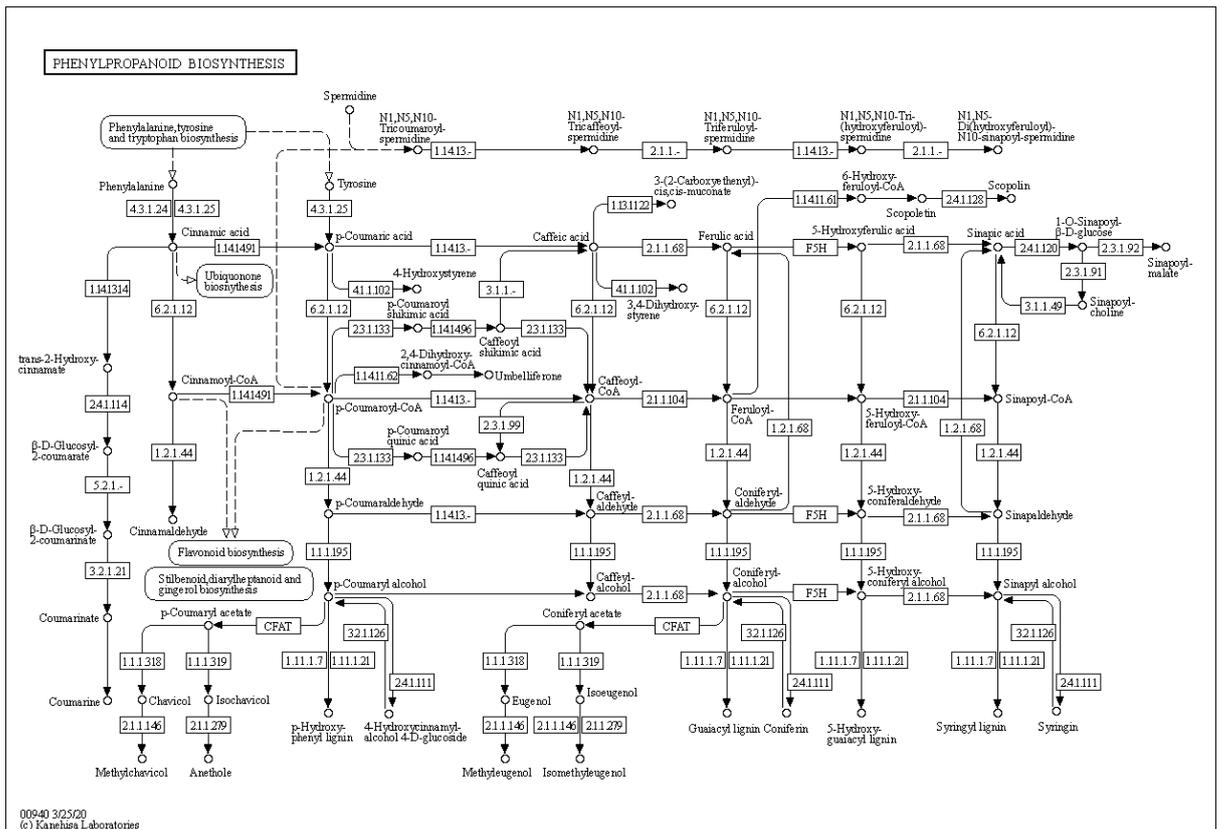




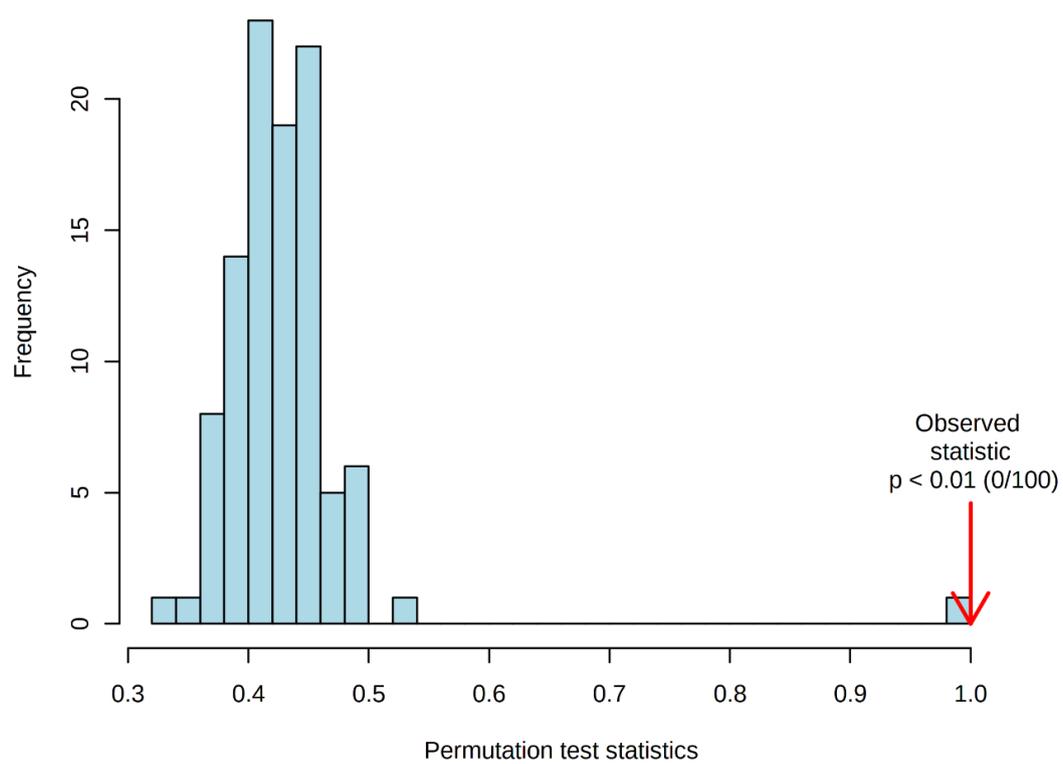
**Supplementary figure 3.** KEGG pathway of the phenylalanine, tyrosine and tryptophan biosynthesis pathway (map00400)



Supplementary figure 4. KEGG pathway of the phenylpropanoid biosynthesis pathway (map00940)



Supplementary figure 5. KEGG pathway of the flavonoid biosynthesis pathway (map00941)



**Supplementary figure 6.** Permutation analyses showing the statistical significance of the PCA and PLS-DA models generated and represented in Figure 7.

## Appendices – Supplementary tables

**Supplementary Table 1.** Sample lists containing sample descriptions of the tissues used for the metabolomics analyses. [https://drive.google.com/file/d/1LvcZBnwKaGDTjka0TPkoJ9AFuF7b\\_dsa/view?usp=sharing](https://drive.google.com/file/d/1LvcZBnwKaGDTjka0TPkoJ9AFuF7b_dsa/view?usp=sharing)

**Supplementary Table 2.** Metabolites detected in the leaf +1 (L1) tissues and their relative intensities. [https://docs.google.com/spreadsheets/d/1hzufQDiFEsmtA\\_bNTjx4fNQ5v38vzYBwliYzUSs8hg/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1hzufQDiFEsmtA_bNTjx4fNQ5v38vzYBwliYzUSs8hg/edit?usp=sharing)

**Supplementary Table 3.** Metabolites detected in the upper internode (I1) tissues and their relative intensities. <https://docs.google.com/spreadsheets/d/1pcfMfuOQPEAlAe62V42-fC8XPi9AhpzzRCkjh2INPs/edit?usp=sharing>

**Supplementary Table 4.** Metabolites detected in the young internode (I5) tissues and their relative intensities. <https://docs.google.com/spreadsheets/d/1884AVEhGndh5BNCrosEV3zrc191fld0UbbVpyRZ-Pg/edit?usp=sharing>

**Supplementary Table 5.** Metabolites detected in the mature internode (I9) tissues and their relative intensities. [https://docs.google.com/spreadsheets/d/1g\\_X052wai1-xjc0hTvmRjDaC2eqzYwKenOSQgr7JK50/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1g_X052wai1-xjc0hTvmRjDaC2eqzYwKenOSQgr7JK50/edit?usp=sharing)

**Supplementary Table 6.** List of 79 metabolites responsible for the separations between the four different anatomical tissues (VIP scores  $\geq 1.0$  from PLS-DA analysis).

Cluster	KEGG ID	KEGG name
1	C01175	1-O-Sinapoyl-beta-D-glucose
	C17210	2-(2'-Methylthio)ethylmalic acid
	C16405	Homoeriodictyol chalcone
	C01750	Quercitrin
	C00079	L-Phenylalanine
	C05577	3,4-Dihydroxymandelaldehyde
	C05576	3,4-Dihydroxyphenylglycol
2	C11950	3-Isopropylbut-3-enoic acid
	C05690	Se-Methylselenomethionine
	C00389	Quercetin
	C15606	1,2-Dihydroxy-3-keto-5-methylthiopentene
	C01134	Pantetheine 4'-phosphate
	C09751	(2R,3R)-3,4',7-Trihydroxyflavanone
	C00197	3-Phosphoglyceric acid
	C07086	Phenylacetic acid
	C00106	Uracil
	C01762	Xanthosine
	C01268	5-Amino-6-(5'-phosphoribosylamino)uracil
	C14865	2-(S-Glutathionyl)acetyl chloride
	C01069	Desulfoglucotropeolin
	C06196	dIMP
	C00341	Geranyl 2-methylbutyrate
	C00307	Citicoline
	C00364	5-Thymidylic acid
	C00365	dUMP
	C00253	Nicotinic acid
	C00379	D-Xylitol
C02637	3-Dehydroshikimate	
C00261	Benzaldehyde	

	C04443	3',4',5,7-Tetrahydroxy-3-methoxyflavone
	C01234	1-Aminocyclopropanecarboxylic acid
	C04713	N6-(delta2-Isopentenyl)-adenosine 5'-monophosphate
	C01852	Secologanin
	C00263	L-Homoserine
	C08650	Isoliquiritigenin
	C00061	Flavin Mononucleotide
	C20490	Cyanidin 3-O-beta-D-sambubioside
	C00423	trans-Cinnamate
	C03765	4-Hydroxyphenylacetaldehyde
3	C06118	4-(4-Deoxy-alpha-D-gluc-4-enuronosyl)-D-galacturonate
	C00072	Ascorbate
	C00388	Histamine
	C01210	N-Methylethanolamine phosphate
	C00719	Betaine
	C00262	Hypoxanthine
	C03373	5-Aminoimidazole ribonucleotide
	C16355	7-Methyluric acid
	C04181	3-Hydroxy-3-methyl-2-oxobutanoic acid
	C05399	Melibiitol
	C07090	Protoanemonin
	C17242	S-(4-Methylthiobutylthiohydroximoyl)-L-cysteine
	C12208	3-O-p-Coumaroylquinic acid
4	C00025	L-Glutamic acid
	C12249	Kaempferol-3-O-glucoside
	C00509	Naringenin
	C08604	Cyanidin 3-glucoside
	C16517	Indolylmethyl-desulfoglucosinolate
	C04431	cis-4-Carboxymethylenebut-2-en-4-olide
	C05909	Leucodelphinidin
	C02591	Sucrose 6-phosphate
	C05908	Delphinidin
	C05143	Dhurrin
	C00015	Uridine 5'-diphosphate
5	C00089	Sucrose
	C00315	Spermidine
	C01477	Apigenin
	C01617	Taxifolin
	C00482	Sinapic acid
	C02514	3-Fumarylpyruvate
	C00170	5'-Methylthioadenosine

---

C00544	Homogentisic acid
C00811	4-Hydroxycinnamic acid
C00242	Guanine
C05631	Eriodictyol
C01494	trans-Ferulic acid
C00148	L-Proline
C00127	Oxidized glutathione
C04302	N-(5-Phospho-D-ribosyl)anthranilate
C05623	Quercetin 3-glucoside

---

**Supplementary Table 7.** List of significantly expressed transcripts from all tissues (L1, I1, I5 and I9), collection points (C1, C2, C3 and C4) and fields (F1 and F2). The values are represented as log<sub>10</sub> intensity. [https://drive.google.com/drive/u/1/folders/1mntI2fnHH4f9chLf171opGH4\\_meyVwDu](https://drive.google.com/drive/u/1/folders/1mntI2fnHH4f9chLf171opGH4_meyVwDu)