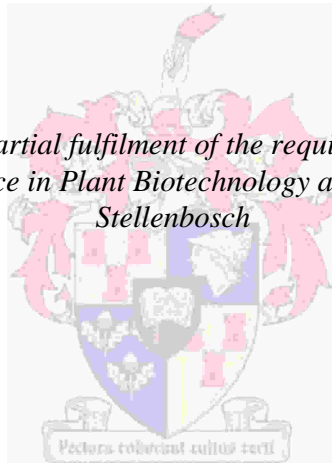# Correlating metabolite and transcript profiles in transgenic sugarcane lines

by
Riaan Neethling de Witt

Thesis presented in partial fulfilment of the requirements for the degree
Master of Science in Plant Biotechnology at the University of
Stellenbosch

Supervisor: Dr. Jan PI Bekker
Co-supervisor: Prof. Jens Kossmann
Faculty of Agricultural Sciences
Department of Genetics
Institute for Plant Biotechnology

November 2012

Declaration

By submitting this thesis/dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

RN de Witt

March 2013

Abstract

The current agricultural climate has initiated an increase in studies aimed at increasing crop yields and overall value. Sugarcane is a key crop in the sucrose production industry and a candidate for biofuel production. A number of transgenic lines have been produced with the aim of increasing the overall value of the crop *via* different approaches. Transgenic lines in which UDP glucose dehydrogenase (UGD) was repressed were previously shown to have increased sucrose content and altered cell wall composition. The aim of this study was to delve deeper into the mechanisms which confer the sucrose content increase by analysing transcript and metabolite profiles. This study found a number of genes to be significantly differentially expressed which have functions in sucrose metabolism, signalling and cell wall metabolism. Transcripts of the *myo*-inositol oxidation pathway and trehalose metabolism were found to be differentially regulated. Transcripts involved in lignin and flavonoid biosynthesis were identified as being down regulated in the transgenic lines which suggests alterations in lignin content to be present. It is suggested that the increase in sucrose is due to a decrease in sucrose breakdown rather than an increase in biosynthesis. This is supported by the identification of several transcripts representing cellulose synthase being down regulated which suggests decreases in sucrose utilization for cell wall polymer synthesis. The metabolome data confirmed the results from the transcriptional profiling by identifying *myo*-inositol as one of the primary discriminating compounds. Sorbitol and raffinose levels were shown to be altered in the transgenic lines which are presumed to be due to the increase in sucrose content. This study suggests future transgenic studies to consider the alteration of the *myo*-inositol oxidation pathway enzymes as a means to increase sucrose content.

Opsomming

Studies wat fokus op hoër gewas opbrengs en waarde is in die afgelope paar jaar grootliks beinvloed deur die huidige landbou bedryf klimaat. Dit het tot gevolg gehad dat meer studies in hierdie veld fokus. Ñ belangrike gewas in die die sukrose industrie, asook ñ kandidaat vir biobrandstof produksie is suikerriet. Menigte transgeniese suikerriet lyne is al geproduseer met die oog daarop om die sukrose stoor kapasiteit te verhoog deur verskillende metodes. In ñ vorige studie met hierdie doel in oog is UDP-Glukose dehidrogenase (UGD) af gereguleer in transgeniese suikerriet plant lyne. Hierdie transgenies suikerriet plant lyne het ñ hoër sukrose inhoud gehad as die kommersiële variteit waarvan die transgeniese plante produseer

is. Die doel van hierdie studie was om die meganismes agter die sukrose toename in diepte te bestudeer. Die transkriptoom was met behulp van Affymetrix mikromatriks plaatjies en die polêre metaboloom en sukrose inhoud deur NMR, GC-MS, HPLC en biochemiese analise bestudeer. Die study het ñ aantal transkripte gevind wat beduidend differensieel uitgedruk is wat funksies het in sukrose metabolisme, seinwerk en selwand metabolisme. Transkripte van die *myo*-inositol oksidasie padweg en trehalose metabolisme was gevind om differensieel uitgedruk te word in die transgeniese plante. Transkripte betrokke by lignien en flavonoied biosintese was af gereguleer in die transgeniese plante wat aandui dat daar moontlike veranderinge in selwand lignien vlakke mag wees. 'N moontlike rede vir die toename in sukrose word toegeksryf aan verminderde degradasie van sucrose en nie 'n op reguleering van sukrose sintese nie. Die aanname word bevestig deur die identifikasie van 'n aantal sellulose sintese transkripte wat af gereguleer was. Dit dui aan dat die gebruik van sukrose vir selwand polimere moontlik verminder is in die transgeniese plante. Die resultate vekry vanaf die metaboloom analises het die resultate van die transkriptoom analiese bevestig deur die identifikasie van *myo*-inositol as 'n primere diskrimineerende metaboliet. Die metaboliese data het aan gedui dat sorbitol en raffinose vlakke veranderd in die transgeniese plante is wat 'n moontlike effek is van die toename in sukrose. Hierdie studie stel voor dat die ensieme van die *myo*-inositol oksidasie padweg die potensiaal het om die sukrose inhoud van transgenies plante te verhoog indien die ensiem aktiwiteit gemodifiseer word.

Acknowledgements

Table of Contents

List of figures and tables

Abbreviations

AIR	Alcohol insoluble residue
bp	Base Pair
CTAB	Cetyltrimethylammonium bromide
CPGR	Centre for Proteomic and Genomic Research
EST	Expressed Sequence Tag
FC	Fold Change
FDR	False Discovery Rate
GC-MS	Gas Chromatography-Mass Spectrometry
HPLC	High Performance Liquid Chromatography
HSQC	Heteronuclear Single Quantum Coherence
kDa	kilo Dalton
LC-MS	Liquid Chromatography-Mass Spectrometry
MHz	MegaHertZ
MIOP	*Myo*-Inositol Oxygenation Pathway
MIOX	*Myo*-Inositol OXygenase
MIPS	*Myo*-Inositol 1-Phosphate Synthase
m/z	Mass-to-Charge ratio
NAD	Nicotinamide Adenine Dinucleotide
NMR	Nuclear Magnetic Resonance
PAD	Pulsed Amperometric Detector
PCR	Polymerase Chain Reaction
PLS-DA	Partial Least Squares Discriminant Analysis
PM	Perfect Match
ppm	Parts Per Million
RIN	RNA INtegrity
RMA	Robust Multichip Average
ROI	Region Of Interest
SASRI	South African Sugarcane Research Institute
SNOP	Sugar Nucleotide Oxidation Pathway
SPP	Sucrose PhosPhatase
SPS	Sucrose Phosphate Synthase
SQRTPCR	Semi Quantitative Reverse Transcriptase PCR

T6P	Trehalose-6-Phosphate
Tris	2-amino-2-hydroxymethylpropane-1,3-diol
SuSy	Sucrose Synthase
TPS	Trehalose Phosphate Synthase
TPP	Trehalose Phosphate Phosphatase
UDP	Uridine DiPhosphate
UDP-Glc	UDP-D-glucose
UDP-GlcA	UDP-glucuronic Acid
UGD	UDP-Glucose dehydrogenase
VIP	Variable Importance in Projection
WT	Wild Type

Chapter 1: General introduction and literature review

1.1 The importance of sugarcane (*Saccharum Spp.*)

Sugarcane is the major crop in the sugar production industry, responsible for more than 70% of world sucrose production with sugar beet second (Castleden *et al.*, 2004; McIntyre *et al.*, 2007; Suprasanna *et al.*, 2011; www.illovo.co.za). The sugar industry is one of the largest agricultural industries in Africa with South Africa producing the larger portion of the continent's sugar (Hassan, 2008; Snyman *et al.*, 2008). The demand to produce more sucrose from the same amount of plant material while at the same time reducing harvest, transport and processing costs is what drives the research efforts to produce higher sucrose yielding sugarcane lines.

Sugarcane is a C4 plant species and its production is localized to regions with tropical- and subtropical climates while cooler climates allow sucrose production *via* the C3 species, sugar beet (Sims *et al.*, 2006). In both crops sucrose content is about 15 percent of the dry weight, however, the costs involved in sucrose production from sugar beet are slightly higher compared to sugarcane (Sims *et al.*, 2006). Some producers cultivate both crops where the climate permits. Sugarcane plants reproduce vegetatively *via* rhizomes or nodal bud formation, but commercially grown sugarcane are vegetatively propagated by nodal stem cuttings (Lakshmanan *et al.*, 2005). Conventional breeding and biotechnological alterations have had success in creating higher sucrose yielding plants, but the potential of sugarcane is still largely untapped (Grof and Campbell, 2001). The current commercial sugarcane varieties have been produced exclusively by conventional breeding (Grof *et al.*, 2007; Suprasanna *et al.*, 2011). Estimations are that sugarcane can accumulate sucrose up to more than 25% sucrose per fresh weight (Bull and Glasziou, 1963). Current yields are about half of this estimate.

Sugarcane is not only an important crop in the sugar industry but finds application in a number of other industries. Paper, pharmaceuticals and several alternative products are derived from sugarcane which solidifies its position as an important multipurpose crop and bio-factory (Wang *et al.*, 2005). Due to the decrease in fossil fuel availability, the use of biofuels is considered as a replacement energy source (Pauly and Keegstra, 2008). A number of crops which including maize, sugarcane, sorghum and oilseed rape are currently being investigated or used for biofuel production (Gressel, 2008; Rubin, 2008; Papini-Terzi *et al.*,

2009). One of the key research areas in the domestication of crops for biofuel production is the plant cell wall which is a major carbon sink. The aim of several studies was to modify the plant cell wall lignin content to allow enzymatic digestion of cellulose or to increase cellulose content (Gressel and Zilberstein, 2003; Barrière *et al.*, 2004; Gressel, 2008; Pauly and Keegstra, 2008; Rubin, 2008; Sticklen, 2008).

Means to improve plant cell wall carbohydrate composition plays an important role in domesticating sugarcane towards a multipurpose energy crop, viable for both sucrose and biofuel production. Increasing structural polymers such as cellulose, which can be broken down to hexose and then further converted to biofuel reduces production costs from bagasse and serves as an alternative to the use of sucrose or maize starch as the primary raw material (Sims *et al.*, 2006; Gressel, 2008).

1.2 UDP-Glucose dehydrogenase and the competing pathways central to the synthesis of plant cell wall polysaccharides and sucrose

Uridine 5'-diphosphate-glucose (UDP-Glc) is the substrate for UDP-Glucose dehydrogenases (UDP-α-D-glucose:NAD$^+$ 6-oxidoreductase, UGD, EC 1.1.1.22) (Grof and Campbell, 2001; Grof *et al.*, 2007). In sugarcane UGDs are cytosolic proteins of 52 kilo Daltons (kDa) (Turner and Botha, 2002) which oxidize UDP-Glc to UDP-D-glucuronic acid (UDP-GlcA) together with the reduction of two NAD$^+$ molecules (Loewus and Loewus, 1983; Hinterberg *et al.*, 2002; Kärkönen, 2005) (Figure 1.1).

UGD activity has been reported predominantly in young, developing tissues such as immature internodes, root tips and nodules but also to a lesser degree in older tissues such as mature and maturing internodes (Seitz *et al.*, 2000; Johansson *et al.*, 2002; van der Merwe, 2006; Klinghammer and Tenhaken, 2007). Immature internodes are defined as meristematic and elongating internodes while mature internodes are defined as storage internodes (Bieleski, 1960; Lingle, 1989).

Amino acid sequences for UGD isoforms are highly conserved amongst plant species and show some similarity to animal sequences (Johansson *et al.*, 2002). *Arabidopsis* contains at least four UGD isoforms and one pseudogene (Klinghammer and Tenhaken, 2007) while there are two *UGD* isoforms known to be expressed in other plant species (Robertson *et al.*, 1996; Kärkönen *et al.*, 2005). Using a reference sugarcane UGD sequence (van de Merwe

14

2006), three additional putative UGD isoforms (UGD-1, -2 and -3) were identified in the TIGR-SoGI, from sugarcane (Bekker, 2007).

UDP-GlcA, the product of UGD, is a key metabolite in nucleotide sugar interconversion and a precursor of UDP-xylose, UDP-arabinose and UDP-galacturonic acid, cell wall biosynthesis and the production of hemicellulose and pectin polymers (Loewus *et al.*, 1962; Seifert 2004). UDP-glucuronic acid is also suggested to play a role in secondary cell wall synthesis (Kärkönen *et al.*, 2005). In plants, two independent pathways exist to produce UDP-GlcA from either UDP-Glc *via* the sugar nucleotide oxidation pathway (SNOP) or from *myo*-inositol (MI) *via* the myo-Inositol oxygenation pathway (MIOP) (Loewus and Murthy, 2000) (Figure 1.1).

Current opinion is that UDP-GlcA is primarily produced in pollen and seedlings *via* the MIOP while during later stages of plant development *via* the SNOP (Reiter and Vanzin, 2001; Pieslinger *et al.*, 2009; Pieslinger *et al.*, 2010). In the SNOP UDP-glucose is converted to UDP-D-glucuronic acid by the action of UGD. The inhibitor of UGD, UDP-xylose, is then produced by UDP-glucuronate decarboxylase (EC 4.1.1.35, UXS) from UDP-GlcA with the release of carbon dioxide (Harper and Bar-Peled, 2002). The SNOP product (UDP-GlcA) and downstream products (e.g. the cell wall polysaccharide precursor UDP-xylose) regulate the pathway by inhibiting UGD (Feingold, 1982; Stewart and Copeland, 1998) *via* feedback inhibition (Hinterberg *et al.*, 2002).

The MIOP involves a few additional enzymatic conversion steps to convert glucose-6-phosphate to UDP-glucuronic acid. Glucose-6-phosphate is first converted to myo-inositol-1-phosphate by L-myo-inositol 1-phosphate synthase (EC 5.5.1.4, MIPS). Myo-inositol-1-P is then converted to myo-inositol together with the release of phosphate *via* inositol-phosphate phosphatase (EC 3.1.3.25, IMPase). This reaction is followed by the synthesis of D-glucuronic acid and water *via* myo-inositol oxygenase (EC 1.13.99.1, MIOX) from myo-inositol and oxygen. D-glucuronic acid and adenosine triphosphate (ATP) is then converted to glucuronic acid-1-phosphate and adenosine diphosphate (ADP) *via* the action of glucuronokinase (EC 2.7.1.43). Glucuronokinase is inhibited by a number of substrates and products of the MIOP and the theory is that glucuronokinase is a glucuronate metabolism regulator which makes it another possible target for manipulation (Pieslinger *et al.*, 2010; Endres and Tenhaken, 2011; Reboul *et al.*, 2011).
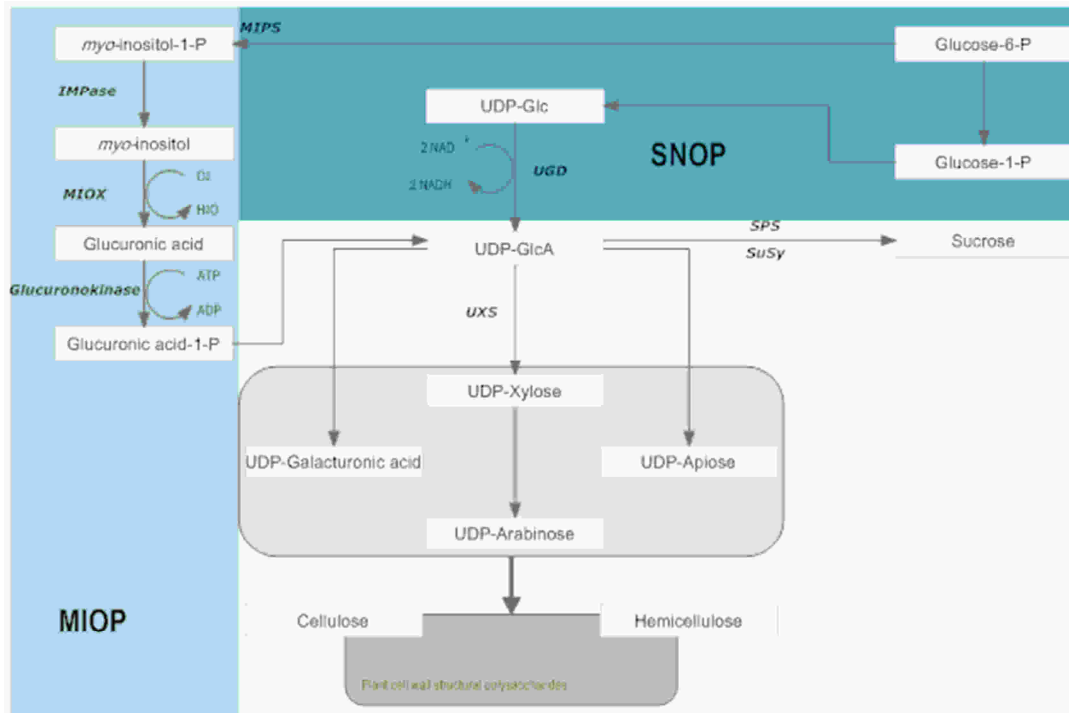
Figure 1.1: Representation of the SNOP and MIOP for cell wall structural polysaccharide synthesis. UDP-Glucose dehydrogenase converts UDP-D-Glucose to UDP-D-Glucuronic acid in the SNOP. Enzyme reactions of the SNOP are in the dark blue box. Glucuronate-1-phosphate uridylyltransferase converts D-Glucuronic acid-1-phosphate to UDP-D-Glucuronic acid in the MIOP. The MIOP is seen as a scavenger pathway and enzyme reactions of this pathway are indicated in the light blue box. The complete reaction scheme of sucrose synthesis *via* sucrose phosphate synthase (SPS) and sucrose synthase (SuSy) has been incorporated into a "single reaction" for simplicity reasons but are in fact two distinct pathways with multiple intermediates. In the complete reaction of SPS, UDP-glucose is metabolised to sucrose-6-phosphate and then to sucrose. Sucrose synthase incorporates UDP-glucose and fructose as substrates to produce sucrose and UDP. Modified from Oka and Jigarni (2006).

Glucuronokinase activity was first reported from *Phaseolus aureus* (mung bean) extracts (Neufeld *et al.*, 1959). Subsequent reports from *Lilium longiflorum* indicate glucuronokinase to be expressed and have activity in most plant tissues (Dickonson, 1982; Pieslinger *et al.*, 2009; Pieslinger *et al.*, 2010). The first complete gene sequence for glucuronokinase was generated from the *de novo* sequencing of peptides from *Lilium longiflorum* pollen (Pieslinger *et al.*, 2010). The reported sequence from *Lilium longiflorum* led to the identification and further caracterization of glucuronokinase in *Arabidopsis thaliana*.

Both the MIOP and SNOP coincide at UDP-glucuronic acid which is produced by glucuronate-1-phosphate uridyltransferase (EC 2.7.7.44) from uridine tri-phosphate (UTP)

16

and glucuronic acid-1-phosphate and the release of diphosphate (Figure 1.1) (Loewus and Loewus, 1983).

1.3 The plant cell wall building blocks

Sugarcane bagasse remains after sugar extraction and as already mentioned, has several downstream uses. The bagasse is primarily the remainder of the cell wall which consists of 40-50% cellulose and 25-35% hemicellulose (Sun *et al.*, 2003) and different portions of pectin and lignin (Pauly and Keegstra, 2008).

The plant primary cell wall consists of cellulose microfibrils, embedded in a matrix of hemicelluloses and pectins (Cosgrove, 2006). A cellulose microfibril is formed when several (1,4)-β-D-glucan chains crystallize, the bundling and crystallization of which is spontaneous (Cosgrove, 2006). Cellulose is produced by the enzyme cellulose synthase (2.4.1.12, CesA) of which a number of isoforms have been identified in plants (Pear *et al.*, 1996; Richmond and Somerville, 2000; Doblin *et al.*, 2010). Cellulose synthases are plasma membrane bound (Roelofsen, 1976; Brown, 1996; Pauly and Keegstra, 2008). The cellulose synthase-like (Csl) enzymes have also been implicated in cellulose (Richmond and Somerville, 2000; Doblin *et al.*, 2010) and hemicellulose synthesis due to the putative sugar binding protein motif shared between CesA and Csl enzymes (Cosgrove, 2006). This putative sugar binding motif is however not found in glucan synthase-like (GSL) proteins which are also involved in the synthesis of cell wall polysaccharides (Richmond and Somerville, 2000).

Studies have shown that during the maturation of sugarcane stem internodes, the expression of genes typically related to cell wall metabolism are altered (Casu *et al.*, 2007; Papini-Terzi *et al.*, 2009). Young internodes show increased cellulose synthase transcript abundance compared to mature internodes however, some cellulose synthase transcripts still show high transcript levels in mature stem tissues (Casu *et al.*, 2007; Papini-Terzi *et al.*, 2009). Cellulase transcript levels are speculated to be elevated in mature sugarcane internodes because the stalk cell wall requires modification even after elongation has finished as a result of sucrose accumulation (Casu *et al.*, 2007).

Hemicelluloses are the amorphous heteropolymers of monosaccharides such as xylose, arabinose and mannose (Sun *et al.*, 2003), which together with cellulose form the plant cell wall. The primary hemicelluloses found in plants are xyloglucans, arabinoxylans and mannans with xylans being the most abundant hemicelluloses found in plant cell walls (Sun

et al., 2003; Pauly and Keegstra, 2008). Pectins, unlike cellulose and hemicelluloses, function in the expansion of microfibrils to accommodate cell growth. The primary pectins in plant cell walls are rhamnogalacturonan I, homogalacturonan and xylogalacturonan, but several others exist (Cosgrove, 2006). As tissue matures the cell wall needs to expand and change to allow for the increased growth and in some cases change of cell function. Endo-(1,4)-β-D-glucanases have been hypothesized to function in primary cell wall loosening to allow for the rapid growth during tissue maturation (Cosgrove, 2000; Cosgrove, 2006). Endo-(1,4)-β-D-glucanases, together with expansins (Sampedro and Cosgrove, 2005) and xyloglucan endotransglycolase-/hydrolases form part of a group of enzymes called "wall loosening" agents which mediate the molecular alterations of the cell wall to reduce cell wall stress and allow for expansion (Cosgrove, 2006).

In order to synthesise the cell wall structural polysaccharides, glycosyltransferases utilise nucleotide sugars as glycosyl donors for polymer synthesis. These nucleotide sugars are synthesised from monosaccharide 1-phosphates by pyrophosphorylases (Kotake *et al.*, 2004). A number of nucleoside diphosphate sugars (NDP-sugars) can be synthesized *de novo* from UDP-Glc and GDP-mannose or *via* phosphorylated sugars by salvage pathways (Kotake *et al.*, 2004). Studies on *Arabidopsis* nucleotide sugar synthesis mutants (*ara* and *mur*) have shown that the availability of nucleotide sugars influences cell wall polysaccharide composition and the importance of salvage pathways as well as kinases and pyrophosphorylases in normal cell wall development (Bonin *et al.*, 1997; Kotake *et al.*, 2004; Kotake *et al.*, 2010). In *Arabidopsis*, the MUR1 gene encodes an isoform of GDP-D-mannose-4,6-dehydratase (EC 4.2.1.47, GDP-mannose dehydrogenase) which is involved in the conversion of GDP-mannose to GDP-fucose as observed in the xyloglucan cell wall component of the *mur1 Arabidopsis* mutant (Reiter *et al.*, 1993; Reiter *et al.*, 1997) which has reduced L-fucose residues (Bonin *et al.*, 1997; O'Neil *et al.*, 2001; Pauly *et al.*, 2001; Cosgrove, 2005). The MUR4 gene of *Arabidopsis* functions in the *de novo* synthesis of arabinose as seen by the 50% reduction in arabinose content in some organs in the *mur4* mutant which can be rescued by the exogenous application of arabinose (Reiter *et al.*, 1997; Burget and Reiter, 1999). The *mur4* gene encodes a UDP-D-Xylose 4-Epimerase localized to the Golgi apparatus (Burget *et al.*, 2003). The *Arabidopsis ara* mutant is deficient in arabinose kinase activity and show reduced arabinose content. However, little or no physiological effects on growth and development are evident in *ara* mutants (Dolezal and Cobbett, 1991). Current models suggest that the enzymes of the *de novo* and salvage

18

pathways finely regulate nucleotide sugar synthesis in a collaborative manner (Yamamoto *et al.*, 1988; Kotake *et al.*, 2004).

Lignin is produced by the phenylpropanoid pathway and is a major component of the secondary cell wall and flavonoids (Li *et al.*, 2010). The primary compound of the phenylpropanoid pathway is phenylalanine. Phenylalanine is derived from chorismate which is one of the products of the shikimate biosynthesis pathway (Vanholme *et al.*, 2010; Weng and Chapple, 2010). The lignin content of cell walls affect the enzymatic digestion of cell walls for use in industrial processes content as observed by several studies on the alteration of lignin content (Gressel and Zilberstein, 2003; Barrière *et al.*, 2004; Gressel, 2008; Pauly and Keegstra, 2008; Rubin, 2008; Sticklen, 2008). Lignin serves as a barrier to protect the cell wall components from microbial enzymatic digestion (Vanholme *et al.*, 2010). Li *et al.*, 2010, revealed that lignification is essential for the growth of plants in their study on *Arabidopsis*. The decrease of lignin content has in past studies led to reduced growth rates of plants. The split between flavonoid and lignin biosynthesis is at the *p*-coumaroyl CoA by the enzymes chalcone synthase (EC 2.3.1.74; CHS) and hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl transferase (EC 2.3.1.133; HCT) respectively (Figure 1.2). HCT has been identified as one of the key enzymes in the phenylpropanoid pathway (Hoffmann *et al.*, 2003). Flavonoids function as pigmentation, defence and light protection compounds and over 6000 flavonoids have been identified thus far (reviewed by Schijlen *et al.*, 2004). Flavonoids are of interest to crop improvement studies as they are beneficial to human health (Colombo *et al.*, 2006). Colombo and colleagues, 2006, reported flavonoid concentrations in sugarcane juice of 0.6mg.mL^{-1}. The first enzyme in flavonoid biosynthesis is CHS which produces naringenin chalcone which is converted by chalcone isomerase (EC 5.5.1.6; CHI) to form naringenin which is a flavanone. Naringenin chalcone can also spontaneously isomerize to naringenin (Holton and Cornish, 1995).
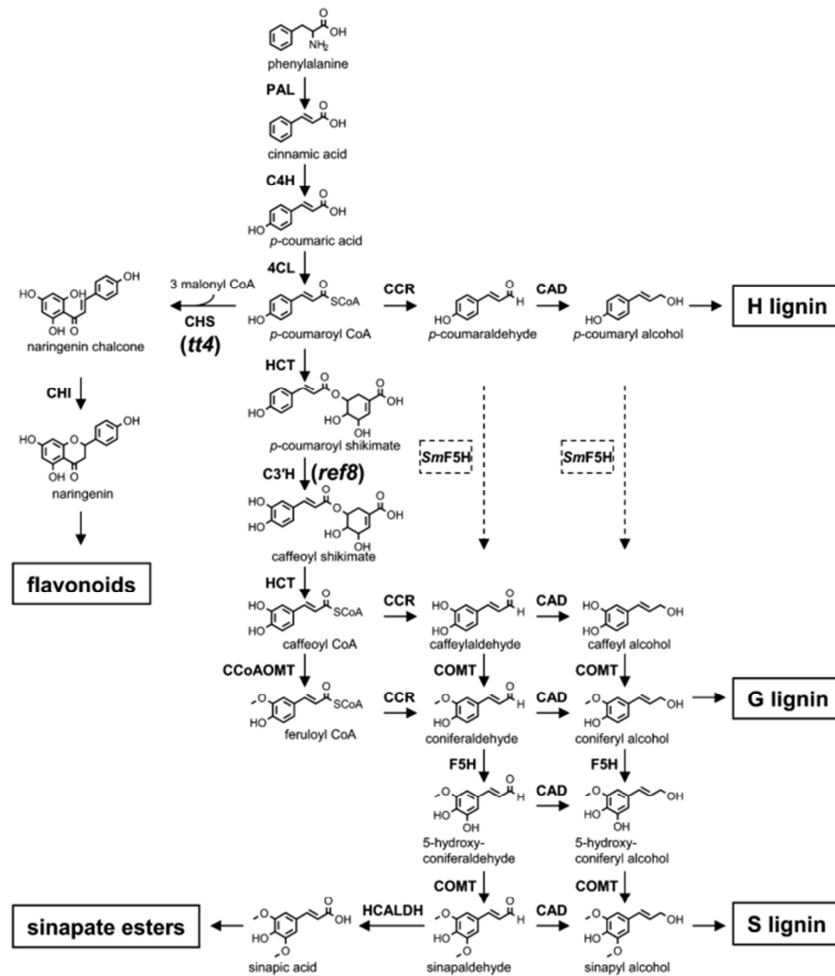
Figure 1.2: Flavonoid and lignin biosynthesis pathways (Li *et al.*, 2010). PAL: phenylalanine ammonia lyase; C4H; cinnamate 4-hydroxylase; 4CL: 4-coumarate:CoA ligase; HCT: hydroxycinnamoyl CoA:shikimate hydroxycinnamoyl transferase, C3'H: p-coumaroyl shikimate 39-hydroxylase, CCoAOMT: S-adenosyl-L-methionine:trans-caffeoyl-CoA 3-O-methyltransferase, CAD: (hydroxy)cinnamyl-alcohol dehydrogenase; CCR: cinnamoyl-CoA reductase; CHI: chalcone isomerase; COMT: caffeate O-methyltransferase

1.4 Sucrose phosphate synthase and sucrose synthase

Partitioning of carbon from UDP-glucose to cell wall polysaccharides is only one of the "source to sink" routes available for carbon. Another possibility is the pooling of carbon into the production of sucrose by the enzymatic action of sucrose synthase (EC 2.4.1.13, SuSy) and/or sucrose-phosphate synthase (EC 2.4.1.14, SPS). Both SuSy and SPS are key regulators of plant sucrose metabolism. Sucrose and its intermediates have multiple functions in plants which have made the genes involved in the metabolism thereof the focus of numerous studies aimed at increasing crop value throughout the development of improved crop varieties (Baxter *et al.*, 2003; Castleden *et al.*, 2004; reviewed by Sachdeva *et al.*, 2011).

In germinating seeds, sucrose is produced by SPS and sucrose-6-phosphate phosphatase (EC 3.1.3.24, SPP) from starch (Castleden *et al.*, 2004). SPS reversibly converts fructose-6-phosphate and UDP-Glc to sucrose-6-phosphate during sucrose synthesis in plants. Sucrose-6-phosphate is then irreversibly converted to sucrose by SPP. Several isoforms of SPS have been identified in dicots and monocots and found to be primarily expressed in various tissues in early developmental stages (Castleden *et al.*, 2004). SPS isoform expression levels have been shown to vary between tissue types in monocots (Grof *et al.*, 2006). The three SPS isoforms found in dicots have been reported to be expressed in different tissues and levels (Chen *et al.*, 2005).

Five SPS families have been identified in several monocot species (Castleden *et al.*, 2004). The five SPS families common among monocots are designated as A, B, C and D (divided into D_1 and D_2) (Castleden *et al.*, 2004). The D family has been termed as "unusual" due to the short linker region between the glucosyltransferase and sucrose-phosphatase-like catalytic domains in comparison to the A, B and C SPS families. The D family is the only SPS family found in monocots but not in dicots (Castleden *et al.*, 2004). The SPS D family has been reported to represent the larger portion of SPS expression in stem tissue (Grof *et al.* 2006). The SPS A family also show high expression levels in stem tissues. SPS families A, D1 and D2 have been reported to be the primary contributors to SPS expression in stem tissue in the day (Grof *et al.* 2006). SPS C is suggested to be the primary active isoform during the night (Grof *et al.*, 2006). The lack of regulation over the D family is speculated to be the reason for the accumulation of sucrose in the mature internodes of sugarcane plants (Grof *et al.*, 2006). The high expression of family A is attributed to the sequence similarity it shares with family D (Castleden *et al.*, 2004; Grof *et al.*, 2006).

SPS enzyme activity has been shown to increase with tissue maturity with a high correlation between sucrose accumulation and enzyme activity (Botha and Black, 2000). Other agriculturally important traits linked to SPS enzyme activity are plant height and yield (Zhu *et al.*, 1997; Ishimaru *et al.*, 2004).

Sucrose synthase is involved in the synthesis and breakdown of sucrose in plants but is also implicated in other functions such as phloem transport and cellulose synthesis (Geigenberger *et al.*, 1993; Amor *et al.*, 1995). The ratio of sucrose breakdown to synthesis by SuSy has been shown to increase as sugarcane internodes mature (Botha and Black, 2000; Schäfer *et al.*, 2004). This observation as well as the role SuSy plays in a number of other pathways

21

makes this enzyme a key target for further study and manipulation. While the current study focused on mature tissue, it should be noted that the enzyme is involved in the futile sucrose cycle in mature tissues (Moore, 1995).

1.5 Sugar signalling in plants

In short, sucrose is produced in the leaf photosynthetic tissues and transported by sucrose transporters to the apoplast. Here, importers transfer sucrose from the apoplast into the phloem which now mediates the sucrose transport to sink tissues where sucrose is unloaded for storage, either *via* sucrose transporters (SUTs) or plasmodesmata. Sucrose concentration gradients are thought to be responsible for the movement of sucrose from source to sink tissue (van Bel, 2003). Sucrose is then cleaved by invertases into hexoses for transport into storage cells (Patrick, 1997). The ratio of sucrose to hexose is hypothesized by Grof *et al.*, (2007) to indicate which of sucrose breakdown or synthesis is occurring in a tissue type. The movement of apoplastic sucrose out of the apoplast is deterred by the lignified cell walls around the vascular bundles (Welbaum and Meinzer, 1990) in mature tissue.

Sucrose can accumulate up to concentrations of almost 650mM in mature storage tissues in some sugarcane varieties (Welbaum and Meinzer, 1990). The storage of sucrose in such high concentrations have been shown to induce stress related responses as reported by the up-regulation of stress related genes in high sucrose yielding sugarcane varieties (Casu *et al.*, 2004). Transport and localization of sucrose are key interest fields as they serve as important regulatory points and may be targets for genetic manipulation to improve crop yields (Rae *et al.*, 2005).

Sugars have powerful signalling properties in plants (Paul *et al.*, 2001; Rolland *et al.*, 2006; Hanson *et al.*, 2008). Rolland *et al.*, (2006) review a number of studies in which transcription factors and amino acid related genes frequently have altered transcript levels when sugar concentrations are altered from untransformed NCo310 or unstressed plant levels. An example of this is seen by the up-regulation of asparagine synthesis when a plant is experiencing sugar starvation. Sugar starvation leads to the activation of primary metabolism, such as amino acid synthesis as secondary metabolism shuts down (Rolland *et al.*, 2006). Transcription factors are activated by changes in the cellular environment, such as increases in sucrose and hexose. In *Arabidopsis* the transcription factor bZIP11 has been reported to directly regulate *ASN1* and *ProDH2* which functions in amino acid metabolism (Hanson *et al.*, 2008). Opinion is that transcriptional regulation of hormone concentrations by sugars is

likely as demonstrated by the interaction of glucose with ethylene signalling (Rolland *et al.*, 2001; Price *et al.*, 2004).

A non-reducing sugar which has been implicated by several studies in signalling is trehalose (Reviewed by Paul, 2007 and Ramon and Rolland, 2007). Trehalose phosphate synthase (EC 2.4.1.15, TPS) and trehalose phosphate phosphatase (EC 3.1.3.12, TPP) are the two enzymes involved in the synthesis of trehalose from UDP-Glc and glucose-6-phosphate (G6P) (Paul *et al.*, 2001). While multiple trehalose synthesis pathways are known, eukaryotes only have one pathway namely, the TPS/TPP pathway (Figure 1.3) (Avonce *et al.*, 2006).
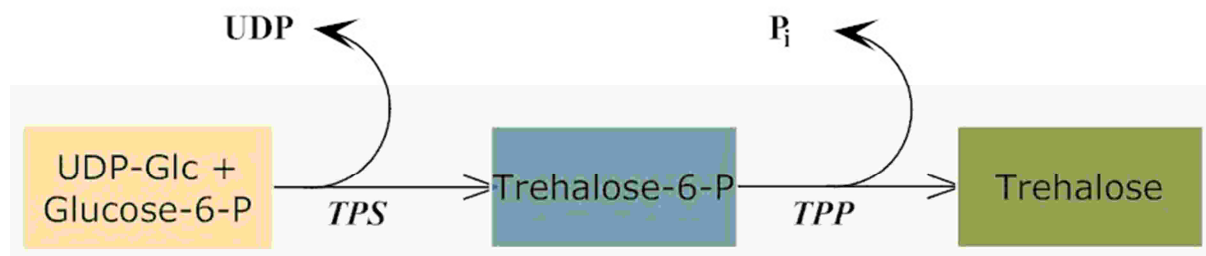
Figure 1.3: TPS/TPP pathway, adapted from Avonce *et al.*, (2006). TPS, Trehalose phosphate synthase TPP, Trehalose phosphate phosphatase.

Trehalose has in recent years seen much research interest as studies on model plant organisms yielded results potentially beneficial to crop plants when trehalose metabolism was altered (Pilon-Smits *et al.*, 1998; Garg *et al.*, 2002). Trehalose has been shown to regulate CHO metabolism in plants (Avonce *et al.*, 2005; Best *et al.*, 2011). In *Nicotiana tabacum* the inhibition of trehalose breakdown leads to the partitioning of carbon away from sucrose synthesis and into cellulose and hemicellulose (Best *et al.*, 2011). The current opinion is that in most plants, trehalose is not accumulated at significant levels (Garg *et al.*, 2002; Avonce *et al.*, 2006).

The raffinose family oligosaccharides (RFOs) are non-reducing sugars frequently found in higher plants and able to accumulate to high levels. Raffinose serves as the predominant soluble carbohydrate in plants as well as being a storage molecule in seeds (Kuo *et al.*, 1988) and has also been shown to be involved in cold desiccation tolerance (Keller and Pharr, 1996; Peters *et al.*, 2007; Peters and Keller, 2009). RFOs are synthesized from galactinol which is the product of galactinol synthase (EC 2.4.1.123, GAS, GolS) which uses the substrates, *myo*-inositol and UDP-galactose. A galactosyl moiety is transferred from galactinol to sucrose by the action of raffinose synthase (EC 2.4.1.82) to produce raffinose and *myo*-inositol. The

23

addition of another galactosyl moiety to raffinose by stachyose synthase (EC 2.4.1.67) yields the tetrasaccharide stachyose and *myo*-inositol (Figure 1.4). Thus far, no functions other than donating galactosyl moieties have been assigned to galactinol. Thus galactinol synthase is presumed to be a key regulator in RFO synthesis. Studies on *myo*-inositol 1-phosphate synthase identify this enzyme as a potential key regulator of galactinol and raffinose levels (Keller and Pharr, 1996; Keller *et al.*, 1998; Lehner *et al.*, 2008). RFOs are catabolized by α-galactosidases which remove the galactosyl moieties when there is no need for high levels of RFOs in a plant (Carmi *et al.*, 2003). The study by Carner *et al.*, 2004, showed that there is a possibility that the levels of *myo*-inositol coupled with sucrose levels affect RFO levels. Their study was aimed at explaining the inconsistent results of RFO content from studies focused on the over-expression of galactinol synthase. Their study on mutant Barley lines showed that the lines with increased RFO content also had an increase in sucrose, galactinol and *myo*-inositol levels.
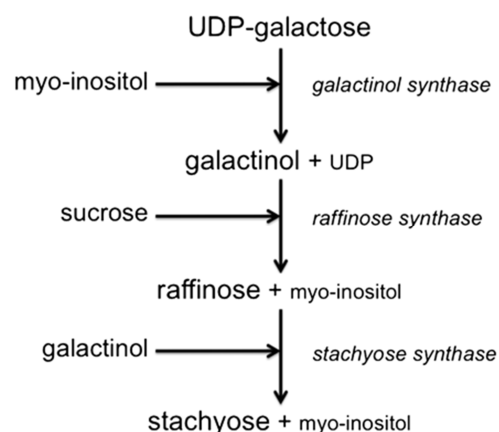
Figure 1.4: Raffinose biosynthesis pathway (Unda *et al.*, 2011)

Discussing sugar signalling in plants cannot be done without mentioning sucrose non-fermenting1 (Snf1) Related protein Kinases (SnRKs) and the central role they plays in sugar signalling and the coordination of carbon metabolism. Phosphorylation of SnRKs and other signalling proteins creates a motif for the binding of 14-3-3 proteins. The 14-3-3 proteins have been implicated in signal transduction pathways *via* their binding to signal-transduction proteins (Mhawech, 2005; Rolland, 2006). Their actions are affected by phosphorylation of either themselves or the binding partner (Aitken, 1996; Moorhead *et al.*, 1999). 14-3-3 proteins were first discovered as abundant brain proteins (Moore and Perez, 1967). Studies in plants show 14-3-3 proteins to directly bind to the phosphorylation site of nitrate reductase (NR) (Moorhead *et al.*, 1996). Sucrose-phosphate synthase and trehalose-6-phosphate

synthase have been shown to be 14-3-3 binding proteins (Moorhead *et al.*, 1999). The study by Cotelle *et al.* (2000) implicated 14-3-3 proteins in the metabolic changes occurring in sugar starved plant cells. Questions still remain about the function and action of 14-3-3 proteins and how they fit into the broader scope of plant signalling (Mhawech, 2005; Rolland, 2006).

1.6 Microarray technology and the analysis of transcript levels

Microarray technology has made possible the cost effective high-throughput analysis of gene expression profiles to elucidate gene function and interactions (Lange *et al.*, 2010). Microarrays generate a transcript level snapshot or expression profile of a sample which can then be compared to other expression profiles in order to determine how the two samples differ in their transcriptional profile. This is where the broad approach and high throughput of microarrays are at their most powerful, namely the screening of genes to find the key regulators. Genes identified as potential key regulators can then be further investigated in order to finally alter for organism improvement.

Microarray analyses rely heavily on statistical methodologies and a number of assumptions. The main assumption made when analysing microarray data is that the majority of gene transcript levels between two samples are the same and only a small number of gene transcript levels differ significantly. Due to this assumption, the masses of data created by the platform needs to be statistically analysed and validated by other techniques to determine if a gene is differentially expressed or not. There are other considerations and assumptions which need to be made which will not be discussed.

A number of array platforms exist to date and the Affymetrix GeneChip Sugarcane Genome Array (Affymetrix, Santa Clara, CA, USA) was selected for this study. This study focused on the robust power of the statistical language R (R Development Core Team, 2011) and its plethora of analysis packages available for free to conduct the analysis.

1.7 The plant metabolome and analysis techniques

Metabolites are of great importance when investigating the effect a variable has on an organism. Changes in metabolite levels can supply the data needed to create a hypothesis or describe how metabolite pathways interact and alter. As with transcriptome profiling, a

metabolite profile can be generated for a sample which is a representation of the sample metabolite levels for the particular variable under investigation.

Metabolite profiling is viable if a subset of metabolites; such as the polar or non-polar metabolites are investigated (Kruger *et al.*, 2008). A metabolite profile contains the exact quantities of the subset of metabolites in a sample. The profiles of different subsets from the same sample put together forms the metabolite profile of a particular sample under investigation (Kruger *et al.*, 2008). A simpler approach is metabolite fingerprinting in which statistical analyses are employed to compare metabolite data and identify underlying features which separate different samples into groups. Fingerprinting makes use of raw metabolite spectral data to identify features which differ between samples (Ward *et al.*, 2006; Scott *et al.*, 2010). These features can then be further analysed to determine the metabolites and their quantities within a sample (Lisec *et al.*, 2006; Kruger *et al.*, 2008). The investigation of metabolites compared with transcripts and proteins is simplified in that no prior sequence data is needed and the results are relatively less time consuming to obtain to name but a few benefits (Ludwig and Viant, 2009).

Several methods and technologies with which to investigate metabolites have been developed as the field has developed. Nuclear Magnetic Resonance (NMR) spectroscopy and Gas Chromatography-Mass Spectrometry (GC-MS) were used in this study due to the nature of the samples and the robust and versatile properties of these methods (Lisec *et al.*, 2006; Ward *et al.*, 2006).

NMR is both quantitative and non-degrading which sets it apart from other techniques such as GC-MS where the metabolites are destroyed and not available for subsequent analysis. Metabolites need to be derivatized for analysis by GC-MS to remove polar groups (Wilson and Walker, 2005, Ward *et al.*, 2006). GC-MS on the other hand is able to detect metabolites at far lower concentrations than NMR (Wilson and Walker, 2005; Lisec *et al.*, 2006; Wishart *et al.*, 2008). It is often the case that a metabolite is present in such low quantities that the changes in its concentration can't be measured by a method with too low sensitivity. This is especially true of metabolites which are toxic to an organism in relatively low amounts. GC-MS has also been more extensively used to determine metabolite profiles and thus variation in experimental setup is reduced due to the robust protocols and analyses available (Lisec *et al.*, 2006).

Liquid chromatography mass spectrometry (LC-MS) is another method extensively employed in metabolite studies. The method is sensitive and the analytes need no modifications such as derivitization which makes the analysis of non-volatile and temperature sensitive compounds possible with LC-MS. Both NMR and GC-MS can identify a broad range of metabolite classes but LC-MS has the more versatile range when compared to GC-MS. However, LC-MS is not as suited to the analysis of complex samples as NMR or GC-MS (Georgi and Boos, 2006).

1.8 The UGD repressed transgenic sugarcane lines

This study focused on transgenic sugarcane lines created at the Institute for Plant Biotechnology (IPB, Stellenbosch University, South Africa) (Bekker, 2007), in which UGD was repressed ('antisense' *UGD*). The research hypothesis was that reduction of UGD activity would alter carbon flux into sucrose. The commercial sugarcane (*Saccharum officinarum*) variety selected for the repression study was the *Saccharum* hybrid cultivar NCo310, first released in South Africa in 1945 (Nuss and Brett, 1995). Embryogenic calli were transformed *via* microprojectile bombardment with a vector containing a full length *UGD* cDNA sequence.

Northern blot and reverse transcriptase polymerase chain reaction (RT PCR) revealed several of the transgenic lines had lower transcript levels of *UGD* present than the untransformed NCo310 plants. The UGD repressed transgenic plants were grown under greenhouse conditions for further analysis. Several of the UGD repressed transgenic lines showed an increase in the amount of sucrose accumulated as well as alterations of cell wall components. The increase in sucrose was present in young, maturing and mature sugarcane internodes.

SPS and SuSy activity was investigated in order to determine the cause of the increased sucrose content in the transgenic sugarcane plants. SPS was shown to have increased activity in transgenic plants with increased sucrose content. Glucose-6-phosphate, an activator of SPS, was found to be increased in the transgenic lines. SuSy activity was however unchanged in young and mature internode material.

The downstream products of UDP-Glc were investigated in order to determine what effect the repression of UGD would have on cell wall composition. The uronic acid content was investigated to determine this. Against expectations, the uronic acid content was increased in the transgenic plants in both young and mature internode tissue. This finding initiated the

27

investigation into the MIOP enzymes of the transgenic sugarcane plants. The hypothesis at this stage was that the MIOP is responsible for the production of UDP-GlcA needed for the production of cell wall components as the SNOP output is reduced in the transgenic plants. A SQ RT PCR analysis of MIOX revealed that transcript levels up to five-fold higher than untransformed NCo310 plants. MIOX activity was shown to be increased in the leaf roll and maturing internode tissue. These findings supported the hypothesis that the MIOP is compensating for the repression of the SNOP (Charalampous and Lyras, 1957; Loewus *et al.*, 1962; Loewus and Murthy; 2000). The results also led to the assumption that glucuronokinase transcript and enzyme activity should in theory be increased (Labate *et al.*, 2008).

Further investigation of the cell wall glucose component revealed increases to be present in the transgenic plants. The increase in glucose was observed to be the result of an increase in the cellulose component of the cell wall.

1.9 Aims of the study

The sugarcane UGD repressed transgenic sugarcane plants created by Bekker (2007) showed an increase in sucrose content and an unexpected modification of the cell wall cellulose component which made the plants viable biofuel candidates. Bekker hypothesized that the MIOP was now the primary pathway responsible for the synthesis of UDP-GlcA which was supported by experimental results. Building on the knowledge gained by the study of Bekker (2007), this study aimed at improving our knowledge of sucrose metabolism in sugarcane.

In order to further improve sugarcane varieties, microarray technology was employed to construct transcript profiles of the transgenic sugarcane to be used in the screening for possible targets for future manipulation. Metabolite fingerprints generated from crude samples *via* NMR and GC-MS analysis was investigated as a possible means to aid in the search of candidates for manipulation when combined with transcript data.

Chapter 2: Transcriptional profiling of transgenic sugarcane lines with repressed UDP-Glucose Dehydrogenase activity

2.1 Abstract

The bagasse which remains after the sugar extraction processes from sugarcane stalks can be used as energy source in the form of liquid biofuel or in industrial purposes previously discussed. The demand from the sustainable energy industry (biofuels) is to develop crops with traits which enhance their suitability for use in biofuel production. A research focal point in this field is the manipulation of plant cell wall composition to facilitate low cost fermentation in biofuel production.

Transcriptional profiles of the transgenic sugarcane lines in which UGD was repressed were generated using microarrays. Several genes were selected for further investigation due to their potential to further increase sucrose yields. A transcriptional profile correlating to high sucrose accumulation was observed in the UGD repressed transgenic lines, in accordance with other reports that focused on improving sucrose yield in commercial sugarcane varieties (Watt *et al.*, 2005; Casu *et al.*, 2007). Several transcripts involved in cell wall alteration were also found to be significantly differentially expressed.

2.2 Introduction

2.2.1 Affymetrix microarrays

To investigate the gene expression profiles of UGD repressed transgenic sugarcane lines, GeneChip microarrays from Affymetrix were employed. The sugarcane GeneChip platform contains 8,236 *Saccharum officinarum* probe sets of which 6,024 are distinct genes (UniGene) from the Genbank UniGene build 5 (2004) mRNAs. 12 sugarcane control probe sets are also included of which four are hybridization-, five, poly-A- and three housekeeping gene controls. These controls are needed to determine if the hybridization procedure was successful.

Microarrays work on the principle that complementary single stranded oligonucleotides will bind to each other in a solution. Single stranded cDNA fragments or shorter oligonucleotides complementary to gene fragments are localized in an array on a support. Affymetrix arrays use 25mer oligonucleotide "probes", known as perfect match probes (PM), to represent each gene. In order to minimise variation in hybridization efficiency, each gene is represented by

10-20 oligonucleotides (Naidoo *et al.*, 2005). These oligonucleotides are designed by an algorithm to represent known or predicted open reading frames (ORFs). Control oligonucleotide probes which detect cross-hybridisation are included for all probes. The control oligonucleotides, known as mismatch (MM) probes, contain a single base pair change (Naidoo *et al.*, 2005). The location of each localized oligonucleotide and the gene it represents is annotated.

Affymetrix GeneChip arrays use a single colour detection system where biotin labelling of probes and subsequent staining with a streptavidin phycoerythrin conjugate produce the fluorescent signals where complementary nucleotide binding of the probes has occured. After this process the array is scanned for the signal intensities of all probes and compared to control array intensities. The average intensity for the entire set of probes recognising the same sequence is used to generate the signal intensity for the specific recognition sequence. The intensity of the signal is a measure of the amount of a transcript bound to the probe and thus also an indication of the amount present in a sample (Naidoo *et al.*, 2005). Changes in the intensities of the signals between control and sample reveal alterations in transcript abundance.

Normalization of arrays is essential for data comparisons amongst arrays and involves the flagging of features, background removal and hybridization efficiency calculations, the result of non-specific binding (Stekel, 2003). Normalization corrects for any signal errors or bias which may occur due to the experimental procedure. The type of normalization used depends both on the experimental design as well as the array type and manufacturer. Results from arrays need to be verified with, but is not limited to procedures such as semi-quantitative RT PCR, quantitative PCR (Q-PCR) or other omics data (Celis et al, 2000).

The experimental design of array studies greatly influences the quality of the data for downstream analysis. Replicates and pooling are two of the "grey" areas of microarray analysis. The advancement of microarray technology and the standardization of equipment and protocols have in recent years made the use of technical replicates less common in studies focused on biological variance (Göhlmann and Talloen, 2009). Biological replicates introduce more variance than technical replicates (Göhlmann and Talloen, 2009). The number of replicates used depends on the nature of the question which needs to be answered but the general consensus is, "the more the better".

2.2.2 Transcriptional profiling in plants

The redirection of carbon fixed by photosynthesis towards storage tissues and bypassing of other "sinks" is currently one of the primary focus points of manipulation studies (Grof and Campbell, 2001). Of particular interest are the enzymes involved in sucrose synthesis, transport and hydrolysis as well as sink intermediate metabolites (Grof et al., 2007). Carbon reserve partitioning in the plant cell wall are of importance for biofuel production as well as sucrose accumulation as the cell wall is a major carbon sink, the composition of which affects biofuel yields (Sticklen, 2008). Central to the production of most cell wall polymers is UDP-D-glucose (UDP-Glc) which is also involved in sucrose synthesis. The UGD transgenic sugarcane lines had a significant effect on cell wall composition and sucrose content which is due to the repression of UGD and the alteration of transcripts involved in sucrose metabolism, signalling and transport (Charalampous and Lyras, 1957; Loewus et al., 1962; Loewus and Murthy; 2000; Bekker, 2007).

The transcriptional profile of a plant will change as the conditions the plant is exposed to changes. Field trial and greenhouse grown plants will have significant differences due to the different environmental conditions. Some plant lines may perform well under greenhouse conditions but not under external environmental conditions (Bekker, 2007; Baafi and Safo-Kantanka, 2008; Panthee et al., 2012). Related to this is the developmental stage of the plant or tissue being investigated as the expression level of a gene can alter between developmental stages (Botha and Black, 2000; Castleden et al., 2004). It is for this reason that the whole network of transcripts related to a targeted alteration be investigated in order to investigate how developmental and environmental conditions alter transcriptional patterns and the link between them.

2.3 Methods

All chemicals were obtained from Sigma-Aldrich unless otherwise indicated.

Microarrays, chemicals and machinery required for the analysis were from Affymetrix.

Primers synthesized by Integrated DNA Technologies (IDT, Whitehead Scientific)

2.3.1 Plant material

Sugarcane (cultivar NCo310) was engineered to down-regulate UGD activity using an antisense construct (pAUGDf510; Figure 2.1) containing a full-length sugarcane *UGD* cDNA

sequence under the control of the constitutive maize ubiquitin and cauliflower mosaic virus 35S promoters. The *UGD* cDNA was identified from a previous research project (Bekker, 2007). Transformants were produced *via* micro projectile bombardment of sugarcane variety NCo310 embryogenic calli with the pAUGDf510 and pEmuKN vectors. The pEmuKN vector contains the antibiotic selection gene for neomycin phosphotransferase (*nptII*). A number of transformants were identified by PCR of either regenerating calli or young leaf material DNA. Plantlets were transferred to autoclaved soil to harden off when they reached 3cm in length. . Control plants (untransformed NCo310) did not go through the transformation process but was regenerated in the same manner and exposed to the same conditions as the transformed plants. Transgenics with increased sucrose content were then transferred to field trial conditions. Transgenic plants were grown in the eastern region of South Africa in proximity to the city of Durban, which, due to the sub-tropical climate, has year round optimal growth conditions for sugarcane.
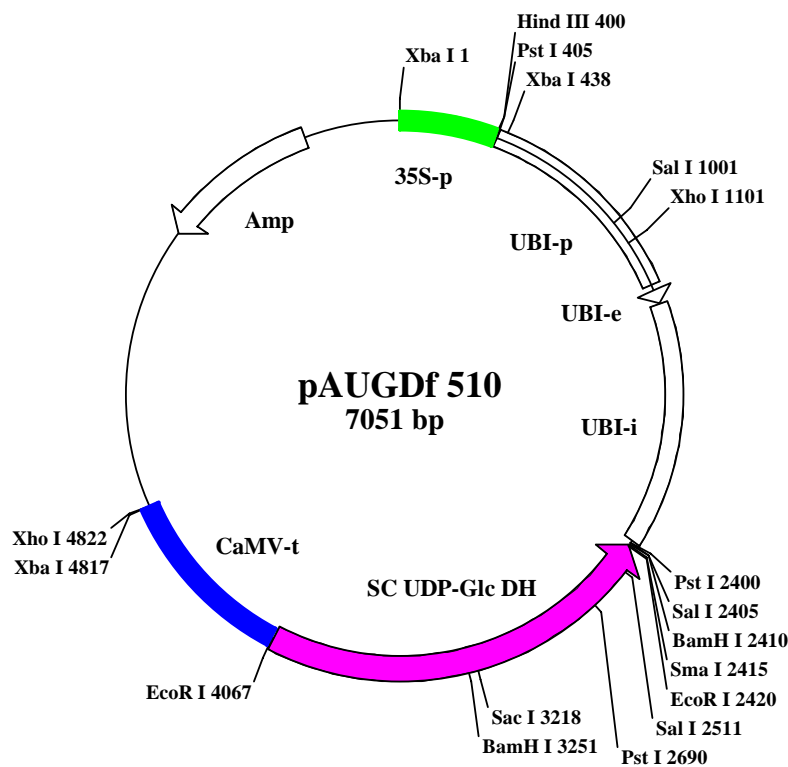
Figure 2.1: pAUGDf510 (7051bp) plasmid map. The plasmid is based on the pUBI510 plasmid. The plasmid contains the full length UGD gene insert of 1647bp in an antisense orientation. The CaMV terminator is used as termination sequence. Selection mediated by Ampicillin (bacterial) resistance on the second vector, pEmuKN, which is not shown. Vector map provided by JPI Bekker.

Plant material for transcript and metabolite analysis were from plants in their 5th growth cycle. The plants were cultivated on the terrace at the South African Sugarcane Research

Institute (SASRI) – (Mount Edgecombe, Durban, South Africa). Plants on the terrace are grown in large river-sand-containing containers able to facilitate the growth of a number of stalks which are exposed to monitored external environmental conditions.

Plant material used for the transcript analyses were 12 months of age (one growth cycle under the southeast environmental conditions of the KwaZulu-Natal province, South Africa) at the time of harvest.

Sampling occurred during September (spring, Southern Hemisphere) harvesting period of 2009. Plants were randomly harvested from the terrace containers at mid-day when sucrose is known to be most abundant in the stalk. Each transgenic sugarcane line was represented by three biological replicates which included a control set of wild type (WT) NCo310 stalks. Internodes 3-4 and 10 were selected from each cane sample and flash frozen in liquid nitrogen. Internode numbers were determined as previously described (Meinzer and Moore, 1988). Internodes 3-4 and 10 were classified as young and mature internodes respectively.

To verify if the UGD repressed transgenic lines retained their high sucrose content, sugar concentrations were determined *via* HPLC-PAD analyses (Chapter 3). Frozen internode material was ground using an analytical mill (IKA, A 11 basic, IKA®-Werke GmbH & Co. KG, Staufen, Germany). Three biological replicates were ground under liquid nitrogen and stored at -80°C until needed for transcript profiling.

The transgenic lines selected for the study were UGD 7, UGD 8, UGD 10 and UGD 18, all having showed increased sucrose content as indicated by the HPLC analysis (Figure 3.1). The control plants were wild type NCo310 plants.

2.3.2 RNA extraction

Total RNA was extracted from each sample using the CTAB (2% CTAB; 2% polyvinylpyrolidone (PVP); 100 mM Tris-HCL (pH 8.0), 25 mM ethylenediaminetetraacetate (EDTA), 2 M NaCl, 3% β-mercaptoethanol (added individually per sample just prior to extraction) method (White *et al.*, 2008). Ground material (200 mg) was added to 1.2 mL of CTAB buffer pre-heated to 65°C. Samples were vortexed for 30 s and then incubated for 30 min at 65°C with vortexing every 5 min. Two chloroform isoamylalcohol extractions were done and the RNA was precipitated by addition of 2 M LiCl for 16 h at 4°C. Samples were centrifuged (13 000 x g, 60 min). The resulting pellet was washed with 70% EtOH and centrifuged (13 000 x g, 10 min). The pellet was then resuspended in 20-50 µL nuclease free

33

water. A DNase (DNase I, RNase-free; Fermentas, Thermo Fisher Scientific Inc, USA) treatment followed to remove any DNA contamination as stated by the manufacturer's specifications. The RNA was quantified *via* Nanodrop spectrophotometer (Thermo Fisher Scientific Inc., USA) and analysed on 1% agarose gels stained with ethidium bromide to determine integrity. Transgenic lines which did not pass RNA integrity checks (UGD 8 and UGD 10) were omitted from further analysis as any degradation would influence results obtained from transcript profiling. Samples were stored at -80°C.

2.3.3 Affymetrix GeneChip hybridization

Affymetrix GeneChip Sugar Cane Genome Arrays (Affymetrix Inc., USA) were employed to investigate the transcriptional profiles of the different transgenic sugarcane lines at the Centre for Proteomic and Genomic Research (CPGR - Institute of infectious Diseases and Molecular Medicine, University of Cape Town, South Africa). The CPGR staff were responsible for the hybridization of RNA to the Affymetrix arrays. Only RNA from the mature internode (internode 10) samples was used in the microarray analysis. A complete sample set of 3 young internode (internode 3-4) biological replicates per line were not available for transcriptional analysis.

The RNA to be hybridised was re-quantified and analysed for degradation and impurities (salts and DNA) with both a Nanodrop and BioAnalyser (Agilent Technologies Inc., USA). The RNA samples were then amplified as per Affymetrix 3' IVT Express protocol together with a control to determine assay efficiency. 100 ng of complementary RNA (cRNA) was used for the assay as recommended by the manufacturer.

Biotin-labelled cRNA was synthesized *via in vitro* transcription of first and second strand cDNA. The quality of the amplified samples was assessed followed by fragmentation and a fragmentation efficiency test with the BioAnalyser. Successfully fragmented samples were hybridized to the arrays for 16 hours.

Following hybridization, the arrays were washed and stained using the Affymetrix GeneChip Fluidics Station 450 and stained chips scanned with the GeneChip® Scanner 3000 7G. Affymetrix Expression Console software v 1.1 was used for preliminary quality control and signal analysis. The .CEL files containing the average of a probe sets raw intensity values were supplied for further analysis. Additional quality control was done with Bioconductor (Gentleman *et al.*, 2004). Bioconductor packages used were affy (Gautier *et al.*, 2004),

Biobase (Gentleman *et al.*, 2004), limma (Smith, 2005), affylmGUI (Smyth, 2004; Wettenhall *et al.*, 2006), affyQCReport (Parman *et al.*, 2005), simpleaffy (Miller, 2005) and all dependencies

In total, nine arrays were used for the nine mature internode samples, each sample hybridized to a single array. The experimental design was such that each sugarcane line was represented by three biological replicates with no technical replicates included. The nine samples were hybridized in two batches. The first set of 6 hybridizations comprised two biological replicates per line. Samples of each batch were hybridized on the same day. The RNA samples were hybridized in random order to negate technical variability for each batch. The second batch of hybridizations followed several months after the first batch of mature internode hybridizations. The second batch was comprised of the remaining biological replicates of the mature internode samples.

2.3.4 Array quality control and data analysis

The statistical analysis was done at CPGR by Andrew Einhorn for the first data set which included 2 arrays per line (6 arrays in total). When the complete dataset was available a similar workflow was used to analyse the results from all the arrays.

Arrays which passed quality control were subsequently subjected to robust multichip average normalization (RMA) (Irizarry *et al.*, 2003). During this process background and interchip normalisation takes place as well as probe summarisation. The RMA normalization is applied in three steps which are, (i) PM probe specific correction with a model based on the sum of the signal and noise is equal to the observed signal, (ii) normalization of the corrected PM probes by quantile normalization (Bolstad *et al.*, 2003) and finally, (iii) median polish of values to calculation expression values.

A filter was applied to the data to separate the significant from the non-significant data. Differentially expressed genes were selected according to a modification of the p-value by the False Discovery Rate (FDR) algorithm (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001) of all probe values to identify if they were significant compared to the wild type. A $p < 0.05$ was defined as significant. The Bonferroni (Miller, 1981) correction was also applied to the unmodified p-values in a separate analysis.

After a list of differentially expressed probe sets were generated the normalised data was arranged according to probe set fold change values and the probe sets which showed large fold changes (fold changes larger or smaller than 1.5) were selected.

The probe sets of interest, identified by the filtering and statistical analysis, were then annotated *via* the National Centre for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/) and The Gene Index Project sugarcane index (SoGI), hosted by the Dana-Farber Cancer Institute (DFCI) available at (DFCI-http://compbio.dfci.harvard.edu/index.html), (SoGI-http://compbio.dfci.harvard.edu/cgi-bin/tgi/gimain.pl?gudb=s_officinarum), expressed sequence tag (EST) database using the Basic Local Alignment Search Tool (BLAST- http://blast.ncbi.nlm.nih.gov/) algorithm for similar sequences in cases where no annotation was supplied by Affymetrix. The probe set identifier was used to search for the probe set recognition sequence, annotation and design clusters *via* a NetAffx query (http://www.affymetrix.com/analysis/netaffx/index.affx) on the sugarcane array which was used in the BLAST searches. The annotation file created by Casu (2007) and a recent (2011) annotation by BLAST2GO (http://www.blast2go.com) was used to annotate the majority of transcript representative probes.

In order to investigate the expression patterns of a number of genes of interest in graphical format, the MapMan v3.5.0 BETA (Usadel *et al.*, 2005) software was used. The normalised probe set fold change values were used as input. A custom metabolic pathway image was created for the plant cell wall synthesis pathway and additional genes were added. The annotation file by Casu (2007) was used as annotation for the probe sets.

2.3.5 Array data validation

Semi-quantitative RT PCR was employed to validate the microarrays. cDNA was synthesized from 1µg of total RNA for each sample and used to optimize the SQ RT PCR conditions. The RevertAid[TM] H Minus first strand cDNA synthesis kit (Fermentas) was used for the reverse transcription reactions with oligo (dT) 18 primers. The cDNA reactions were set up in thin wall PCR tubes and the incubation done in a thermocycler as per the manufacturer's protocol. cDNA was stored at -80°C.

Expression data generated by the microarrays were validated for a number of selected genes (Table 2.1) which have representative probes on the microarray. These genes were selected based on the fold change values of their representative probe sets obtained from the

microarray data and gene function. The criterion for selection was a fold change value of higher than 1.5 or lower than 1.5 or genes involved in sucrose metabolism. Probe sets were selected for which annotation was available at the time of writing.

Actin was selected as a housekeeping control to normalize values from spot densitometry as actin is constitutively expressed (Huggett *et al.*, 2005; Li *et al.*, 2010). Negative (water) controls were included to determine non-specific amplification.

 A set of CTAB based RNA extractions were done on the different transgenic sugarcane lines as described previously. Samples were treated with DNase I (Fermentas), followed by cDNA (Fermentas) synthesis from a total of 1µg RNA per sample. The following primers were used for semi-quantitative RT-PCR (Table 2.1); Actin, which served as the template amount housekeeping control; Sucrose-phosphate synthase isoforms A, B, C, D1 and D2 (SPSA-D); Trehalose-6-phosphate synthase probe 2 (TPS probe 2) and TPS, Glucose 6-phosphate phosphate translocator (Glc6P PT) an F-Box protein and glucuronokinase. The primers for oxime monooxygenase, TPS, TPS probe 2, the F-box protein and the glucose-6-phosphat phosphate translocator were designed from the same sequences from which the microarray probes were designed (Casu *et al.*, 2007), available from the Affymetrix NetAffx (www.affymetrix.com) portal. The web based primer design software Primer3Plus (Untergasser *et al.*, 2007) was used in the design of the primers. The primers for SPS were taken from the study by Grof *et al.*, (2006). Actin and glucuronokinase PCR products were in previous studies in this lab shown to amplify the correct product *via* sequencing of the PCR product. Primer sets were optimized individually before full scale SQ-RT-PCR analysis.

Each PCR contained 2 µL of cDNA which was amplified with 1 unit of BIOTAQ (Bioline.Ltd) in the supplied reaction buffer. The primer, dNTP and $MgCl_2$ concentrations were as per manufacturer protocol in a 50 µL reaction volume. The PCR cycles were as follows: 94^oC for 2 min, followed by 94^oC for 30 s, (primer annealing temperature) for 30 s and 72^oC for 30 s (25-35 cycles). This was followed by a 72^oC final extension step for 5 min. The number of cycles were optimized for each primer set to ensure the amplification was within the exponential phase. Each PCR was repeated three times, once for each biological replicate. The SQ-RT-PCR was again repeated several months after the first set of PCRs for five of the primer sets (data not shown).

Table 2.1: Primer pairs used for Semi-quantitative RT-PCR validation of microarray results. The amplification sizes and the representative probe sets from the microarrays are shown below.

Primer name	Forward and Reverse primer sequence	Probe set	Product Size (bp)
Actin	**Fw:** TCACACTTTCTACAATGAGCT **Rev:** GATATCCACATCACACTTCAT		600
SPS A	**Fw:** TCTTCTCAAACCCCCGTAAGC **Rev:** CAGTTCACGATGTTCACCAAATG	SOF.2826.1.S1_AT	100
SPS B	**Fw:** GGCAGACCATAGCGAAGCTC **Rev:** CAGTTCACGATGTTCACCAAATG	SOF.119.1.S1_AT	100
SPS D2	**Fw:** CCATTGCGGATGCACTATATAAAC **Rev:** TTCGGGCCACGAGAATTG	SOF.5325.2.S1_AT	100
TPS2	**Fw:** CTGTCCATAGCTGATTCCTTACG **Rev:** GCTGCGACATGTGTCCTTT	SOFAFFX.974.1.S1_AT	195
TPS	**Fw:** AGCAAGGCCAAGTACTACC **Rev:** ACCAGAACATTCCTAGAGCA	SOF.3217.1.S1_A_AT	263
Oxime monooxygenase	**Fw:** ACTACTACGGCTCGCACTT **Rev:** GCGGCAGTAAAACACATAAT	SOF.83.1.S1_AT	396
F-Box protein	**Fw:** TCTACACGGTTCAATACGGCTCCCT **Rev:** ACTCCCTGGCCGCCAACAGC	SOF.2770.2.S1_AT SOF.2770.2.S1_X_AT	450
Glc6P PT	**Fw:** AGGAGCTCAGGAACACTGGAAGCT **Rev:** TCGCTGGTGATCCTCCTCCCC	SOF.3931.2.A1_S_AT	500
Glucuronokinase	**FW:** ATGCTGCAACGATGGCGGCC **Rev:** TCAGGCTCTCTCTTTTGAATCTCGG	SOFAFFX.529.1.S1_AT	580

The PCR products were analysed on ethidium bromide-stained 1% agarose gels. The PCR amplifications for the SPS transcripts were run on acrylamide gels as well as agarose gels due to the small amplicon. Each primer set produced only a single band. Gel images were created with a CCD camera (Alpha Innotech Corp). Intensity values were calculated with the AlphaEaseFC software (v 3.2.1 Alpha Innotech Corp) as an Integrated Density Value (IDV) and normalized against actin. Spot densitometry values were compared to the fold change values obtained from the microarrays to compare the results from the two methods.

2.4 Results

2.4.1 RNA extractions for microarray hybridization

A number of plant RNA extraction kits from several manufacturers were tested beforehand to determine RNA yield and quality and compared to the Cetyltrimethylammonium bromide (CTAB)-method (White *et al.*, 2008). The CTAB based method was preferred over kit-based RNA extraction methods.

Analysis of RNA using a Bioanalyser and Nanodrop revealed that all samples were of high quality and quantity and sufficient to be hybridized to microarrays. The RNA integrity (RIN) values from the Bioanalyser was approximately 8 for all samples. High RIN values indicate the presence of high quality RNA which is still intact. Degraded RNA samples will have low RIN values in the region of 3.

2.4.2 Array results

Signal histograms (Figure 2.2 A) of each array were generated and overlaid to investigate and compare probe intensity distributions. An RNA degradation plot (Figure 2.2 B) was drawn to determine if any degradation of RNA occurred. The combined results of the signal intensity histogram and the RNA degradation plot revealed that the RNA hybridization process was successful as no major deviations were observed in either of the signal histogram or RNA profiles. Boxplots (Figure 2.3 A; Figure 2.3 B) of the log expression values were generated before and after normalisation to determine if the normalization process was successful. Boxplots of the raw intensity values revealed, in agreement with the signal histogram and RNA degradation plots, that the hybridization process was successful. A batch effect is visible between the two hybridization sets as can be seen from the lower mean of all the second hybridization samples (blue boxplots) (Figure 2.3 A). The RNA quality control analysis was followed by an investigation into physical deviations of the arrays such bubbles or scratches. Again no physical damage or deviations were observed for any of the arrays. This concluded the RNA and hybridization quality control analysis with a pass for all arrays.
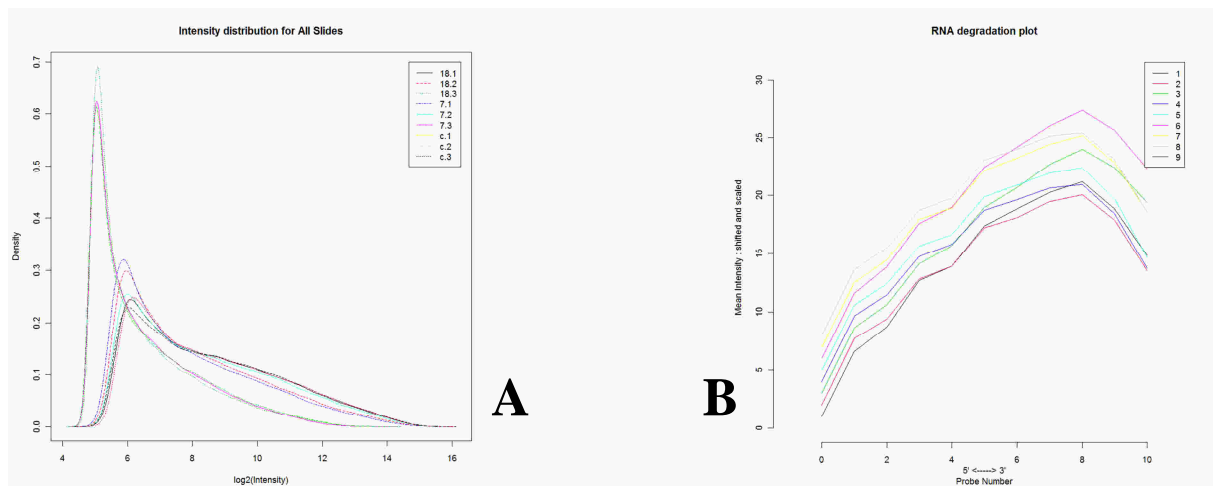
Figure 2.2: (**A**) Intensity distribution histogram for all mature internode microarrays of both the PM and MM probes. Intensity histograms illustrate the number of probes which have certain intensity. (**B**) RNA degradation plot of mature internode samples. RNA plots indicate if there are any differences in the RNA profiles amongst arrays. No significant deviations in RNA profiles were observed.
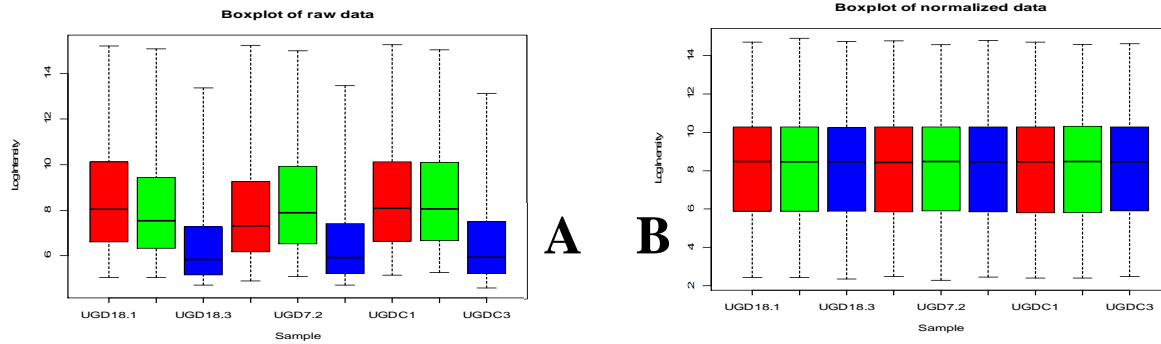
Figure 2.3: (**A**) Boxplots of the raw intensity values for each array in $\log_2$ space. The red and green boxplots represent the first hybridization event while the blue boxplots represent the arrays hybridized with the second event. (**B**) Boxplot of RMA normalized data. The plot indicates successful normalization.

A Pearson Correlation plot (Figure 2.4) was generated to determine the similarity between array biological replicates. The UGD 7 array set has the lowest (92%) and the untransformed NCo310 array set the highest (98%) correlation from the complete array set.
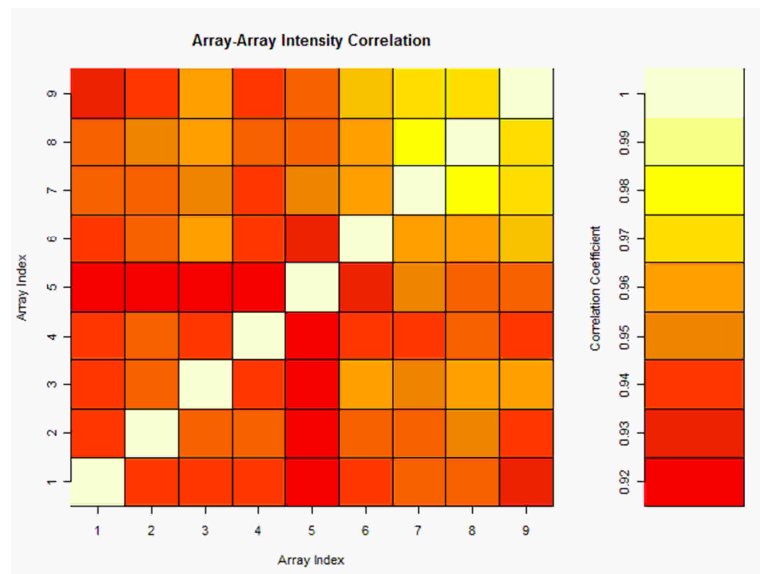
Figure 2.4: Pearson correlation plot of UGD mature internode microarrays. Arrays 1-3 are the replicates of UGD 18, 4-6 UGD 7 and 7-9 the untransformed NCo310 /Control.

A degree of inter-array variability was detected between biological replicate arrays. This is especially true for the two transgenic lines when investigation the fold change values of the 300 most differentially expressed probe sets (Figure 2.5 A) and a PCA plot thereof (Figure 2.5 B). The PCA plot indicates the second biological replicate of the UGD 7 array set does not cluster with the other two replicates from this set or the transgenic arrays.

Comparison of the expression profiles of UGD 18 and UGD 7 indicate that no gene up-regulated in one transgenic line is down-regulated in the other (data not shown). The Venn

diagrams (Figure 2.6) indicate the number of transcripts compared to the untransformed NCo310 line which are differentially up- or down-regulated and the number of transcripts similar between the two transgenic lines.

A FDR modification of the p-value combined with a 1.5 fold change threshold analysis of the RMA normalized data revealed 11 genes were significantly differentially expressed in UGD 7 and 98 in UGD 18 compared to the untransformed NCo310 line (Appendix Table A1). The data was also analysed with no FDR modification of the p-value or fold change cut-off value (Table 2.2). The rationale behind the removal of the fold change threshold is that some gene transcript levels do not show large deviations from the baseline but the enzymes they represent may still have high activity.
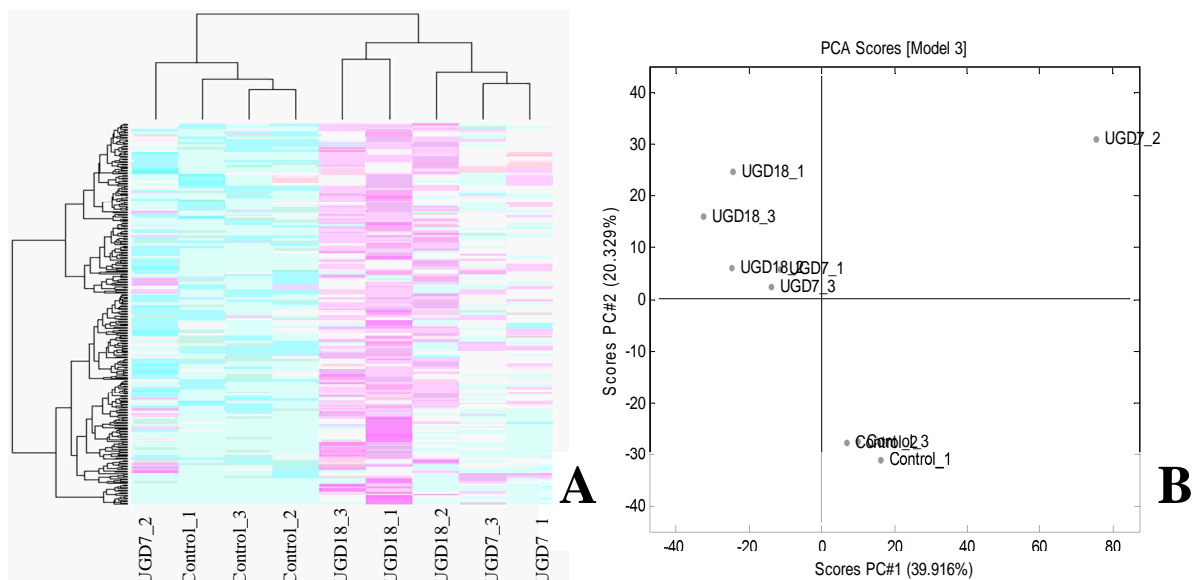
Figure 2.5: (**A**) Heatmap of the top 300 differentially expressed probe sets. Steel blue indicates low expression (transcript abundance) while pink indicates high expression. (**B**) PCA score plot of the top 300 differentially expressed, normalized probe sets.
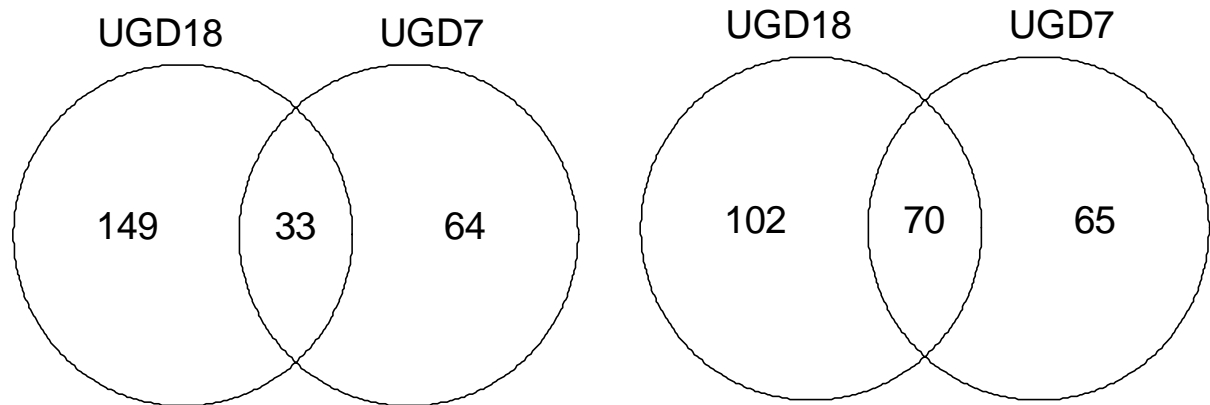
Figure 2.6: Venn diagram of up regulated (left) and down regulated (right) probe sets from transgenic vs. untransformed NCo310 lines. A fold change value of 1.5 was used as the cut-off value. In UGD 18, 182 probe sets were up-regulated and in UGD 7, 97 probe sets were up-regulated. Thirty three of the probe sets were similar between the two transgenic lines. From the UGD 18 dataset, 172 probe sets were down-regulated and from the UGD 7 dataset, 135 probe sets were down-regulated. Seventy of these probe sets were similar between the two transgenic lines.

Table 2.2: A subset of significantly differentially expressed probe sets from UGD 7- and UGD 18 against untransformed NCo310 normalized data comparisons. Probe sets are grouped according to molecular function of the transcript they represent. Only probe sets which could be annotated are displayed. FC, fold change; Group, UGD 7 or UGD 18; FDR 1.5 FC, probe sets significantly differentially expressed by FDR analysis with a 1.5 fold change threshold. Probe set is significantly different if the p-value < 0.05.

* Probe sets which are significantly differentially expressed as indicated by a FDR modification of the p-value and a 1.5 fold change cut-off threshold.

Annotation	Class	Probe set	FC	Group	FDR 1.5 FC
Cell wall					
Cellulose synthase (UDP-forming)	Cellulose synthesis	sofaffx.220.1.s1_at	-1.81	7	
Cellulose synthase (UDP-forming)	Cellulose synthesis	sof.5118.1.s1_at	-2.35	7	
Xylanase inhibitor	Cell wall modification	sofaffx.1646.1.s1_at	-3.01	18	
Xylanase inhibitor	Cell wall modification	sofaffx.220.1.s1_at	-1.74	18	
Alpha-1,4-glucan-protein synthase (UDP-forming)	Cell wall modification	sof.2854.1.s1_at	-1.80	18	*
Cellulose synthase (UDP-forming)	Cellulose synthesis	sof.3822.1.s1_at	-1.77	18	*
Cellulose synthase-like protein	Cellulose synthesis	sof.5118.1.s1_at	-2.43	18	
Cellulose synthase (UDP-forming)	Cellulose synthesis	sofaffx.220.1.s1_at	-1.74	18	
Xyloglucan endo-beta-1,4-glucanase	Cell wall modification	sof.5235.1.a1_a_at	-2.03	18	
Blight-associated protein p12 (Expansin)	Cell wall modification	sof.3338.1.s1_at	3.51	18	
Xyloglucan endo-beta-1,4-glucanase	Cell wall modification	sof.5217.1.s1_at	-2.71	18	
Xyloglucan endo-beta-1,4-glucanase	Cell wall modification	sof.3569.3.s1_a_at	-1.93	18	

Rhamnose biosynthetic enzyme	Rhamnose biosynthesis	sof.970.1.s1_a_at	-1.80	18	
Dtdp-glucose 4,6-dehydratase		sof.970.2.s1_at	-1.51	18	
Glucan endo-1,3-beta-D-glucosidase	Cell wall modification	sof.3550.1.s1_at	3.97	18	
Major CHO metabolism					
Tyrosine phosphatase	Starch degradation	sofaffx.40.2.s1_at	2.60	18	
Minor CHO metabolism					
Myo-inositol-1-phosphate synthase	*Myo*-inositol biosynthesis	sofaffx.699.1.s1_at	1.66	7	
Myo-inositol-1-phosphate synthase	*Myo*-inositol biosynthesis	sof.3427.1.s1_at	1.93	7	
Myo-inositol-1-phosphate synthase	*Myo*-inositol biosynthesis	sofaffx.699.1.s1_s_at	1.59	7	
Din10 (dark inducible 10) hydrolyzing o-glycosyl compounds	Raffinose synthesis	sofaffx.542.1.s1_at	-1.87	7	
Trehalose-6-phosphate synthase	Trehalose metabolism	sof.4709.1.s1_at	-0.74	7	
Trehalose 6-phosphate phosphatase	Trehalose metabolism	sof.3105.2.a1_at	-3.99	7	
Trehalose 6-phosphate phosphatase	Trehalose metabolism	sof.3653.1.s1_at	-2.20	7	
Trehalose-6-phosphate synthase	Trehalose metabolism	sofaffx.1973.1.s1_at	1.94	18	*
Trehalose-6-phosphate synthase	Trehalose metabolism	sof.3217.1.s1_a_at	1.59	18	
Trehalose 6-phosphate phosphatase	Trehalose metabolism	sof.3105.2.a1_at	-2.91	18	
Misc					
Beta-glucosidase isozyme 2 precursor	gluco-, galacto- and mannosidases	sof.1984.1.s1_at	3.86	7	
4-hydroxyphenylacetaldehyde oxime monooxygenase -- Cytochrome	Oxidation/reduction	sof.83.1.s1_at	2.11	7	*
Cell wall invertase		sof.512.1.s1_at	-2.04	18	
Endo-1,3-beta-glucanase	gluco-, galacto- and mannosidases	sof.4799.1.s1_at	2.75	18	
Beta-glucosidase isozyme 2 precursor	gluco-, galacto- and mannosidases	sof.1984.1.s1_at	2.49	18	
4-hydroxyphenylacetaldehyde oxime monooxygenase -- Cytochrome	Oxidation/reduction	sof.83.1.s1_at	2.51	18	*
NADPH--cytochrome p450 reductase	Oxidation/reduction	sof.2084.1.s1_at	-2.27	18	
Secondary metabolism					
Hyoscyamine 6-dioxygenase	Alkaloid biosynthesis	sof.116.1.s1_at	-2.14	7	
Hydroxycinnamoyl transferase (HCT)	Lignin biosynthesis	sof.122.1.s1_at	1.54	7	*
Chalcone isomerase	Lignin biosynthesis	sof.3905.1.s1_at	-1.68	18	
Chalcone synthase	Lignin biosynthesis	sof.3101.1.s1_at	3.66	18	
Chalcone synthase	Lignin biosynthesis	sof.702.1.s1_at	-2.53	18	
Hyoscyamine 6-dioxygenase	Alkaloid biosynthesis	sof.116.1.s1_at	-1.83	18	
Anthocyanidin reductase	Flavan biosynthesis	sof.572.1.s1_at	1.51	18	
Caffeoyl- o-methyltransferase 1	Lignin biosynthesis	sof.4258.1.s1_x_at	2.38	18	
Caffeoyl-coa O-methyltransferase	Lignin biosynthesis	sof.5198.1.s1_a_at	-1.60	18	*
Caffeoyl-coa O-methyltransferase	Lignin biosynthesis	sof.5198.2.s1_a_at	-1.69	18	*
Caffeoyl- o-methyltransferase 1	Lignin biosynthesis	sof.1122.2.s1_a_at	-1.54	18	
Hydroxycinnamoyl transferase (HCT)	Lignin biosynthesis	sof.122.1.s1_at	2.36	18	*
Hydroxyanthranilate hydroxycinnamoyltransferase (HHT)	Lignin/ Phytoalexin biosynthesis	sofaffx.409.1.s1_at	-1.54	18	

Laccase	Lignin biosynthesis	sof.4682.1.s1_at	-2.23	18	
Transport					
Glucose-6-phosphate phosphate translocator 2	Sugar transport	sof.161.1.s1_at	-2.38	7	
Glucose-6-phosphate phosphate-translocator	Sugar transport	sof.3931.2.a1_s_at	-2.39	7	
Glucose-6-phosphate phosphate translocator 2	Sugar transport	sof.161.1.s1_at	-3.83	18	*
Glucose-6-phosphate phosphate-translocator	Sugar transport	sof.3931.2.a1_s_at	-3.71	18	*

The analysis of the UGD 7 and UGD 18 array data revealed probe sets representing inositol-3-phosphate synthase (EC 5.5.1.4; MIPS or Myo-inositol-1-phosphate synthase) to be slightly up-regulated when compared to untransformed NCo310 (Table 2.2). MIPS synthesize the first precursor of the MIOP, *myo*-inositol 3-phosphate, which is used in the production of glucuronic acid as the pathways final product. Further investigation of the microarray data for transcripts of the MIOP found that transcripts of glucuronokinase were not significantly differentially regulated.

When investigating cell wall synthesis and degradation, probe sets representing transcripts of cellulose synthase and cellulose synthase-like genes were found to be highly down-regulated. Several probe sets representing transcripts involved in the modification of cell walls were found to be significantly differentially expressed.

Transcripts of lignin biosynthesis, representing hydroxycinnamoyl CoA:shikimate/quinate hydroxycinnamoyl transferase (EC 2.3.1.133; HCT) to be slightly up-regulated (Table 2.2). Probe sets representing enzymes further downstream of HCT namely caffeoyl CoA 3-O-methyltransferase (EC 2.1.1.104; CCoAOMT) were found to be slightly down regulated. A probe set representing Hydroxycinnamoyl-CoA:hydroxyanthranilate N-hydroxycinnamoyltransferase (EC 2.3.1.-; HHT) involved in phytoalexin biosynthesis was found to be slightly down-regulated. HHT uses various hydroxycinnamoyl-CoAs to catalyze the acetylation of hydroxyanthranilate in the synthesis of avenanthramides. The pathway is linked with the lignin biosynthesis pathway by coenzyme A (CoA) which is a precursor in both pathways. Laccase (EC 1.10.3.2) which is believed to be involved in cell wall lignification (reviewed by Mayer and Staples, 2002; Turlapati *et al.*, 2011) in higher plants was also observed to be highly down regulated. In the flavonoid biosynthesis pathway which branches out of the lignin biosynthesis pathway, probe sets representing chalcone synthase (EC 2.3.1.74; CHS) was highly up-regulated and down regulated while transcripts for the

enzyme directly downstream thereof, chalcone isomerase (EC 5.5.1.6; CHI), were slightly down-regulated.

When investigating sugar metabolism, a probe set representing rhamnose and raffinose synthesis (din10 hydrolyzing o-glycosyl) was found to be slightly down regulated (UGD 7 FC: -1.87). The DIN10 sequence from *Arabidopsis* was used as query sequence in a BLAST analysis to determine if the annotation was correct. The BLAST results indicated that the DIN10 protein had high sequence similarity with a number of stachyose and raffinose synthases but also to alpha galactosidases from *Arabidopsis* (Peters *et al.*, 2010).

Of interest are the transcripts representing trehalose 6-phosphate phosphatase which are down regulated while those of Trehalose 6-phosphate synthase are up-regulated. Both trehalose and trehalose-6-phosphate are known to be powerful signalling molecules of the plant sugar levels. The alteration of these enzymes suggests that the signalling of sugar levels in the transgenic plants may be altered which would allow for the accumulation of sucrose to such high levels. The majority of sugar transporters significantly differentially expressed had lower transcript levels in the transgenic plants than untransformed NCo310 (Appendix Table A1). Starch degradation was up-regulated while synthesis was down regulated which indicate the direction of carbon away from storage compounds. An investigation into sucrose metabolism found no SPS transcript levels to be significantly differentially expressed however, SuSy transcripts were found to be down regulated in both transgenic lines, but not significantly. It is known that relatively few genes involved in carbohydrate metabolism are expressed in the mature internode tissue of sugarcane (Casu *et al.*, 2003). Thus the increased sucrose content might be the result of decreased degradation rather than increased synthesis. However this remains to be verified and the increased sucrose content could also be due to both decreases in degradation and increased synthesis. Interesting is the highly down regulated transcripts of a Glucose-6-phosphate phosphate-translocator as G6P is involved in sugar signalling.

Probe sets which had a fold change value of larger (positive or negative) than 1.5 versus the untransformed NCo310 line was grouped according to their functional class annotation for the two transgenic lines (Table 2.3). Attempts at assigning annotations to probe sets with none were unsuccessful.

Table 2.3: The number of probe sets found to be differentially expressed between untransformed NCo310 and transgenic lines; grouped according to molecular function. Abbreviations: CHO: Carbohydrate, post trans: post translational modification, Up reg: up regulated, Down reg: down regulated.

Functional class	UGD 18 total	UGD 18 Up reg	UGD 18 Down reg	UGD 7 total	UGD 7 Up reg	UGD 7 Down reg
Amino acid metabolism	8	7	1	4	3	1
Cell wall	20	3	17	11	8	3
CHO metabolism	4	2	2	9	5	3
Hormone metabolism	6	5	1	3	2	1
Lipid metabolism	7	3	4	3	2	1
Protein degradation	11	8	3	8	5	3
Protein post trans	8	2	6	8	0	8
Secondary metabolism	11	5	6	2	1	1
Signalling	9	0	9	9	0	9
Stress	18	15	3	3	3	0
Transport	10	3	7	5	2	3
Unknown	186	92	94	123	30	93

2.4.3 RNA extraction and semi-quantitative Reverse Transcriptase PCR

An amplification product of the correct size compared to a product size standard was observed for all samples with no non-specific amplification products (Figure 2.7). The negative controls (water) did not amplify any products.

Figure 2.7: Agarose gel images of the semi-quantitative RT PCR amplifications for the specific primer sets. The biological replicates are grouped with untransformed NCo310 first followed by the UGD 7 group and finally the UGD 18 group of three.

Table 2.4: Semi-quantitative RT-PCR spot densitometry fold change compared to the microarray fold change results for untransformed NCo310 and transgenic UGD 7 and UGD 18 lines. SQRT PCR FC, semi-quantitative RT PCR fold change. Significant diffirences determined by the Student's t-test. Values calculated as mean $\pm$ SEM, n=3. * $p < 0.05$

Primer set	SQRT PCR FC UGD 7	Array FC 7	SQRT PCR FC UGD 18	Array FC 18
F-box protein	0.07	1.29	0.16*	2.92 *
Glucose-6-P translocator	-0.08	-2.39 *	-0.12*	-3.71 *
Glucuronokinase	-	-0.06	0.05*	0.14
Oxime monooxygenase	0.07	2.11 *	0.07	2.51 *
SPS A	-0.01	0.06	0.04*	0.12
SPS B	0.01	0.35	0.04	0.29
SPS C	-0.06	-	-0.03	-
SPS D1	0.01	-	0.01	-
SPS D2	0.03	0.00	0.04	0.15
TPS	0.02*	0.79	0.05*	1.59 *
TPS probe 2	0.02	0.57	0.05*	1.03 *

Of the different SPS family transcript levels, only SPS A of the UGD 18 line indicated a significant increase in transcript levels from the untransformed NCo310 line. The significant increase in SPS A transcript levels from UGD 18 is not observed in the microarray data. The remaining SPS transcript levels of the semi-quantitative RT PCR and microarray data are in accordance with each other with no significant changes in transcript abundance. No probe sets were found which represent SPS C or SPS D1 on the microarrays. SPS A transcript levels are known to increase from the leaves up to internode 7 while SPS B and C transcripts are primarily found in leaf tissues (Grof *et al.*, 2006). Thus, there could be reason to believe that SPS A transcript levels are increased in the transgenic lines which could lead to increased sucrose biosynthesis. Without enzyme activity data the assumption that UGD 18 SPS A transcript levels are increased based on the low sensitivity of semi-quantitative RT PCR, cannot be supported.

TPS, TPS probe 2, the oxime monooxygenase and the F-box protein probe sets indicate the transgenic lines to have increased transcripts compared to the untransformed NCo310 which is also observed in the semi-quantitative RT PCR results. The decrease in the transcript abundance of the glucose-6-phosphate phosphate translocator was observed for both of the datasets with UGD 18 showing the largest decrease when compared to the untransformed NCo310 line. The glucuronokinase results from the microarray data however did not indicate the same significant increase as the semi-quantitative RT PCR results. Numerous attempts to amplify a fragment from UGD 7 cDNA failed (Table 2.4).

It should be noted that the UGD isoform repressed by Bekker (2007) is not detectable by either western blot or RT-PCR in mature internodal tissues of sugarcane (van der Merwe, 2006; Bekker, 2007). For this reason the transcript levels of UGD were not investigated *via* semi-quantitative RT PCR.

2.5 Discussion

2.5.1 Transcriptome profiling

To gain a better understanding of the genes involved in the observed sucrose increase and modified cell walls of the UDP-glucose dehydrogenase repressed sugarcane plants (Bekker, 2007), microarrays were employed to generate transcriptional profiles. Two transgenic lines with the largest increase in sucrose levels and an untransformed NCo310 line (UGD 7, UGD 18 and NCo310 respectively) were selected for the transcript analysis. Nine Affymetrix GeneChip Sugarcane Genome Arrays were used to which three biological replicates per line were hybridized. A batch effect was present due to the extended timespan between the two hybridization events (Bolstad *et al.*, 2003; Chen *et al.*, 2011) as observed by the slight deviations in signal intensity plots (Figure 2.2 A; Figure 2.3). All quality control analysis revealed the hybridization events to be successful and the batch effect negateable after normalization *via* RMA (Figure 2.2 B; Figure 2.4) (Bolstad *et al.*, 2003; Ness, 2007; Chen *et al.*, 2011).

A heatmap and PCA plot (Figure 2.5 A; Figure 2.5 B) of the top 300 differentially expressed transcripts as well as a Pearson correlation plot (Figure 2.4) indicates the levels of a particular transcript will vary between biological replicates. This result is expected as the individual plants will differ from each other. A comparison of the UGD 7 biological replicates' transcriptional profiles indicates that the second biological replicate (UGD7-2) has an altered profile (Figure 2.5 A; Figure 2.5 B). The "change" in the transcriptional profile of UGD7_2 is suspected to be the result of post-harvest treatment or mechanical or biological damage during growth. The sample was retained as part of the UGD 7 transcript dataset based on the assumption that the key transcripts involved in the sucrose accumulation and cell wall modification would still be differentially expressed.

In a separate analysis of the microarray data (data not shown), the UGD7-2 biological replicate transcript data was excluded and the remaining replicates of UGD 7 combined with the UGD 18 replicates to create a "UGD transgenic" group which was then compare to the

untransformed NCo310 group. The rationale was that transcript levels altered by the repression of UGD would be the similar in both transgenic lines and thus key targets for manipulation could be identified in this manner when compared to the untransformed NCo310 line. The analysis however revealed that there was too much variation between the two transgenic lines to be group together as too few probe sets were identified to be significantly different. The transgenic lines were reported as all having undergone "unique transformation events" which was confirmed by DNA (Southern) Blots (Bekker, 2007). Differences observed in transcript levels may be due to the non-targeted insertion of the antisense construct *via* biolistics. Thus the insertion copy number and the location of insertion could be implicated in the differences between the transgenic lines (Kohli *et al.*, 1999, Butaye *et al.*, 2005).

2.5.2 Array validation and expression confirmation

In order to confirm the microarray transcript data were not influenced by the technique or methods used (Moreau *et al.*, 2003; Rockett and Hellmann, 2004; Adjaye *et al.*, 2004; Dallas *et al.*, 2005) semi-quantitative RT PCR analysis was done on a selected set of genes (Table 2.1). The transcripts to be analysed by semi-quantitative RT PCR were selected from the microarray data to include transcripts which were highly differentially expressed or did not deviate significantly from untransformed NCo310 transcript levels in order to reduce bias. Actin was used as the housekeeping gene for normalization primarily because primers for this gene were already available at the IPB and the product which these primers amplify known to be correct. Other housekeeping controls such as 16S and GAPDH were also considered.

In general the semi-quantitative RT PCR results (Figure 2.7; Table 2.4) positively correlated with the array results. The SPS transcript levels were all near untransformed NCo310 levels except for SPS C for which no representative probe could be found on the microarrays. No transcript found to be up-regulated in one dataset (microarray or semi-quantitative RT PCR) was down-regulated in the other and vice versa (Table 2.4), except for SPS A. The semi-quantitative RT PCR indicates a slight decrease (FC of -0.01) in SPS A transcripts in the UGD 7 transgenic sugarcane lines when compared to the untransformed NCo310 while the microarray data indicates a slight increase (FC of 0.06). The suspicion is that semi-quantitative RT PCR sensitivity is too low to accurately detect such small changes in transcript levels.

In overview, the semi-quantitative RT PCR results confirmed that the array results are an accurate measure of transcript levels and that the signals observed on the arrays were not due to experimental error or technique.

2.5.3 Genes of interest

In order to screen for probe sets which were differentially expressed between the two UGD repressed transgenic lines and the untransformed NCo310 line, a fold change filter of 1.5 was applied to the microarray probe set intensity data. A p-value threshold of $p < 0.05$ was then applied to identify probe sets which were significantly differentially expressed between the transgenic and untransformed NCo310 lines. A FDR modification of the p-value was applied to the p-values in an attempt to remove false positives. However, the number of probe sets which were found to be significantly differentially expressed with the FDR modification was too low to be of use and the FDR modification was removed. In total, 129 probe sets were found to be significantly differentially expressed in the UGD 7 transgenic line when the above filters without the FDR modification were applied. The UGD 18 line revealed 251 probe sets to be significantly differentially expressed. An analysis of probe sets up- and down regulated in the UGD transgenic lines indicated probe sets up-regulated in one line was not down regulated in the other and vice versa. The number of probe sets found to be differentially expressed was within the range of findings from other studies on sugarcane (Watt *et al.*, 2005; Casu *et al.*, 2007).

When the significantly differentially expressed probe sets of the two transgenic lines were compared, 33 up-regulated- and 70 down-regulated probe sets were found to be similar when a fold change cut-off of 1.5 and a p-value threshold of $p < 0.05$ (Figure 2.6) was applied. Of the 103 significantly differentially expressed probe sets similar to both UGD lines, only seven were identified to be significantly different from the untransformed NCo310 line and could be assigned a putative annotation when a FDR modification of the p-value was applied (Table 2.2). The deviation in the transcriptional profile of biological replicate UGD7-2 from the remaining UGD 7 replicates explains why few probe sets are differentially expressed in this transgenic line.

One of the transcripts highly up-regulated in both UGD repressed transgenic lines was that of 4-hydroxyphenylacetaldehyde oxime monooxygenase (CYP71E1) (EC 1.14.13.68) (Table 2.2), a cytochrome P450 which is involved in oxidation / reduction reactions and the production of the cyanogenic glucoside, dhurrin ((*S*)-4hydroxymandelonitrile β-D-glucoside)

(Dunstan and Henry, 1902). Dhurrin is believed to be involved in plant defence (Osbourn, 1996; Tattersall *et al.*, 2001). The glucose unit of UDP-Glc is cleaved and attached to p-hydroxymandelonitrile to produce dhurrin by cyanohydrin β-glucosyltransferase (EC 2.4.1.85) which prevents the release of hydrogen cyanide from (*S*)-4-hydroxymandelonitrile in the plant cell (Nielsen and Møller, 2000). Of note is a NADPH-cytochrome p450 reductase representing probe set which was found to be down regulated in one of the transgenic lines. Several of the lignin biosynthetic enzymes are known to by cytochrome P450s and the probe set mentioned above might be one of these enzymes, but cannot be confirmed without further evidence (reviewed by Weng and Chapple, 2010).

Of particular interest to this study are the transcripts representing cellulose synthase (UDP-forming), cellulose synthase-like and lignin biosynthesis transcripts which are in general down-regulated. Current opinion suggests that the pathways for secondary cell wall synthesis might be regulated by the transcript abundance of the genes involved in these pathways (Taylor, 2008). These findings, combined with down-regulation of cell wall modifying transcripts such as xyloglucan endo-beta-1,4-glucanase, glucan endo-1,3-beta-D-glucosidase, expansin and auxin associated proteins (Català *et al.*, 2000) suggest that the plant cell wall structure of the UGD repressed lines may be altered to allow for increased sucrose accumulation (Table 2.2; Table 2.3; Appendix Table A1). Redirection of carbon away from cell wall biosynthesis allows more carbon to be allocated to sucrose accumulation as sucrose is used as substrate for cell wall biosynthesis (Papini-Terzi *et al.*, 2009). This data is supported by other studies which employed microarray technology to investigate transcript levels of high sucrose accumulating sugarcane lines (Casu *et al.*, 2007; Papini-Terzi *et al.*, 2009).

The down-regulation of enzymes involved in hemicellulose breakdown such as xyloglucan endo-beta-1,4-glucanase and a xylanase inhibitor hints that the xylose fraction of hemicellulose may be increased. However, no transcripts representing xylose synthesizing genes were found to be significantly differentially expressed to support the speculation mentioned above.

Interesting is the finding that transcripts representing caffeoyl-coa O-methyltransferase which is involved in lignin biosynthesis is down regulated 1.6 fold in UGD 18. This finding is in contradiction to a similar study (Casu *et al.*, 2007). The down regulation of cellulose synthase and cellulose synthase-like transcripts (Table 2.2) in this study is again not observed by other

reports (Casu *et al.*, 2007). The observed differences are most likely the result of the repression of UGD and not the increase in sucrose content. Two chalcone synthase representing probe sets are found to be significantly differentially expressed, one being up-regulated 3.66 fold in UGD 18 and the other down regulated -2.53 fold in UGD 18. The enzyme downstream thereof, chalcone isomerase was down regulated -1.68 fold in the same transgenic line (UGD 18). It has been suggested that there is competition between the flavonoid and lignin biosynthetic pathways for a common substrate (4-coumaroyl-CoA) due to the up-regulation of both CHS and HCT which are the key enzymes of the two diverging pathways (Weng and Chapple, 2010). Enzymes downstream of both CHS and HCT are however down regulated. Several studies have shown the down-regulation of one gene involved in lignin biosynthesis leads to the down regulation of lignin biosynthesis but also alters other metabolite pathways (Rohde *et al.*, 2004; Sibout *et al.*, 2005; Dauwe *et al.*, 2007; Li *et al.*, 2008; Nakashima *et al.*, 2008; Vanholme *et al.*, 2008; reviewed by Vanholme *et al.*, 2010). Silencing of HCT in *Arabidopsis* leads to a reduced size phenotype due to flavonoid accumulation *via* redirection of metabolic flux into flavonoid biosynthesis (Besseau *et al.*, 2007). This phenotype was not observed in the UGD repressed transgenic lines. Another enzyme which forms part of the phenylpropanoid pathway is hydroxycinnamoyl-CoA:hydroxyanthranilate N-hydroxycinnamoyltransferase (HHT) which uses *p*-coumaroyl-CoA or feruloyl-CoA as substrate to produce phytoalexins which are believed to be involved in plant pathogen defence (Yang *et al.*, 2003). A probe set representing HHT was found to be down regulated -1.54 fold in the UGD 18 transgenic line. The down regulation of any HHT transcripts was however not observed in transgenic line UGD 7. The down regulation of lignin synthesizing genes has been shown to increase the fermentability of lignin containing tissues during biofuel production (Chen *et al.*, 2007). However, lignification is essential for normal plant growth and development in *Arabidopsis* (Li *et al.*, 2010). The hypothesis is that an alternate biochemical pathway exists to produce lignin (Li *et al.*, 2010). The synthesis of hemicellulose is also altered as observed by the decrease rhamnose biosynthesis and xyloglucan endotransglycolase transcript levels.

The down regulation of transcripts representing cell wall modifying enzymes indicates that no further deposition of cell wall polymers is occurring as cell expansion is reduced which supports the assumption that cell wall synthesis is reduced (Cosgrove, 2000; Li *et al.*, 2003). Callose synthase and endo-1,3-beta-glucanase transcripts were up-regulated while callose isomerase transcripts were down-regulated in the mature internodes. The down-regulation of

callose isomerase is expected as there is no growth in mature sugarcane tissues (Jacobsen *et al.*, 1992), which is in accordance with the results from Casu *et al.*, 2007. They do however state that the finding of callose synthase being up-regulated is unexpected. The finding that transcripts of endo-1,3-beta-glucanase which removes callose from the cell wall is up-regulated is also in concurrence with the results from Casu *et al.*, 2007. Sucrose accumulation is believed to cause osmotic damage to cell walls which leads to the up-regulation of cell wall biosynthesis genes in order to repair the damage (Casu *et al.*, 2007). This hypothesis is supported by the down regulation of polymer depositing and cross linking enzymes and the up-regulation of an expansin transcript in this dataset. It is assumed that the cell wall needs to expand and loosen in order to accommodate the increased sucrose amounts.

Trehalose 6-phosphate phosphatase transcripts were highly down regulated in both transgenic lines (Table 2.2) while trehalose 6-phosphate synthase was up-regulated only in UGD 18. Trehalose levels have been reported to positively correlate with sucrose levels (Glassop *et al.*, 2007). Trehalose can accumulate to near 5 nmol.g^{-1} in mature sugarcane internodes (Bosch, 2005). The increase in sucrose content also increases trehalose content and thus the transcriptional regulation of trehalose biosynthesis. Together with a number of MAP kinase probe sets found to be highly down regulated in both transgenic lines (Table 2.2; Appendix Table A1), signalling pathways seem to be altered and essential for sucrose accumulation (McCormick *et al.*, 2008).

It has been suggested that the sucrose concentration of the cytosol is mediated by T6P to the chloroplast (Pellny *et al.*, 2004; Paul, 2007). The genes involved in the sugar signalling pathway and trehalose synthesis have been identified as potential targets to release the limiting mechanism in sucrose production (McCormic *et al.*, 2008). Reports indicate genes related to starch metabolism to be up-regulated due to changes in the ratio of TPP:TPS transcript levels in sugarcane plants with increased sucrose content (McCormic *et al.*, 2008).

An interesting finding is the down regulation of a probe set annotated as Din10 (dark inducible 10). In *Arabidopsis thaliana* the *din* genes (Azumi and Watanabe, 1991) are induced by darkness and affected by the plant sugar levels (Fujiki *et al.*, 2000, Fujiki *et al.*, 2001). *Din*10 is classified as a seed imbibition protein due to its similarity to a protein from *Cicer arietium* found in imbibed seeds (Fujuki *et al.*, 2001). However, the two proteins also share similarity with stachyose- and raffinose synthases from other plants (Fujiki *et al.*, 2001). *Arabidopsis AtSIP2*, previously thought to be a raffinose synthase was recently

functionally characterised and found to be an alkaline alpha-galactosidase specific for raffinose (Peters *et al.*, 2010). Based on similarity from amino acid sequence alignments, DIN10 would appear to be most likely an alpha galactosidase. Due to the annotation of this probe set in the current study, this probeset has been flagged to be a possible raffinose-, stachyose synthase or an alkaline alpha-galactosidase. Alterations in raffinose levels are observed in the UGD repressed transgenic lines as discussed in chapter 3 which supports this decission.

Raffinose has been shown to accumulate in sugarcane mature internode tissues up to a maximal concentration of about 20 mg.g^{-1} (fresh weight) (Glassop *et al.*, 2007). The physiological role of raffinose in sugarcane is still a matter of speculation but it could legitimately serve as a storage compound or have a function in signalling (Glassop *et al.*, 2007). Raffinose has been reported to positively correlate to regards to sucrose concentration in sugarcane (Glassop *et al.*, 2007) which is observed in the metabolite profiling of the UGD transgenic plants (chapter 3). The reason why a raffinose synthase would be transcriptionally down regulated in a sugarcane line with increased sucrose when trehalose biosynthesis is up-regulated is difficult to explain. There is reason to believe that this probe set as with the *Din*10 annotated probe set could be mis-annotated. It could be possible that enzymes involved in raffinose break down are down regulated but not annotated. If *DIN10* is an alpha-galactosidase then the down regulation of this enzyme and the increased raffinose levels would be in conformance with each other.

Several sugar transporters were found to be down-regulated in the UGD repressed transgenic sugarcane lines, most notably the transcripts of a glucose-6-phosphate: phosphate-translocator which was found to be highly down-regulated in both transgenic lines (Table 2.2; Appendix Table A1). Down-regulation of sugar transporters as internodes mature was previously reported in sugarcane genotypes which accumulate high levels of sucrose (Watt *et al.*, 2005). The consistent down-regulation of a glucose-6-phosphate: phosphate translocator combined with the transcript increases known to be involved in starch degradation (Table 2.2) in the UGD repressed lines suggests that glucose-6-phosphate could be used as a substrate for sucrose or cell wall polymer biosynthesis instead of storage as starch (Kammerer *et al.*, 1998; Streb *et al.*, 2009, Geigenberger, 2011). No probe sets representing transcripts involved in glucose-6-phosphate metabolism were found to be significantly differentially expressed and thus more information is needed to support this assumption.

A large number of transcripts were found to be differentially expressed are generally thought to be involved in stress tolerance and defence (Table 2.3). One possible explanation is the osmotic stress that high sucrose concentrations cause (McCormic *et al.*, 2008). However, it has also been suggested that the signalling pathways for sucrose biosynthesis and abiotic stress are not completely independent of each other (Ehness *et al.*, 1997) which could also account for our findings. Reports have also associated the transcriptional regulation of several stress related transcripts to sucrose content in high sucrose accumulating sugarcane lines (Watt *et al.*, 2005; Casu *et al.*, 2007).

As previously suggested (Bekker, 2007), the myo-inositol oxidation pathway (MIOX) appears to supply the intermediates needed to produce the cell wall components in UGD repressed lines. Precursors for the MIOP would thus be needed in higher quantities to accommodate the decreased production of the SNOP. Transcripts for *myo*-inositol phosphate synthase (MIPS) were found to be significantly up-regulated in UGD 7, supporting this assumption. A slight up-regulation of MIPS in UGD 18 was also observed although the increases were not significant. A search for probe sets representing MIOX and UDP-Glc PPase could however not identify any with high sequence similarity. The transcripts for glucuronokinase were not significantly differentially expressed between any of the transgenic lines and the untransformed NCo310 line. Enzymatic acitivity assays to confirm that no alteration in glucuronokinase acitvity was present was not done. However, it must be noted that transcript level and protein activity do not always correlate with each other (Gibon *et al.*, 2004). The alteration observed in the MIOP transcripts is supported by the metabolite data which indicates *myo*-inositol levels to differ significantly between the untransformed NCo310 and UGD repressed transgenic sugarcane.

Although a number of studies have focused on microarray technology to identify candidate genes for manipulation in sugarcane, the findings have not yielded much success (Casu *et al.*, 2004; Watt *et al.*, 2005; Casu *et al.*, 2007). One of the primary concerns is the large number of probe sets for which no annotation is yet available. A complete genome sequence for sugarcane would undoubtedly enhance such studies.

2.6 Conclusion

This study set out to identify differentially regulated genes in the mature stem tissue of two transgenic sugarcane lines in which UGD was transcriptionally repressed. The goal was to identify what effect the repression of UGD had on the global transcript profile of the

transgenic sugarcane lines. The Affymetrix microarray platform was used to acquire the transcriptional profiles of these lines. The profiles of each of the two transgenic lines were individually compared to that of the untransformed NCo310 line and a list of differentially expressed genes generated.

Microarray results were verified with the use of semi-quantitative RT PCR of selected genes which were normalised to actin. The semi-quantitative RT PCR and microarray results correlated with each other for all genes investigated, with the exception of SPS A transcripts. The reason for the deviation of SPS A results between the two methods is most likely due to lower sensitivity of semi-quantitative RT PCR compared to microarrays.

In summary, a number of transcripts involved in cell wall metabolism and modification, signalling, stress and transport were found to be differentially expressed (Table 2.3). The general results of this study are supported by other reports which have focused on sugarcane with high sucrose concentration and found similar transcript expression patterns.

This study identified transcripts involved in trehalose, *myo*-inositol and raffinose metabolism to be of interest to future studies targeting increased sucrose and more efficient bagasse fermentability. Of particular interest from our findings were a group of genes involved in lignin biosynthesis and breakdown. Alterations of lignin content in plant cell walls are known to alter fermentation efficiency in biofuel production. Our findings suggest that enzymes involved in flavonoid and lignin biosynthesis compete for the common substrate *p*-coumaroyl-CoA (the branching point of the two pathways). Both HCT (lignin biosynthesis) and CHS (flavonoid biosynthesis) transcripts were found to be increased in the transgenic sugarcane lines but genes involved in downstream biosynthesis in both these pathways were down regulated.

Transcripts of cellulose, lignin, rhamnose and raffinose synthesis and those involved in the breakdown of the xylose component of hemicellulose were down-regulated. Transcripts related to the breakdown of starch and the synthesis of *myo*-inositol were up-regulated in the UGD repressed transgenic sugarcane plants. These findings are in accordance with those of previous report (Bekker, 2007). The findings from the previous and this study suggests enzymes involved in the MIOP to be potential candidates for future manipulation, although additional data needs to confirm this assumption.

57

In conclusion, this study has successfully analysed the transcriptome of transgenic sugarcane plants with repressed UGD. A number of genes involved in cell wall metabolism and some which may cause increases in sucrose have been identified as candidates for future manipulation in sugarcane.

2.7 Future work

During the harvest season of September 2011 a new batch of young internode samples were collected and analysed for sucrose content. After confirming a sucrose content increase in both transgenic lines (compared to untransformed NCo310 plants), the transcriptomes were analysed by microarrays. A new procedure was implemented which includes an increase in the amount of biological samples to diminish biological variance and to enable efficient pooling of samples. The hope is that the pooling of samples will improve the significance and accuracy of the microarray transcript level results due to a reduction in biological variance (Hanson *et al.*, 2008). This data has yet to be fully processed.

Acquiring transcript profiles of young internode tissue will assist in the identification of genes with altered transcript levels in more metabolically active tissue types (compared to mature internodes) as well as enable the comparison of sucrose synthesizing genes in tissue which differentially accumulate sucrose.

Chapter 3: Metabolome fingerprinting

3.1 Abstract

Plants accumulate a broad range of metabolites, during exposure to biotic and abiotic stress. Analyses of metabolite profiles provide key information to unravelling how plants overcome these unfavourable conditions.

The amounts to which a plant will accumulate a specific metabolite depends on its function, the plant species and the environmental conditions the plant is exposed to. Most metabolites are toxic or inhibitory to plant growth at high concentrations and thus are only accumulated to relatively low levels.

Several analytical techniques have been developed for the determination of metabolite profiles, each of which has advantages and limitations. Consequently, the use of multiple analytical techniques is common place.

In this study HPLC, NMR, GC-MS and biochemical assays were employed to investigate the polar metabolome of sugarcane plants in which UDP-glucose dehydrogenase was repressed.

3.2 Introduction

3.2.1 Plant metabolomics

Plants have been estimated to accumulate more than 200 000 different metabolites (Schauer and Fernie, 2006). A number of spectral methods have been developed to analyse these metabolomes and the availability of publicly available spectral libraries and databases have accelerated the use and further development of such methods (Tukinov *et al.*, 2005; Lewis *et al.*, 2007; Akiyama *et al.*, 2008; Cui *et al.*, 2008).

Metabolomics research has historically received much less attention than other omics fields (transcript- and prote-omics)(Sumner *et al.*, 2003). Investigating how molecular alterations affect the plant transcriptome or proteome might be preferred due to the complications of metabolomic studies. However, although proteins are the effectors of change, transcript and protein levels often do not correlate well with each other (Gibon *et al.*, 2004). Prior knowledge of both the transcriptome and proteome is needed to make further conclusions while this is not always true for metabolome investigations (Sumner *et al.*, 2003).

The driving force behind investigations of the plant metabolome is the agriculture, pharmaceutical and energy industries. Losses in crop yield due to stress (both biotic and abiotic) have promoted the investigation of new methods to increase crop resistance and yield (Schauer and Fernie, 2006). How a plant adapts itself to overcome and thrive in non-optimal conditions is key in producing elite agricultural crop lines.

Considering the above, the metabolome of individual plants should reflect their conditions and modifications to it, be it external or molecular in nature. Using this as the basis for further study, the plant metabolome can be used to generate an extensive and unique fingerprint which is an indication of the plant "environment". While fingerprinting is the simpler approach to metabolome analysis, it still allows for the grouping of samples according to their overall metabolic composition or fingerprint (Kruger *et al.*, 2008).

Metabolite fingerprinting has a number of uses which range from quality control in the industrial sector to correlating profiles with specific genotypes (Kruger *et al.*, 2008). The metabolite fingerprint can only be an accurate representation of the group of metabolites extracted as a number of factors come into play with metabolite extraction procedures (Kruger at al., 2008). The differences between metabolites make the extraction of all metabolites by a single method a non-realistic goal with the current technology (Lin *et al.*, 2007). A subset of the plant metabolome can be extracted according to the extraction method used and a fingerprint generated for comparisons between samples. Data from different extraction procedures can be combined to form a larger fingerprint with the ultimate goal of fingerprinting the entire metabolome.

While care must be taken during extraction procedures and the choice of statistical analysis, the correct harvesting of plant material is of the utmost importance. Metabolic changes in cells can occur at near instant speeds and inhibiting enzymatic action is key in order to obtain a fingerprint which represents pre-harvest tissue (Bieleski, 1964; Roessner *et al.*, 2000; van Bel, 2003; Lin *et al.*, 2007).

It is for this reason that the perchloric acid ($HClO_4$) extraction method has been employed in metabolite extractions for analysis by NMR (Kruger *et al.*, 2008). The procedure inhibits any enzymatic activity at the cost of degrading some acid labile polar metabolites and does not allow the extraction of hydrophobic compounds (Kruger *et al.*, 2008). A broader range of metabolites can in theory be extracted with organic solvents which allow for the extraction of both polar and nonpolar compounds (Kruger *et al.*, 2008). Studies have however shown

several enzymes still retain some activity in organic solvents which might alter metabolic compositions (Bieleski, 1964; Kruger *et al.*, 2008).

Indicating that the generated fingerprint is in fact a representative of the pre-harvest tissue is the subject of much debate (Kruger *et al.*, 2008) and no robust, validated method for sampling exists yet (Lin *et al.*, 2007). Recovery experiments where a known amount of a compound is added to the tissue before metabolite extraction can be used to address this. However, this is impractical when investigating a large number of compounds (Kruger *et al.*, 2008). The reproducibility of an extraction procedure is what most studies focus on.

3.2.2 Nuclear magnetic resonance

Nuclear magnetic resonance (NMR) is based on the principle that compounds have a magnetic field which is generated by the spin of the charged particles it contains. Hydrogen (^{1}H) and ^{13}C atoms have residual nuclear magnetism due to an odd number of nucleons, the magnitude of which can be measured by the application of an external magnetic field (Wilson and Walker, 2005). NMR procedures measure the chemical shifts of protons due to the induction of secondary fields by the applied magnetic field (Wilson and Walker, 2005). The area under the chemical shift peak is proportional to the number of protons a compound contains and thus, NMR methods are also quantitative.

One dimensional (1D) NMR in which only one atom such as hydrogen (^{1}H or proton) or ^{13}C is investigated requires only a few minutes to acquire the spectrum of a sample. The method is thus rapid and also non-degrading to the sample as no pre- or post-analysis modifications are required. Two dimensional (2D) NMR techniques combine two 1D spectra to produce a more detailed spectrum. However, a 2D NMR spectrum of the same resolution as a 1D spectrum of the same sample will require 16-20 hours to generate (Tang *et al.*, 2004; Lewis *et al.*, 2007).

The sensitivity of spectral methods are influenced by the signals from predominant metabolites which may in some instances mask the signals from less abundant metabolites (Tang *et al.*; 2004) which is especially true when working with crude extracts or metabolite mixtures (Lewis *et al.*, 2007).

Crude or complex samples which contain an entire subset of a tissue metabolome, such as those used in this study, cannot accurately be analysed by one dimensional (1D) NMR methods. Peak overlap of multiple metabolites complicates the identification of metabolites

and the estimation of quantities (Lewis *et al.*, 2007; Ludwig and Viant, 2009). The spectral overlap limitations of 1D NMR methods can be overcome by two dimensional (2D) NMR methods such as correlation spectroscopy (COSY), total correlation spectroscopy (TOCSY) and heteronuclear single-quantum correlation (HSQC) (Lewis *et al.*, 2007).

NMR has the advantage over other spectral techniques due to its non-destructive nature; no chemical derivatization of samples required (Wishart, 2007). Thus NMR is better suited to the analysis of sugars than either GC-MS or LC-MS. NMR is also non-biased and easily quantifiable. However, a disadvantage of NMR is the sensitivity of the technique. A large number of compounds fall in the concentration range below 10µM which is the lower detection limit range of NMR with the cut-off at between 1-5µM (Wishart, 2007; Kruger *et al.*, 2008). Metabolites which fall into the lower limit of detection cannot be distiguished from spectral noise.

3.2.3 Gas chromatography-mass spectrometry

In short, Gas Chromatography-Mass Spectroscopy (GC-MS) separates volatile analytes *via* a chromatography column/stationary phase. The time it takes for an analyte to "elute" is recorded and the mass of the analyte measure by the detector. The combination of the data can then be compared to a standard library to identify the compound (Wilson and Walker, 2005). The injected liquid sample is slowly heated to a pre-determined temperature, allowing highly volatile analytes to elute first. This also means that the sample under investigation needs to be thermally stable. The volatile analytes are carried through the stationary phase *via* an inert carrier gas such as nitrogen, argon or helium depending on the column used. Samples need to be derivatized in order keep highly polar compounds from being retained within the column (Wilson and Walker, 2005).

In this study nuclear magnetic resonance (NMR), gas chromatography–mass spectrometry (GC-MS), high-performance liquid chromatography (HPLC) and standard biochemical assays were used to investigate the metabolic changes occurring due the genetic modification of sugarcane. The results from this study were also used in the verification of microarray results which formed part of the transcriptional profiling of transgenic sugarcane lines.

3.3 Methods

3.3.1 Plant material

Plant material was obtained on several occasions to analyse soluble sugar levels. In all samplings, plants with similar physiology were selected. The transgenic plants used in this study were the UGD repressed sugarcane lines previously created at the IPB which were in their 5th growth season (Bekker, 2007). The control plants (untransformed NCo310) were of the same age.

In order to extract and determine the polar metabolome component of the untransformed NCo310 and transgenic sugarcane plants (UGD 7, 8, 10 and 18), mature and young internode tissue was sampled in September (2009) from plants growing under field trial conditions at SASRI, Mount Edgecombe, Kwazulu Natal, South Africa. This was the first sugarcane material sample set. The mature internode samples of transgenic lines UGD 7 and 18 were also used to extract RNA from for use in transcript analysis (Chapter 2).

The sugarcane plants were grown under monitored, external envirnomental conditions. Three biological replicates were obtained per line for each of the mature (internode 10) and young (internode 3-4) internode tissues. Sampling methods and internode determination are as discussed in the transcriptomics plant material section (Chapter 2).

The terrace plants were transferred as stem cuttings from the terrace and cultivated on the field for the 2010 growth season as part of a larger field trial. Field trial plants were grown in large "blocks" simulating field conditions. During the southern hemisphere winter (June to August) period of 2011 a second and third sugarcane material sample set was obtained from the field trial plants. The second sampling was done mid-April and the third sampling mid-July at SASRI. The mid-April sample set consisted of young and mature internode samples from all available UGD lines at SASRI. Six biological replicates were obtained per line per tissue. The mid-July material consisted of young and mature internode samples from the control (untransfomred NCo310) and transgenic lines UGD 7 and UGD 18. Three biological replicates were obtained per line.

A fourth and final sugarcane material sample set was obtained during the 2011 harvesting period (September, spring in the southern hemisphere) from the field trial sugarcane plants at SASRI and whole internode sections sent to the IPB. The sample set consisted of six young internode biological replicates per sugarcane line from the untransformed NCo310 , UGD 7,

UGD 8 and UGD 18. Only the young internode sections were available for sampling as the mature internodes were required by a different study group. The soluble sugar content of the plant material was investigated *via* enzymatic assays.

3.3.2 Metabolite extraction for NMR analysis

The polar subset of metabolites were extracted from ground sugarcane tissue using the perchloric acid extraction method (Kruger *et al.*, 2008). Ground stem material was kept frozen in liquid nitrogen and 3 g of tissue weighed off per sample in 50 mL polypropylene tubes. Ice-cold 3 M perchloric acid ($HClO_4$, 1 mL) solution was added to the samples and the mixture allowed to thaw on ice.

Samples were made up to 15 mL with 1M $HClO_4$ and vigorously mixed. Samples were then incubated on ice for 30 min and centrifuged (25 000 x g at 4°C, 15 min). The supernatant was transferred to a new tube and a volume of 10 mL 1 M $HClO_4$ added to the remaining pellet and centrifuged as described above. The supernatants were pooled after the centrifugation.

The supernatants were then neutralized with a 2 M KOH solution to a pH of between 5.0 and 6.0. A microprobe pH electrode (Hanna instruments, USA) was used as a standard probe proved difficult to use. The cell debris was kept in ethanol for analysis of cell wall uronic acids later. An incubation of the sample on ice for 30 min produced a potassium perchlorate precipitate ($KClO_4$). The pH was again adjusted to between 5.0 and 6.0 and the samples centrifuged (25 000 x g at 4°C, 15 min).

The supernatant was transferred to a new tube, frozen in liquid nitrogen and lyophilized in a freezedrier at -60°C (Virtis, Benchtop K).

Dried samples were re-dissolved in 1 mL of a 25 mM NaH_2PO_4/Na_2HPO_4 phosphate buffer (pH 7.5). The pH was then checked and adjusted to 7.5.

Any precipitated $KClO_4$ was removed by centrifugation (14 000 x g at 4°C, 10 min). Samples were freeze-dried and the dry samples stored at -80°C until needed.

3.3.3 Metabolite extraction for GC-MS analysis

Metabolites were extracted *via* a methanol based extraction method with heating which has been reported to be the most reproducible compared to other extraction procedures (Roessner *et al.*, 2000). Polar metabolites were extracted from ground stem material stored at -80°C. A

100 mg of frozen material was added to 1400 µL of 100% methanol containing 60 µL of ribitol (0.2 mg.mL^{-1} stock in dH2O) as internal standard. The samples were incubated at 70°C for 15 min in a shaking heating block (BIOER, MB-102).

Samples were centrifuged (13 000 x g, 10 min) after which 700 µL of the supernatant was transferred to a new tube. A volume of 300 µL chloroform and 700 µL water was added to the supernatant and the samples vortexed and incubated at room temperature for 10 min to allow for phase separation.

Following the incubation step, samples were centrifuged (13 000 x g, 10 min). 100 µL of supernatant was then transferred to a new tube and dried in a centrifugal vacuum drier (Genevac, EZ-2) without heating. Dry samples were stored over silica gel at -80°C after tubes were flushed with argon.

On the day of analysis, samples were dried for 30 min in a centrifugal vacuum drier prior to derivatization. This was followed by the addition of 80 µL of methoxyamine reagent (20 µg methoxyamine hydrochloride/mL pyridine). Samples were shaken at 37°C for 2 hours. A 140 µL of N-Methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) and 20 µL time standard mixture (n-alkanes, C12-C36) was added to the sample which was shaked at 37°C for an additional 30 min.

A volume of 1 µL derivatized sample was injected using splitless injection. The GC-MS system consisted of the following instruments: an AS 2000 autosampler, trace GC and a quadrupole trace mass spectrometer (MS) (ThermoFinnigan). The chromatography was performed on a 30m Rtx®-5Sil MS column with Integra Guard (inner diameter 0.25 nm, film thickness 0.25 mm, RESTEK). An injection and ion source temperature of 230°C and 200°C respectively was used. Following injection, the GC oven was kept at 70°C for 5 min, after which the oven temperature was ramped by 1°C.min^{-1} to 76°C. This was followed by a second ramp of 6°C.min^{-1} to a temperature of 350°C. The system flow rate was 1 mL.min^{-1}. The system temperature was equilibrated to 70°C prior to injection of the next sample. Mass spectra were recorded at two scans per second with a range of 25-600 m/z (mass-to-charge).

Chromatograms and spectra were evaluated using Xcalibur™ software version 1.2 (Finnigan Corporation 2000) and converted to .netCDF format for further data analysis.

3.3.4 Analysis of NMR data

Dried metabolites were dissolved in 1mL of deuterium monoxide (D_2O) containing TMS at a concentration of 4.5 mM. The weight of the dried metabolites dissolved in the D_2O ranged from 60 mg to 100 mg with the majority of samples containing 100 mg of dried metabolites.

To remove any residual $KClO_4$ still present, samples were centrifuged (13 000 x g at 4^oC, 5min). A volume of 750 µL supernatant was then transferred to a thin wall NMR tube for analysis.

A 600 MHz Varian INOVA NMR at the Central Analytical Facility of Stellenbosch University (CAF-US) was employed to acquire the 2D HSQC spectra of the crude metabolite extracts. Proton (^1H) resonances were acquired at 600 MHz and Carbon thirteen (^{13}C) at 150.88 MHz at a temperature of 25^oC. The number of transients were 16 per spectra.

The FID (Free Induction Decay) files were Fourier transformed with the Vnmr (Varian NMR Spectrometer Systems) macro and converted to UCSF format for analysis in the statistical language R using the package rNMR (Lewis *et al.*, 2009) with default parameters. No phase correction was applied. Spectra were shift corrected to the spectra of sucrose obtained from the Biological Magnetic Resonance Data Bank (BMRB).

One dimensional proton spectra were analysed with Advanced Chemistry Development (ACD) NMR Processor, academic Edition, version 12.01.

3.3.5 Analysis of GC-MS data

Raw GC-MS data was batch deconvoluted using the Automated Mass-Spectral Deconvolution and Identification System version 2.69 (AMDIS), (National Institue for Standards and Technology). AMDIS was configured to automatically identify metabolites using the freely available plant metabolite library from the Golm Metabolome Database (Kopka *et al.*, 2005). Identification was based on retention indexing using the retention times of the added n-alkanes and the similarity of the mass-spectra to those in the Max Planck Institute of Molecular Plant Physiology (MPIMP) database. An ion-retention time (IRt) list based on the AMDIS results was edited in Microsoft Excel 2010 and saved as a text file. MET-IDEA version 2.05 (METabolomics Ion-based Data Extraction Algorithm) was used for ion based quantification of metabolites (Broeckling, 2006). MET-IDEA utilizes directed extraction of ion intensity values based on a list of ion–retention time values. The ions,

representative of specific metabolites, are located based on the corresponding retention time values and corrected for retention time drift, the centre of the peak is isolated based on the ion intensity and peak start and stop points are identified based on shape characteristics. The ion counts within the width of the peak are summed to provide a 'peak area' value which represents the metabolite abundance. Following identification and quantification, all values were normalized to the internal standard ribitol (319 m/z) and the fresh weight of each sample.

3.3.6 Soluble Sugar extraction

Sugars were extracted from 100mg ground material in 70% ethanol (EtOH) and 30% 100 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (pH7.8), 20 mM MgCl2 (30% v/v HM-buffer). The extractions were done in triplicate for each biological sample. Ground material was incubated in 1:10 volumes (fresh weight: extraction buffer) extraction buffer at 70°C overnight. Debris was spun down (13 000 x g, 15 min) and the supernatant filtered if needed. An additional extraction sugar was done on the remaining debris in 500 µL of extraction buffer which was incubated at 70°C for 2 hours to ensure all sugars were removed. The second extraction was pooled with the first and stored at -20°C.

3.3.7 HPLC analysis

A volume of 600 µL soluble sugar extract was dried at 60°C in a rotary vacuum evaporator for 60min to remove the ethanol component of the extraction buffer. The removal of ethanol from the sample eliminated the need for an equilibration period between sample injections. The dried samples were then resuspended in 600 µL of de-ionized water for HPLC analysis. 500 µL of sample was transferred to a glass HPLC vial (Waters Corp.) and vortexed prior to analysis.

Sugar content was analysed by a HPLC connected to a pulsed amperometric detector (HPLC-PAD). A volume of 5 µL was analysed per sample in triplicate. To prevent column overload, several test samples were analysed for sugar content in order to determine if sample dilution was required. A sample run time of 10 min was sufficient for the acquisition of glucose, fructose and sucrose peaks. Peaks for glucose, fructose and sucrose were identified at time points 3.450 min, 3.746 min and 6.285 min respectively from a multi-sugar standard sample. The peak area of each peak of the standard was equal to a concentration of 200 ng/µL of the particular sugar. A detector efficiency control was also included for sugars.

Spectra were autoscaled by the HPLC operational software and normalized to the amount of fresh weight of each sample. The HPLC instrument was operated by Ewald Albertse at SASRI according to protocols well established within the institute.

3.3.8 Alcohol insoluble residue preparation and uronic acids content determination

A 100 mg of tissue was added to 100% EtOH to a concentration of 80% (v/v). Samples were incubated at 70°C for 20 min to extract metabolites. Samples were centrifuged (4000 x g, 10 min) and the supernatant discarded. The extraction was repeated four times. The alcohol insoluble residue (AIR) was washed with acetone and dried in a centrifugal vacuum drier (Genevac, EZ-2) until dry. Samples were immediately used for uronic acid determination.

The uronic acid determination procedure was adapted from the methods of Blumenkrantz and Asboe-Hansen (1973) and van den Hoogen $et\ al.$, (1998). Briefly, 40 μL of hydrolyzed AIR sample was added to microtiter plate (Greiner, F-Bottom, non-UV) well. To each well, 200 μL of 96% H_2SO_4 (Sigma) was added which contained 120 mM sodium tetraborate (Fluka, $Na_2B_4O_7$). After a 15 minute incubation period at RT, the background was measured at 530 nm (BMG LabTech, FLUOstar OPTIMA, Optima software v. 2.10). A volume of 40 μL m-hydroxydiphenyl (Sigma) reagent (100 μL m-hydroxydiphenyl in DMSO (SAARChem) at a concentration of 100 mg.mL^{-1} (in 80% H_2SO_4 (v/v)) was freshly made just before use and added to each well. Samples were thoroughly mixed and incubated at RT for 15 min. After the incubation step the sample absorbance was measured at 530 nm. A galacturonic acid (Fluka) standard curve was established from 0 to 8 μg.

3.3.9 Statistical analysis of GC-MS and NMR spectral data

The data output from AMDIS and MET-IDEA was imported into the web-based metabolomics analysis program MetaboAnalyst version 2.0 (Xia $et\ al.$, 2009; Xia and Wishart, 2011). Data was mean-centered and divided by the standard deviation of each variable (Autoscaled) before multivariate analysis (MVA). MVA consisted of principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) of the transgenic and wild-type metabolomes (Wold, 1975; Wold $et\ al.$, 1983). Random Forest was used to identify metabolites which significantly influence the MVA models (Breiman, 2001).

A similar approach was used to analyse the NMR spectra. Spectral regions of interest were defined according to the default rNMR settings together with a strong filter applied. The absolute area of a ROI was calculated and normalized to the amount of dried metabolite per

sample. The ROIs of all samples were combined into one tab delimited file and used as the input for MetaboAnalyst version 2.0.

The ROIs were peak picked in order to determine the metabolites they represented. The peaklist was used as input in MMCD. A proton tolerance threshold of 0.05 ppm and a carbon tolerance threshold of 0.25 ppm were applied as filter. The peaklists were also used as input into SpinAssign (Platform for RIKEN Metabolomics, PRIMe) (Chikayama *et al.*, 2010, Chikayama *et al.*, 2008) for compound identification and analysis.

3.3.10 Sucrose assays of the sugarcane sample sets

Soluble sugar content was determined enzymatically to assess the current sucrose yield of the sugarcane lines.

The assay was slightly modified from the method described by Bergmeyer and Bernt (1974). Briefly, 50 µL of 10 times diluted sugar extraction was added to 200 µL of assay buffer (100 mM HEPES pH 8.0, 5 mM $MgCl_2$, 1 mM ATP, 1 mM NAD) in a 96 well microtiter plate (Greiner, flat bottom well). 0.5 Units of Hexokinase and Glucose 6-phosphate dehydrogenase was added and the absorbance measured at 340 nm until measurements stablized. Invertase was added to each well and the increase in absorbance due to NADH formation measured. The amount of sucrose was calculated from the change in absorbance values.

Fructose and glucose concentrations were determined as described by Bergmeyer and Bernt (1974) from undiluted sugar extractions.

3.4 Results

3.4.1 HPLC analysis of sugarcane lines

Samples collected from the terrace or field were first analysed for soluble sugar content to determine if a sucrose increase was present in the transgenic lines when compared to the untransformed NCo310 line. The sugarcane lines sampled in 2009 were analysed for sucrose (Figure 3.1), fructose and glucose (Figure 3.2) content *via* HPLC analysis. The lines included a control (untranformed NCo310) and several transgenic sugarcane lines in which UGD was repressed.
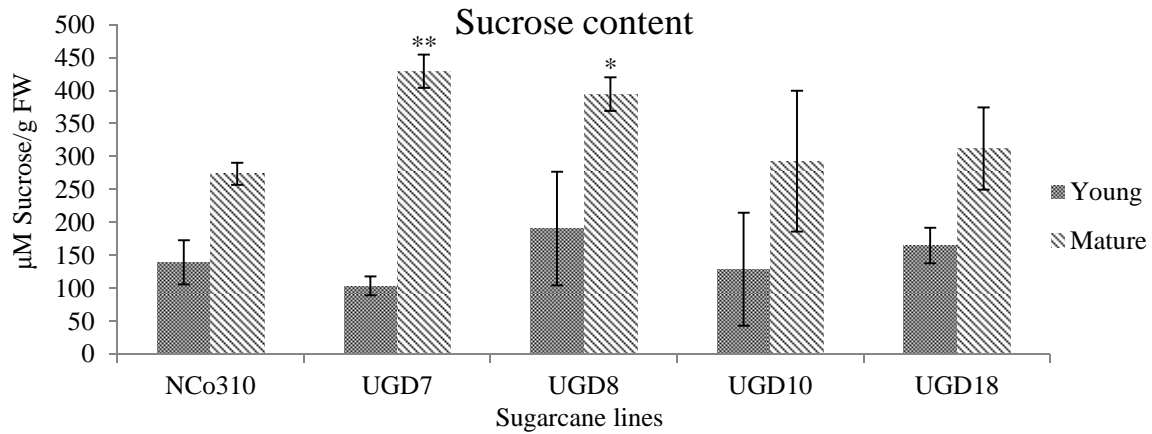
Figure 3.1: Sucrose content of the 2009 samples. Transgenic lines which could not be obtained later in the study are not included. Values displayed as the mean ± SEM. Significant diffirences determined by the Student's t-test. All samples, n=3. * $p < 0.05$; ** $p < 0.01$
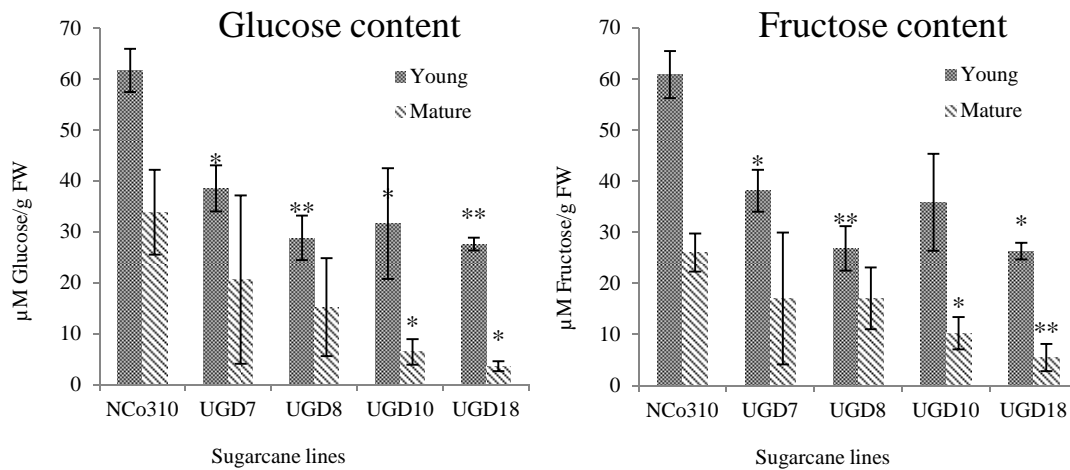
Figure 3.2: Glucose and fructose content of the 2009 samples. Transgenic lines which could not be obtained later in the study are not included. Values displayed as the mean ± SEM. Significant diffirences determined by the Student's t-test. All samples, n=3. * $p < 0.05$; ** $p < 0.01$

The HPLC analysis revealed none of the young internode samples of the 2009 sample set to have significantly increase sucrose content when compared to the control (Figure 3.1). The mature internodes from transgenic lines UGD 7, 8, 10 and 18 of the same sample set had increased sucrose content. Lines UGD 7, UGD 8 and UGD 18 were selected for further analysis as these lines showed the largest increase in sucrose content when compared to the other transgenic lines (Figure 3.1).

A significant increase of 56.9% ± 5.9% (155 ± 9 µM.g^{-1} FW) in sucrose was present in the mature tissue of line UGD 7 and an increase of 44.2% ± 6.3% (120 ± 8 µM.g^{-1} FW) for UGD 8. Transgenic line UGD 18 had a sucrose increase of 14.1% ± 20.0% (38 ± 8 µM.g^{-1} FW) for

70

the mature tissue. Glucose and fructose content was significantly lower in several of the transgenic lines when compared to the untransformed NCo310 plants for both the mature and young internodes (Figure 3.1)

During April of 2011, the soluble sugar content of the sugarcane plants growing in the field was determined by HPLC analysis (data not shown). The mature internode samples of the transgenic lines did not present a significant increase in sucrose content when compared to the untransformed NCo310 line for any of the lines. In order to confirm these results, a second sampling was done during July of 2011.

Figure 3.3: Sucrose content of the July 2011 sugarcane young and mature internode samples as determined by HPLC analysis. Values displayed as the mean ± SEM. Significant diffirences determined by the Student's t-test. All samples, n=3. * $p < 0.1$; ** $p < 0.05$
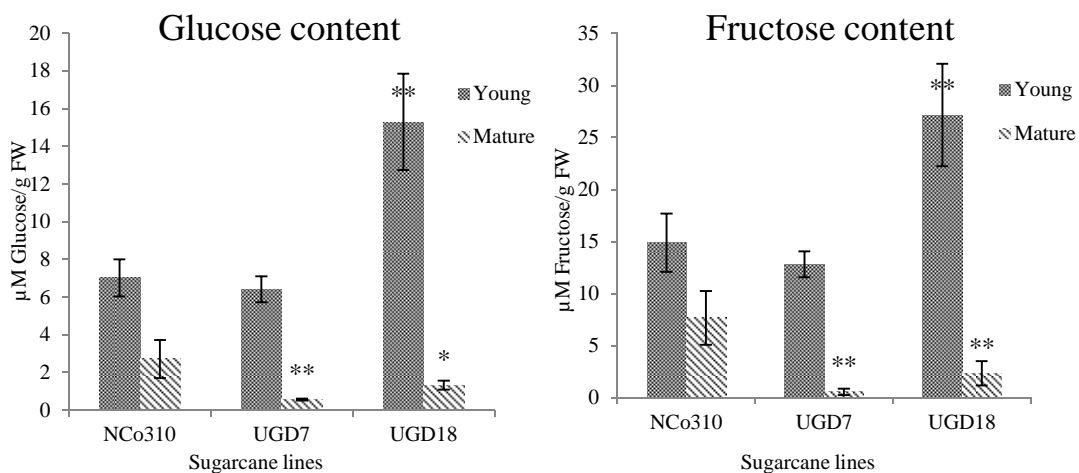
Figure 3.4: Glucose and fructose content of the July 2011 sugarcane young and mature internode samples as determined by HPLC analysis. Values displayed as the mean ± SEM. Significant diffirences determined by the Student's t-test. All samples, n=3. * $p < 0.05$; ** $p < 0.01$

71

The July 2011 dataset indicated a significant increase of 15.1% $\pm$ 8.2% (55 $\pm$ 5 μM.g^{-1} FW) in sucrose content in the mature internode tissue of line UGD 7 (Figure 3.3). The young internode tissue of UGD 18 had an increase of 26.9% $\pm$ 13.1% (21 $\pm$ 3 μM.g^{-1} FW) while the mature internode tissue had a 27.9% $\pm$ 8.3% (101 $\pm$ 8 μM.g^{-1} FW) increase in sucrose when compared to untransformed NCo310 (Figure 3.3). Line UGD 7 showed a decrease in glucose and fructose content from the untransformed NCo310 line (Figure 3.4) which was significant only in the mature internode tissue. Line UGD 18 showed a significant increase in the glucose and fructose content of the young internodes but a significant decrease in the mature internode tissue for both glucose and fructose (Figure 3.4).

Samples from the young internodes had far less sucrose than the mature internodes (Figure 3.1; Figure 3.3) which is to be expected as immature tissues do not actively accumulate sucrose (Whittaker and Botha, 1997; Zhu *et al.*, 1997).

3.4.2 Enzymatic determination of soluble sugar content

Enzymatic assays were employed to determine the soluble sugar content of the April 2011 sample set which previously showed no sucrose increase as determied by HPLC analysis. Soluble sugars were again extracted from the tissues and the results of the enzymatic assay was in agreement with the HPLC results; no significant sucrose increase was present in the transgenic plants compared to the untransformed NCo310 line (data not shown). The sugar extractions and enzymatic soluble sugar determinations were repeated three times to eliminate experimental error. The samples were not used in any further analysis.

The September 2011 young internode sample set soluble sugar content was determined by enzymatic assay. The assay revealed two of the three UGD repressed sugarcane lines had increased sucrose content compared to the untransformed NCo310 line.
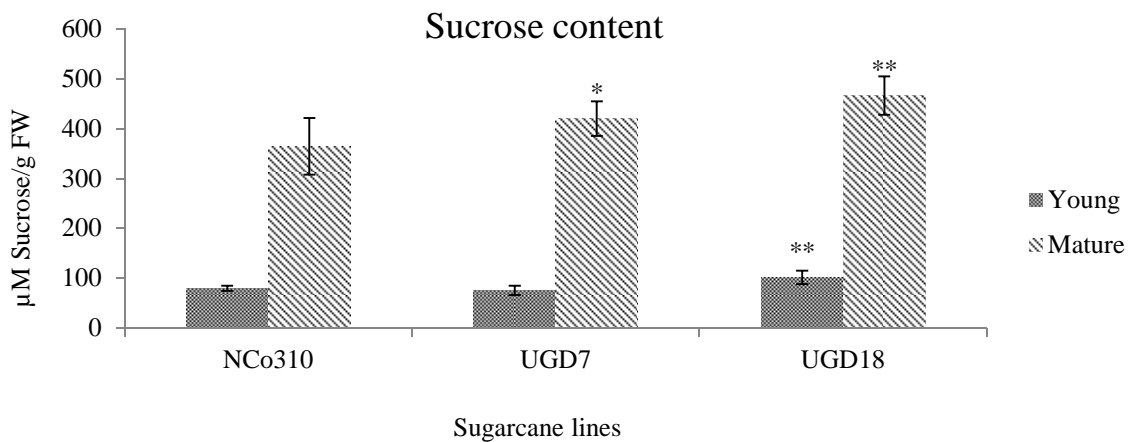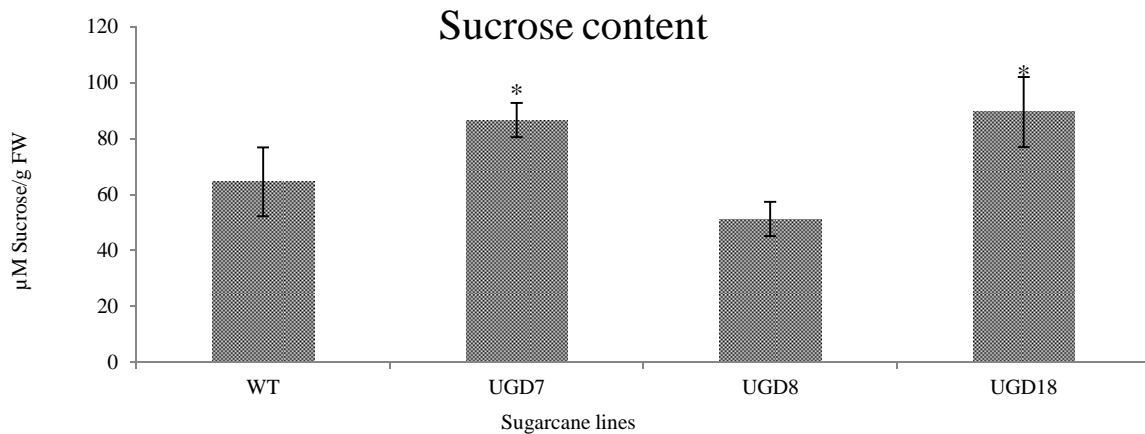
Figure 3.5: Sucrose content of the September 2011 sugarcane young internode samples as determined by enzymatic analysis. Values displayed as the mean ± SEM. Significant diffrences determined by the Student's t-test. All samples, n=5. * p < 0.05

The UGD 7 and UGD 18 transgenic lines presented a significant increase in sucrose of 34.2% ± 9.5% (22 ± 2 µM.g^{-1} FW) and 38.5% ± 19.4% (35 ± 7 µM.g^{-1} FW) respectively when compared to the untransformed NCo310 line (Figure 3.5). The UGD 8 line did not present a significant change in sucrose content compared to the untransformed NCo310 line. This result confirmed that several of the UGD repressed sugarcane lines still have increased sucrose content when compared to the untransformed NCo310 sugarcane line.

3.4.3 GC-MS analysis of mature sugarcane internodes

The mature internode tissue samples from the 2009 sample set was used as material for metabolite extraction in order to determine the polar metabolome component of the sugarcane lines *via* GC-MS. The untransformed NCo310 and transgenic lines UGD 7 and UGD 18 were analysed as these lines were included in the transcriptome analysis and both transgenic lines had increased sucrose content (Figure 3.1).

Partial least squares (PLS) regression and principal component analysis (PCA) was used to construct a multivariate model of the mature internode tissue polar metabolome in order to compare the untransformed NCo310 and UGD repressed transgenic lines according to the differences in metabolite composition (Figure 3.6). Not all variables in a multivariate dataset are important to the analysis. The amount of variables can be reduced to a more "manageable" amount by selecting only those variables which have the greatest influence on the dataset. PCA and PLS both reduce the number of variables in order to identify these variables for further analysis.

In order to determine if the untransformed NCo310 and transgenic sugarcane plants differ with regards to their polar metabolomes, scores and loadings plots were constructed. A significant difference in the polar metabolome of the untransformed NCo310 and UGD repressed transgenic sugarcane lines were observed by the different groupings of the three lines (Figure 3.6 A). In order to determine if the class discriminations are significant a permutation test was done. The permutation test results indicate the difference between untransformed NCo310 and transgenic groups to be significant as the "real class" did not fall within the random class assignments (data not shown) (Bijlsma *et al.*, 2006).

The untransformed NCo310 samples have very similar metabolite profiles and form a compact cluster away from the more spread out transgenic samples. The spread out cluster of the transgenic samples is expected as the two lines are unique and should have different metabolite profiles. Of note is sample UGD7-2 (U7-2 A, B and C) which does not cluster with the other UGD 7 biological replicates. Biological replicates should have very similar metabolite profiles as observed for the untransformed NCo310 and UGD 18 samples but is not the case for the second biological replicate of the UGD 7 transgenic line.

The loadings plot (Figure 3.6 B) indicates which metabolites have the greatest influence on the classification of a sample as either untransformed NCo310 or transgenic. The loadings plot however displays all variables instead of only the most significant varaibles. In order to determine which of the metabolites (variables) had the greatest effect on the classification as transgenic or untransformed NCo310, a variable importance in projection (VIP) and random forests analysis was done.
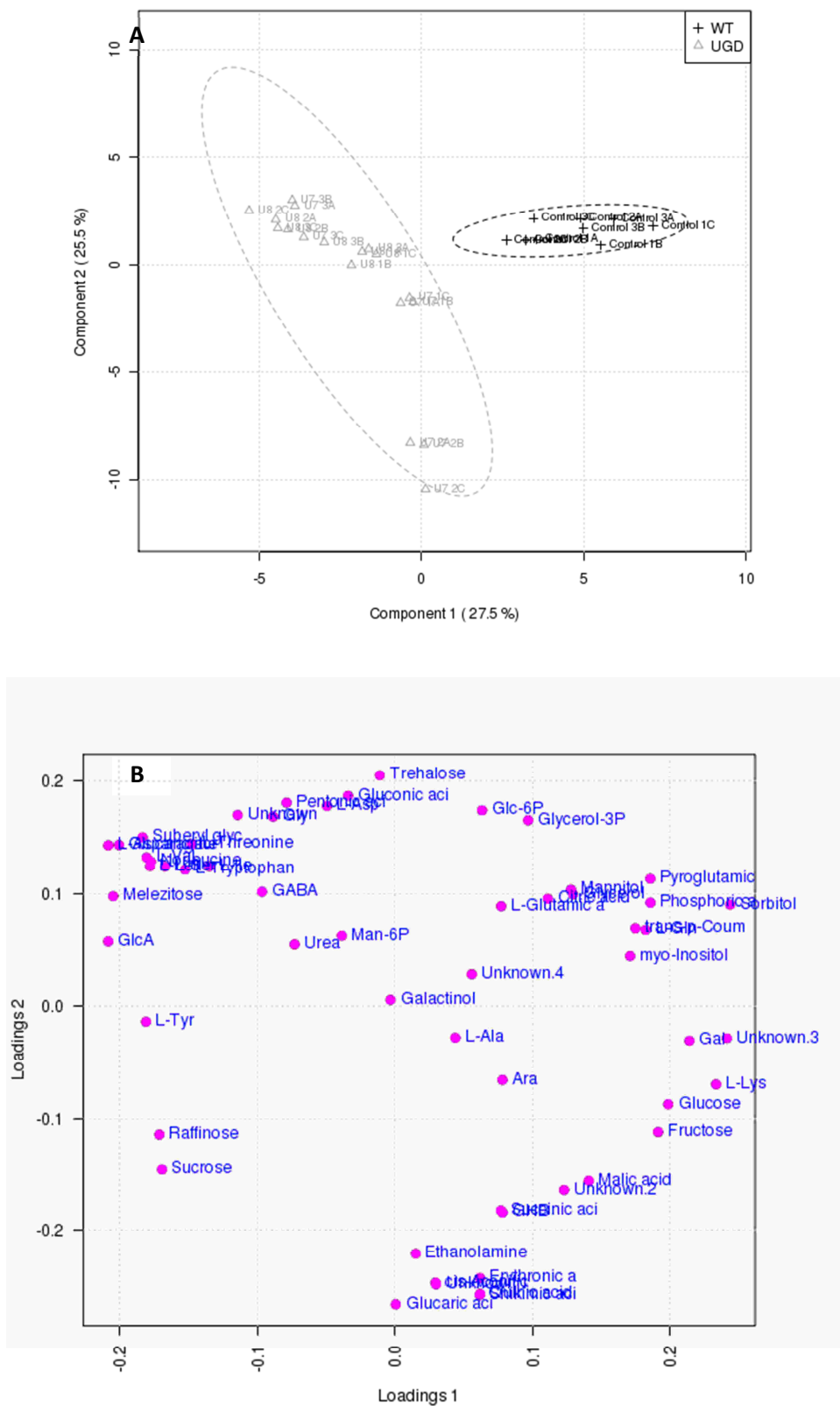
Figure 3.6: Partial least squares regression model of the transgenic and wild type sugarcane polar metabolome. Scores (**A**) and loadings (**B**) plots were constructed with the plsr function from the R package pls (Wehrens and Mevik, 2007). Elipse indicate 95% confidence threshold.
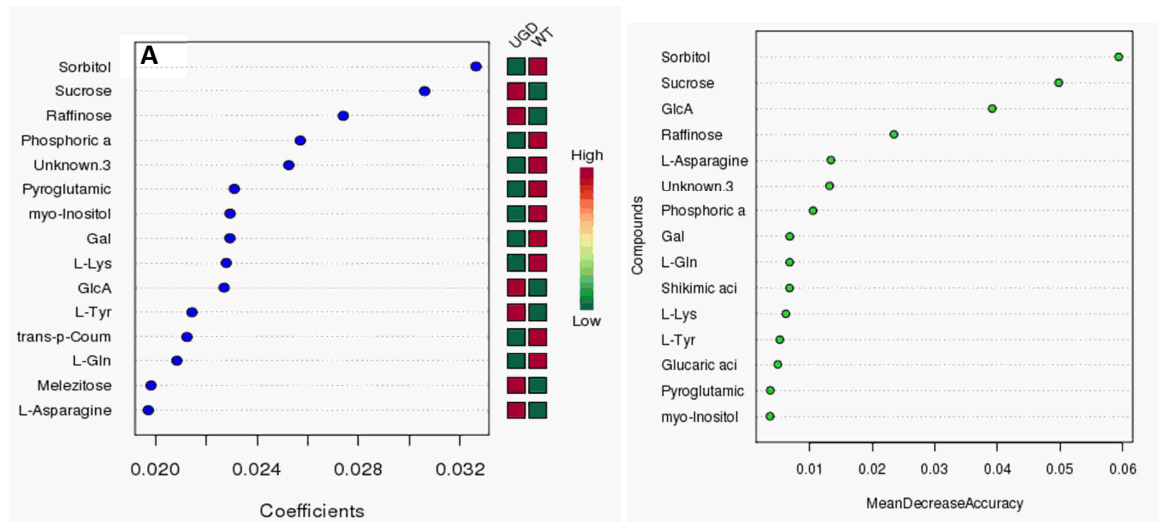
Figure 3.7: Coefficient-based variable importance in projection (**A**) and random forests analysis (**B**) plots. Phosphoric a, phosphoric acid; Pyroglutamic, pyroglutamic acid; Gal, galactose; L-Lys, L-lysine; GlcA, Glucuronic acid; L-Tyr, L-tyrosine; trans-p-Coum, trans-p-coumaric acid; L-Gln, L-glutamine.

The VIP and random forests analysis identified the 15 most important metabolites and the relative concentration of each in the respective class (Figure 3.7 A; Figure 3.7 B). Random forests analysis works well with data which have a low sample number but a large number of variables. Both of the analyses identified sorbitol, sucrose, raffinose and myo-inositol amongst others as important metabolites for the classification of a sample. A volcano plot was constructed to produce a more informative visualization of the significant metabolites (Figure 3.8).

The volcano plot indicated the majority of metabolites to have a fold change within a threshold of two ($2 > x < -2$) and only a few metabolites to fall outside of this range when the untransformed NCo310 and transgenic lines were compared to each other. The addition of a p-value threshold of 0.05 reduced the number of important metabolites to 15. Both sorbitol and glucuronic acid have large fold change values, while sorbitol has the higher significance level. Opposite to this is raffinose which has a high significance value but a small FC magnitude.
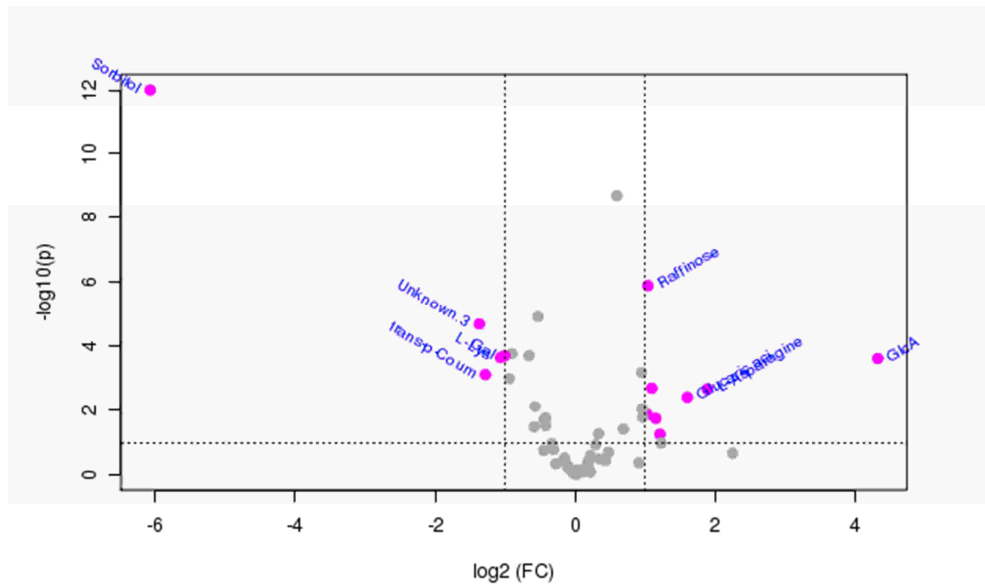
Figure 3.8: Volcano plot of polar metabolites of transgenic and wild type sugarcane showing significantly altered metabolite levels.

3.4.4 NMR analysis of mature sugarcane internodes

Heat-labile metabolites cannot be investigated *via* GC-MS analysis. For this reason NMR was employed to quantify these metabolites. NMR spectra of the mature internode sample set of 2009 were overlaid and regions of interest (ROIs) designated according to the default parameters of the R package rNMR with a strong filter applied. ROIs are assigned to areas with a large number of peaks or where variation between samples accur. The ROIs show that considerable variation exists among biological replicate samples as visible in the peak contour lines (Figure 3.9). Samples of the second untransformed NCo310 (C2_10) and UGD 7 (U7_2_10) biological replicates have low metabolite concentrations due to several low level or missing peaks at the corresponding ROIs. Contour profile deviations between the biological replicates of UGD 18 are also observed (Figure 3.9).

77
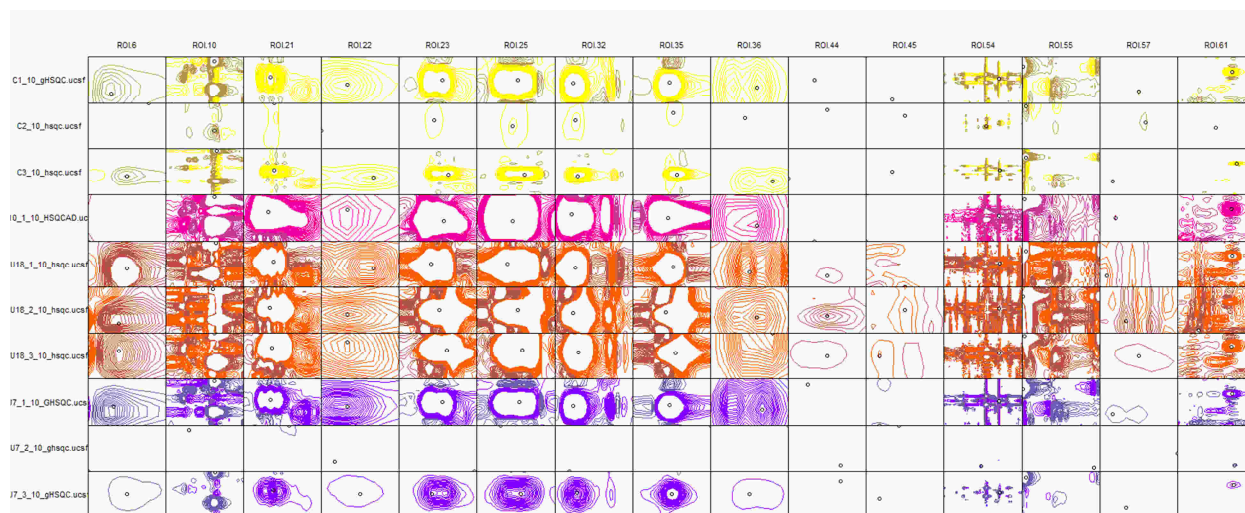
Figure 3.9: Region of interest (ROI) plots for a several ROIs of all of the mature internode sample spectra. ROIs couloured according to line; untransformed NCo310 in yellow, UGD 10 in pink, UGD 18 in orange and UGD 7 in purple. All sugarcane lines are represented by three biological replicates except UGD 10 which is only represented by one biological sample. The ROIs display the peak contours in the specified area.

The deviations observed between replicates are expected as a result of minor pH differences between samples. NMR analysis is extremely sensitive to variations in pH between samples which influences the spectral output.

Regions of interest values were normalized to the constant sum and imported into Microsoft Excel 2010. The data was arranged according to untransformed NCo310 or transgenic groups and saved in .csv format for import into MetaboAnalyst. Data was pre-processed and autoscaled before further analysis.

In order to determine if there was a difference between the polar metabolomes of sugarcane lines a PLS-DA was performed on the NMR ROI data. The samples were grouped into untransformed NCo310 and UGD mature internode groups for the analysis (Figure 3.10).

The analysis of the untransformed NCo310 and UGD groups revealed that there was overlap between the two sugarcane groupings and thus no significant differences between the NMR aquired polar metabolomes (Figure 3.10). Of note are the 2nd biological replicates of untransformed NCo310 (C2_10) and UGD 7 (U7_2_10) which do not cluster with their respective biological replicates, the latter of which is also observed in the GC-MS analysis (Figure 3.10). The loadings plot (Figure not shown) displays the ROIs which have the largest influence on the classification of a sample as untransformed NCo310 or UGD. The young

internode samples from the same sample set were also analysed *via* NMR analysis but due to a lack of replicates, down-stream analysis was not possible (data not shown).
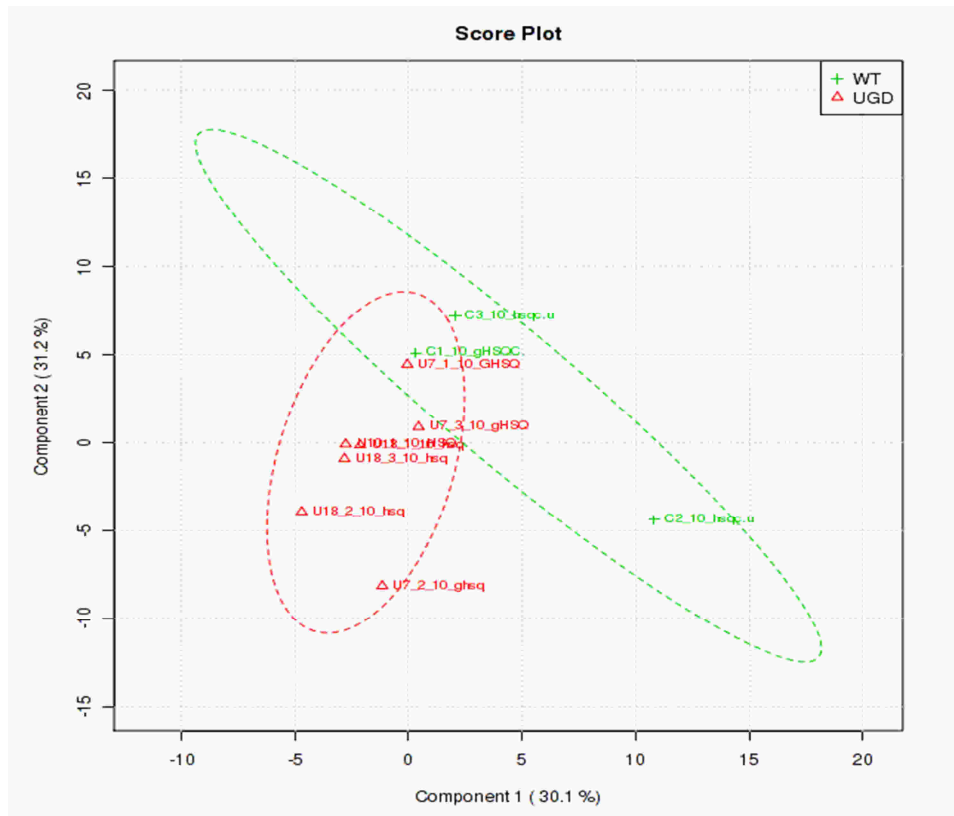
Figure 3.10: Partial least squares regression model of transgenic and wild type sugarcane. Scores plot was constructed with the plsr function from the R package pls (Wehrens and Mevik, 2007).

The results from the PLS-DA was confirmed by a permutation test on the PLS-DA model which revealed that the "real class" was assigned as the random class and thus the two class discriminations (groups) are not significantly different from each other. A VIP analysis was done despite the results which indicate the NMR generated polar metabolite profiles from the untransformed NCo310 and UGD groups not to differ from each other. The VIP plot identified which ROIs to be the most important in distinguishing if a sample is an untransformed NCo310 or UGD sample (Figure 3.11).

79
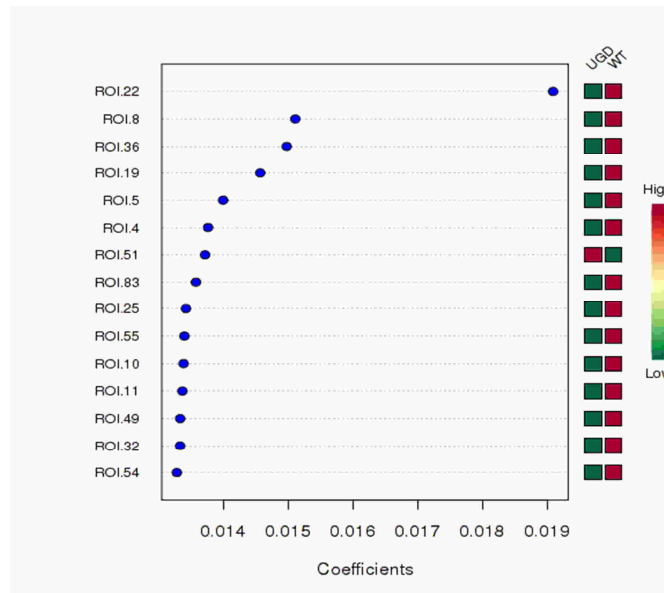
Figure 3.11: Coefficient-based variable importance in projection analysis of the NMR ROI data.

The volcano plot indicated that no ROIs were below the p-value threshold of 0.05 and had a fold change magnitude of greater than 2 or less than -2 (Figure 3.12).
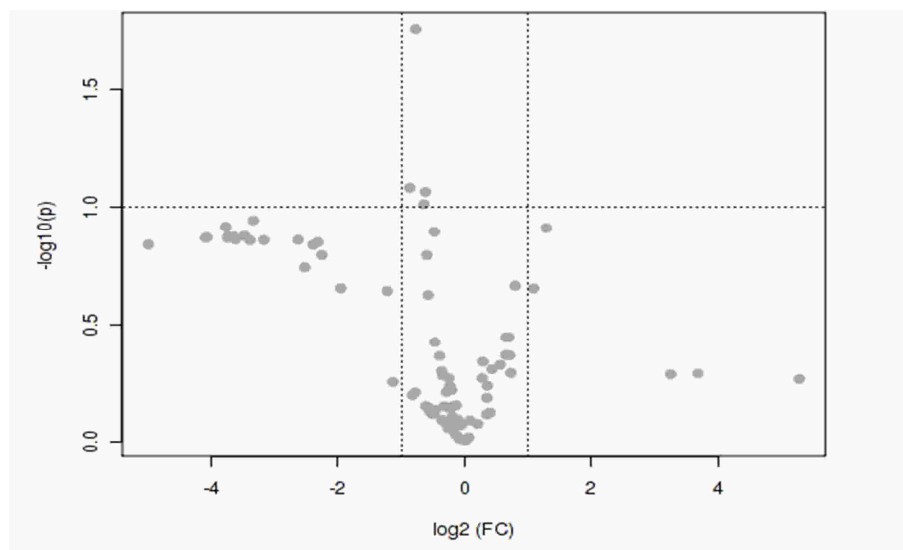
Figure 3.12: Volcano plot of the transgenic and wild type sugarcane ROIs as identified by NMR analysis. Dotted lines indicate the p-value and fold change thresholds.

Due to the low sample number it was not possible to do random forests analysis on the dataset.

In order to determine which metabolites were represented by the 15 ROIs identified by the VIP analysis, the ROIs were peak picked in rNMR and the data imported into both RIKEN

PRIMe and MMC databases. This analysis was done in order to determine if there were any correlations between metabolites identified by the NMR and GC-MS analysis (Table 3.1).

Table 3.1: A subset list of the most significant metabolites identified from the peaklists from the ROIs.

Metabolites	
Cellobiose	Raffinose
Choline	Ribitol
D-Arabitol	L-Arabinose
D-Fructose	Maltose
D-Xylulose	Mannitol
D-Glucose	Melezitose
D-Xylose	Melibiose
D-Glucuronate	*Myo*-inositol
D-Mannose	UDP-glucose
D-Ribose	Sucrose
D-Sorbitol	Trehalose
D-Tagatose	

The NMR peak assignment identified several metabolites which were also identified by the GC-MS analysis as important. Of note are the metabolites *myo*-inositol, sucrose, raffinose, xylose and sorbitol (Table 3.1) which are highly ranked in the GC-MS analysis.

3.4.5 Uronic acid determination

The uronic acid content of the UGD repressed transgenic lines is known to be increased compared to the untransformed NCo310 plants (Bekker, 2007). This increase in uronic acid content is attributed to the increase in MIOX activity (Bekker, 2007). In order to determine if the increase in uronic acid content was still present in the 2009 sample set, uronic acid content was determined *via* enzymatic assay. A significant increase in uronic acid content was present in the young internode tissue of transgenic line UGD 7. Both of the UGD transgenic lines showed a significant increase in the mature tissue when compared to the untransformed NCo310 uronic acid levels. Thus the increase observed in the previous generations of the UGD transgenic sugarcane lines are still present.
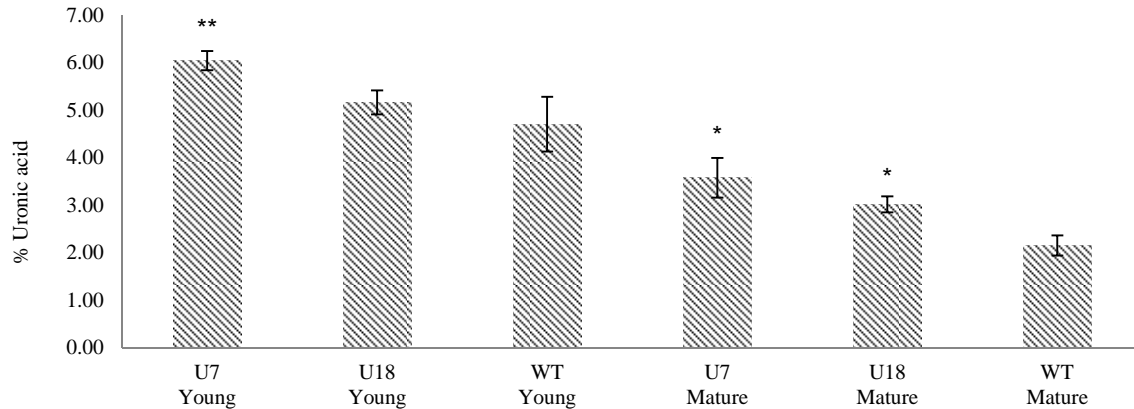
Figure 3.12: Percentage uronic acids in the cell walls of young and mature tissue from the untransformed NCo310 and UGD transgenic sugarcane lines. Uronic acid content was calculated as percentage of alcohol insoluble residue (AIR). Values displayed as mean ± SEM. Significant diffirences determined by the Student's t-test. n=3 for all samples. * P < 0.05; ** P < 0.01

3.5 Discussion

The UGD repressed sugarcane lines investigated previously (Bekker, 2007) showed a significant increase in sucrose content when compared to control (untransformed NCo310) sugarcane lines. The repression of UGD did not only alter the sucrose content of the transgenic lines but also cell wall component levels (Bekker, 2007). These UGD repressed transgenic sugarcane lines were not grown under the same conditions as the plants investigated in this study and thus the need to confirm the previously obtained results. The previous report were on greenhouse grown plants and since then the plants were propagated and moved to field trials under uncontrolled but monitored conditions.

The polar metabolomes of the transgenic and untransformed NCo310 lines were investigated *via* NMR and GC linked MS in order to determine the effect of UGD repression on the metabolome and to verify to what degree the metabolite changes were still present.

Due to the benefits of NMR (Wilson and Walker, 2005; Kruger *et al.*, 2008) over other spectral methods, NMR was selected to determine the polar metabolome fingerprints of untransformed NCo310 and UGD repressed transgenic sugarcane lines. A non-targeted approach was employed in order to determine if NMR could replace other spectral methods in the high throughput analysis of metabolites (Ward *et al.*, 2006; Scott *et al.*, 2010). GC-MS analysis was employed as a more targeted approach to identify metabolites which had significant concentration differences between untransformed NCo310 and transgenic sugarcane lines.

The UGD transgenic lines have great commercial potential due to the increase in sucrose content and thus, to ensure a sucrose content increase was still present, HPLC analysis and biochemical assays were employed to analyse stalk tissue sugar levels. The sucrose content of the UGD repressed transgenic and untransformed NCo310 plants were determined before the metabolomic or transcriptomic analysis in order to determine if the analysis was viable.

3.5.1 Sugarcane sucrose content is increased in the transgenic field trial plants

Sugarcane plants from the untransformed NCo310 and UGD repressed transgenic lines were grown under monitored conditions at SASRI. Sucrose, glucose and fructose content were determined *via* HPLC analysis (Figures 3.1 through 3.4). Several of the transgenic lines had significant increases in sucrose content when compared to the untransformed NCo310 line while others did not (Figure 3.1). Most notably was the sucrose content increase of more than 55% in the mature internode tissue of transgenic line UGD 7. Fructose and glucose content of both the mature and immature internode tissues were significantly reduced in several of the transgenic lines when compared to the untransformed NCo310 line (Figure 3.2). This analysis indicated that the UGD repressed transgenic lines still had a sucrose content increase after several years of propagation under field trial conditions (Bekker, 2007). The transgenic lines which showed the largest increase in sucrose content was selected for metabolome analysis. One concern is the glucose content of the plants which are significantly reduced but was previously significantly increased in several UGD repressed transgenic lines (Bekker, 2007). The decrease in glucose levels could be the result of the analysis technique as an enzymatic assay was previously used to determine the sugar content which is not as sensitive as HPLC determination (Schwald *et al.*, 1988). Another possibility could be that environmental changes induced the alterations in sugar content as observed (Gilbert *et al.*, 2005). The transcriptome analysis revealed transcripts involved in trehalose metabolism to be significantly differentially regulated which could indicate that sugar signalling is altered in the trangenic lines and thus give rise to altered hexose levels (Avonce *et al.*, 2005; Glassop *et al.*, 2007; Best *et al.*, 2011). The reduction in hexose levels is supported by reports on transgenic sugarcane in which neutral invertase activity was reduced which caused an increase in sucrose and a reduction in hexose levels (Rossouw *et al.*, 2010). A third hypothesis which is related to environmental changes is that the transgenic plants are genotypically not as well suited to the environmental conditions as the untransformed plants (Baafi and Safo-Kantanka, 2008; Panthee et al., 2012).

A follow-up analysis was done in April of 2011 to determine if the sucrose content increase was still present in the UGD repressed transgenic sugarcane lines. The follow-up HPLC analysis on the transgenic sugarcane failed to statistically prove any increase in sucrose content to be present in the UGD repressed sugarcane lines when compared to the wild type (Data not shown). The subsequent biochemical assay confirmed the HPLC results which indicated that no significant increase in sucrose content was present in the UGD repressed transgenic lines. Enzyme activity assays were unable to prove that UGD was still repressed due to high variability between repeats (data not shown). A possible reason for the deminished sucrose increase in the transgenic lines may be related to the sucrose storage potential of the plants which had not been reached (the plants were not ripe and were harvested ahead of full ripening) (Inman-Bamber, 1994). The harvesting period for NCo310 is at the end of winter in the late August to September period. Reports on sugarcane sucrose content reveals sucrose levels to be subject to seasonal and weather changes which would explain why the UGD repressed transgenic sugarcane lines showed no significant increase in sucrose content (Inman-Bamber, 1994). It could however be reasoned that even though the plants are not yet within the harvesting stage, a significant increase in sucrose should in theory still be observed due to the underlying molecular manipulations in the transgenic UGD lines. In defense of this, it might be that the mechanisms which allow the increased accumulation of sucrose only activate during the ripening stage when sucrose needs to be stored (Gilbert *et al.*, 2005). Bias can be introduced by the influence of severe environmental conditions on either of the untransformed NCo310 or transgenic lines. Sugarcane lines better suited to overcome harsh environmental conditions will out perform lines not as well suited to these conditions.

In order to confirm the results from the April 2011 sucrose content analysis, another sampling and sucrose content analysis was done in July of 2011. The analysis confirmed a significant increase in sucrose content of the mature internode tissue to be present for both UGD 7 and UGD 18 transgenic lines (Figure 3.3). The young internode tissue of transgenic line UGD 18 had a significant increase in sucrose content which was not observed in UGD 7 (Figure 3.3). Significant decreases in the mature internode tissue glucose and fructose content was observed in both of the transgenic lines (Figure 3.4). The decrease in glucose and fructose content is in concurrence with the observations of the 2009 sample set (Figure 3.2). Of interest is the significant increase in the fructose and glucose content of the young internodes of transgenic line UGD 18 (Figure 3.4). The increase is only observed in the young internode

tissue of UGD 18. The inconsistent fructose levels observed between the different sample sets complicate the explanation thereof.

In September of 2011, another sugarcane sampling was done in order to determine if the sucrose content increase was still present in the transgenic sugarcane lines. The samples consisted of only young internode material for the untransformed NCo310, UGD 7, UGD 8 and UGD 18 sugarcane lines. The UGD repressed transgenic sugarcane line, UGD 8, was included in the analysis as this transgenic line previously showed promise as a high sucrose accumulating line (Figure 3.1). A biochemical assay for sucrose content confirmed transgenic lines UGD 7 and UGD 18 to have significantly higher levels of sucrose compared to the untransformed NCo310 line (Figure 3.5). A decrease in sucrose level was however observed in line UGD 8 (Figure 3.5). The decrease in the sucrose content of transgenic line UGD 8 is not unexpected due to the complex genetics of sugarcane (Birch, 1997; Wei *et al.*, 2003). The suppression of silencing systems over a period in sugarcane is not an uncommon observation (Birch, 1997; Wei *et al.*, 2003; unpublished reports from this lab). The sucrose content analysis confirmed that at least two of the UGD repressed transgenic lines still have increased sucrose content when grown under field trial conditions and after several regeneration stages (Bekker, 2007).

The sucrose increase indicates that the repression of UGD is still active in the transgenic lines and has not undergone silencing in transgenic lines UGD 7 and UGD 18. It also indicates that the more plausible explanation for the low sucrose levels seen in the previous samplings could be that full ripeness of the sugarcane had not been reached (Birch, 1997; Wei *et al.*, 2003).

3.5.2 The NMR analysis identifies several important metabolites

In order to determine the polar metabolome fingerprint of the sugarcane plants, the 600 MHz Varian NMR at the central analytical facility at Stellenbosch University was employed. Metabolites which are significantly altered in the transgenic lines would aid in the understanding of how sugarcane is able to produce and store such high amounts of sucrose compared to other non-sucrose storing plants and how to further improve commercial sugarcane lines. The hypothesis was that the UGD repressed transgenic sugarcane lines would have significant changes in the metabolite levels of numerous metabolites. The sucrose content of the UGD repressed transgenic lines were observed to differ significantly from the

untransformed NCo310 lines and thus the same should also reflect in the metabolite fingerprint analysis.

The polar subset of the metabolome was extracted to reduce the number of metabolites to a more manageable size (Schauer and Fernie, 2006; Kruger *et al.*, 2008). The polar metabolite extraction process was met with some difficulties as the procedure required a large amount of fresh mass and several fine pH adjustments (Kruger *et al.*, 2008). The amount of dried metabolite obtained from 3 g of fresh mass was roughly 100 mg per sample. Extractions had to be done in batches of three or four samples due to availability of the NMR instrument for 2D HSQC spectra acquisition.

No internal standard was included as this might interfere with downstream analysis of the NMR spectral data as well as masking of metabolite peaks which would reduce the number of metabolites the analysis could detect (Kruger *et al.*, 2008; Hays and Thompson; 2009). In order to be able to accurately quantify a specific metabolite, an internal standard must be added and recovery studies done (Wishart, 2007; Xia *et al.*, 2009). As this study aimed only at detecting relative amounts, no internal standard was added. The metabolite levels could thus only be interpreted in relative amounts between samples (Wishart, 2007; Kruger *et al.*, 2008; Xia *et al.*, 2009). A chemometric approached was used to determine metabolite level differences between untransformed NCo310 and transgenic lines. The NMR spectra were grouped and overlaid according to their representative tissue groups and the spectral regions which had either strong signals or large differences in signals between sugarcane lines were designated as regions of interest (Figure 3.9).

The PLS-DA of the ROIs spectral data was capable of identifying an underlying feature in the polar metabolome of the untransformed NCo310 and UGD repressed plants which separates them from each other (Figure 3.10). However, the differences in the polar metabolome as based on the ROI spectral data indicated that the two groups were not significantly different from each other at the 5% threshold (Figure 3.10). Biological replicates are widely spread out from each other which indicated that the metabolite compositions and subsequent spectra have large differences, possibly due to pH deviations or alterations in plant growth and development (Krishnan *et al.*, 2004). Changes in plant growth and development are to be expected due to the nature of the growth conditions. Reports on metabolite extraction procedures for downstream analysis by NMR indicate that the perchloric acid extraction method is the most sensitive to pH changes (Lin *et al.*, 2007). The

pH differences of biological replicate metabolite extractions were observed by peak shifting in the NMR spectra (Lin *et al.*, 2007). Large pH changes have been shown to degrade pH sensitive metabolites during extraction procedures (Friedman and Jürgens, 2000).

The spectral data was analysed to determine which of the ROIs influenced the class discrimination of a sample as untransformed NCo310 or transgenic. The VIP analysis (Figure 3.11) determined the 15 most influential ROIs which were subsequently selected for peak picking and metabolite identification to identify which metabolites (Table 3.1) have altered levels between the untransformed NCo310 and UGD repressed transgenic lines. A volcano plot of all the ROIs indicated that no ROI had a fold change magnitude outside the 2, -2 threshold or within the p-value thresholds of 0.05. This result indicated that there was either no significant difference between the polar metabolomes of the untransformed NCo310 and transgenic plants or that there was too much variation between biological replicates.

When the metabolites identified by the NMR analysis were compared to those identified from GC-MS analysis (Figure 3.7 A; Figure 3.7 B), several metabolites were found to be similar between the two analyses. Sorbitol, sucrose, raffinose, *myo*-inositol and melezitose were identified in both analyses to be important in the separation of the untransformed NCo310 and transgenic groups. The identification of sucrose as one of the primary descriminating metabolites is expected as the UGD repressed transgenic lines were known to have increased sucrose content (Figure 3.1; Figure 3.3; Figure 3.4). The finding that *myo*-inositol is a primary discriminant metabolite is also an expected result as the repression of UGD should in theory alter carbon channeling through the SNOP and MIOP where the MIOP now needs to compensate for the reduced channel through the SNOP (Loewus and Murthy, 2000; Karkonen, 2005; Bekker, 2007). *Myo*-inositol, which is the primary substrate for the MIOP is thus expected to have altered levels. This is supported by the transcript analysis which found transcripts of the MIOP to be significantly differentially expressed.

Sugarcane is known to accumulate a number of different sugars of which sucrose is the primary (Glassop *et al.*, 2007). Melezitose is a non-reducing trisaccharide (α-fructofuranoside) with a sucrose backbone (Besset *et al.*, 2008) which has been reported to accumulate in sugarcane internodes as they mature (Glassop *et al.*, 2007). Reports indicate that legumes secrete the sugar from their leaves (Tsao *et al.*, 2001). Melezitose was at first thought to be synthesized by aphids due to the high levels found in the honeydew they produce (Owen, 1978). Aphids are thought to accumulate melezitose as an osmoprotectant

(Owen, 1978). Melezitose is a common constituent in traditional Chinese medicine used for the treatment of several diseases (Tsao *et al.*, 2001). However, not much is known about the function of melezitose in plants but reports indicate it to function as an osmoprotectant (Xu *et al.*, 2008). This would explain why melezitose levels are altered due to the increased levels of sucrose.

Sorbitol was found to have lower levels in the transgenic lines by the GC-MS analysis and as a primary metabolite in the NMR analysis. Sorbitol is a sugar alcohol and one of the primary photosynthetic products synthesized from either glucose by sorbitol-6-phosphate dehydrogenase (EC 1.1.1.200) or fructose which requires the conversion to glucose-6-phosphate by glucose-6-phosphate isomerase (EC 5.3.1.9). Sorbitol, together with sucrose, is also one of the major phloem translocated products. Sorbitol and sucrose biosynthesis pathways have been shown to not be independent from each other (Zhou *et al.*, 2002; Cheng *et al.*, 2005). Sorbitol-6-phosphate which is an intermediate of sorbitol synthesis has been reported to inhibit SPS and the glucose-6-phosphate activation thereof (Zhou *et al.*, 2002). Reports on transgenic apple in which sorbitol biosynthesis was down regulated revealed that this leads to increases in sucrose and starch content (Cheng *et al.*, 2005). The decrease in sorbitol levels observed in the UGD repressed transgenic lines is thus expected as sucrose levels are increased.

The finding that raffinose levels were significantly different between the untransformed NCo310 and transgenic sugarcane lines is in line with other reports which indicate that raffinose levels are linked to *myo*-inositol, sucrose and galactinol levels (Carner *et al.*, 2004). Both sucrose and *myo*-inositol levels are altered in the UGD repressed lines and thus the alteration of raffinose levels would be expected.

Several other metabolites were identified by the peak picking of the 15 ROIs but which were not identified by the GC-MS analysis. Two metabolites of interest to this study are the strong signalling sugar trehalose and the hemicellulose backbone sugar xylose. The importance of trehalose has already been discussed and xylose is an important constituent of the cell wall. Transcript data suggest that xylose breakdown is reduced but no alterations in synthesising genes are observed. Further analysis is required in order to verify that these metabolites have significantly altered levels between the untransformed NCo310 and UGD repressed transgenic sugarcane lines. The vision was that more metabolites would have been identified by the NMR analysis. The large amount of fresh mass needed for the analysis of low level

metabolites made multiple metabolite extractions non-feasible which is one of the major draw backs of metabolite analysis *via* NMR.

Even though the NMR data is statistically not significant, the underlying basis for the changes between the untransformed NCo310 and UGD transgenic lines can still be visualised. There is evidence which suggest an alteration in the cell wall and sugar components of the transgenic sugarcane plants compared the untransformed NCo310 plants.

3.5.3 The GC-MS results verify the NMR analysis

Due to the low significance of the NMR results, GC-MS analysis was employed to validate the results. The amount of fresh mass needed for GC-MS analysis is a few milligrams in contrast to the three grams needed for NMR analysis. However the technique is not applicable to heat-labile metabolites and additional costs are involved compared to NMR.

As with the NMR data analysis, the PLS-DA was able to detect underlying features which separate the untransformed NCo310 and transgenic samples according to their polar metabolome (Figure 3.6 A; Figure 3.6 B). The volcano plot identified relatively few metabolites which had a fold change magnitude (Figure 3.8) greater than 2 and a p-value within the 0.05 threshold between the untransformed NCo310 and transgenic groups. The metabolites which had the largest influence on the grouping of a sample according to the relative abundance of the metabolite were glucuronic acid, raffinose, sucrose, sorbitol and myo-inositol along with several amino acids as determined by the VIP and random forests analysis (Figure 3.7 A; Figure 3.7 B). A permutation test revealed that the "real class" did not fall within the random class and thus the class discrimination of a sample as untransformed NCo310 or transgenic is statistically significant.

The VIP and random forests analysis both identify sorbitol levels to be the primary factor when assigning a sample as untransformed NCo310 or transgenic. The UGD repressed transgenic sugarcane lines had reduced sorbitol levels compared to the untransformed NCo310 sugarcane line. This is confirmed by the NMR analysis which identified sorbitol as one of the metabolites which had differences in concentration between the untransformed NCo310 and UGD repressed lines. Higher levels of both sucrose and raffinose were observed in the UGD repressed lines while *myo*-inositol levels are reduced when compared to the untransformed NCo310 lines. These findings are supported by the NMR analysis. Interesting is the finding that glucuronic acid levels are increased in the transgenic lines as proposed by

the previous study on the UGD repressed sugarcane lines (Bekker, 2007). This evidence together with the reduced levels of *myo*-inositol supports the hypothesis that the MIOP enzymes are up-regulated as glucuronic acid is the downstream product of this pathway (Bekker, 2007).

A metabolite detected by GC-MS analysis but not the NMR analysis was galactose. Galactose is an epimer of glucose and toxic to cells at low levels (Maretzki and Thom, 1977). Reports indicate that adding free galactose to plants lead to the incorporation of the metabolite into the cell wall or use in sucrose synthesis (Maretzki and Thom, 1977). The GC-MS results indicate reduced levels of galactose in the UGD repressed transgenic lines which could suggest increased incorporation into the cell wall or increased breakdown for use in the synthesis of other compounds. Transcript data to confirm either of the possibilities are however not observed and thus this is only speculation.

Further investigation of metabolites involved in plant cell wall function revealed a decrease in the levels of trans-p-coumaric (4-hydroxycinnamic) acid in the UGD repressed plants. Trans-p-coumaric acid is an early phenylpropanoid pathway substrate for lignin and flavonoid biosynthesis which is produced from coumaric acid by cinnamate 4-hydroxylase (EC 1.14.13.11; C4H). Reports have shown that reducing C4H activity leads to decreases in lignin content (Schilmiller *et al.*, 2009; Bjurhager *et al.*, 2010). The reduction in the levels of p-coumaric acid could indicate that the enzyme activity of C4H is reduced which would lead to a reduction in lignin content of the UGD repressed transgenic lines.

3.5.4 Uronic acid content support the MIOP hypothesis

The previous study on the UGD repressed transgenic sugarcane lines reported that most of the tissues tested had significant increases in uronic acid content. Uronic acid content was investigated because uronic acids are derived from the products of UGD (Bekker, 2007). The increase in uronic acid content was attributed to the up-regulation of the MIOP (Bekker, 2007). The uronic acid content of the 2009 sample set was determined in order te verify that the sugarcane lines still had an increase.

Uronic acid content of the cell walls of both young and mature sugarcane internodes revealed significant increases in uronic acid levels to be present in the transgenic sugarcane lines (Figure 3.12). The magnitude of the increase in uronic acid levels weren't as great as those observed in the previous study.

90

The significant increase in uronic acid content of both the young and mature internode tissues together with the increase in sucrose content shows that UGD is still repressed in the transgenic sugarcane as similar metabolite changes were observed in the previous study. It is presumed that based on the uronic acid levels that the MOIP is still up-regulated in the transgenic sugarcane lines.

3.6 Conclusion

The UGD repressed transgenic lines were previously reported to have increased sucrose and cell wall uronic acid levels. The increase in the cell wall uronic acid levels was attributed to the up-regulation of the MIOP enzymes (Bekker, 2007).

The metabolite fingerprints of the transgenic and untransformed NCo310 sugarcane lines were constructed in order to determine if the UGD repressed transgenic sugarcane lines still had an increase in sucrose and cell wall uronic acid levels. NMR, GC-MS, HPLC and enzymatic assays were employed to construct the metabolite profiles. To better understand the metabolic changes occuring in the UGD repressed lines the metabolic fingerprints were compared to each other.

The NMR results were however not statistically significant. The PLS-DA model was unable to distinguish between untransformed NCo310 and transgenic samples although the two groups did form two clusters. The low significance of the NMR results is attributed primarily to pH changes as NMR metabolite analysis and the perchloric acid extraction method have been reported to be extremely sensitive to changes in pH (Krishnan *et al.*, 2004 Kruger *et al.*, 2008).

The NMR analysis identified several metabolites which had altered levels between the untransformed NCo310 and transgenic sugarcane lines. Sucrose was identified as one of the primary metabolites which had altered levels between the untransformed NCo310 and transgenic sugarcane lines which is to be expected. Two other metabolites of interest are sorbitol and *myo*-inositol which are expected to have altered levels due to the increase in sucrose levels and the repression of UGD. Interesting is the finding that raffinose is identified as one of the primary metabolites in the analysis. It is speculated that the increase in sucrose led to the increase observed in the amount of raffinose as reported by other studies (Carner *et al.*, 2004).

Previous reports suggest that the levels of lignin and cellulose are altered in the UGD repressed transgenic sugarcane lines (Bekker, 2007). The NMR analysis was however unable to identify either lignin or cellulose and thus the reports could not be validated.

The aim of the non-targeted NMR approach was to identify the maximum number of compounds without the interference of internal standards (Krishnan *et al.*, 2004). The relatively low sensitivity of the technique to metabolite levels reduced the number of metabolites observed (Table 3.1). It was expected that more metabolites would be identified by the NMR analysis than was detected. The conclusion is that a non-targeted NMR approach to metabolome profiling is applicable with enough replicates, starting material and more robust extraction method (Krishnan *et al.*, 2004; Lin *et al.*, 2007).

Gas chromatography-mass spectrometry was employed to investigate the polar metabolome in a more targeted approach than the NMR analysis. Fewer compounds could be identified by the GC-MS analysis but the results had increased significance compared to the NMR results.

The PLS-DA analysis clearly seperated the untransformed NCo310 and transgenic groups from each other based on their metabolite profiles. The GC-MS analysis identified several of the metabolites which were also identified by the NMR analysis as important metabolites. The GC-MS analysis was also able to assign a relative metabolite level to each of the important metabolites which could not be done with the NMR analysis. The finding that the NMR and GC-MS analysis identified several of the same metabolites as important verifies the results of the two techniques.

Glucuronic acid, raffinose, sucrose, sorbitol, p-coumaric acid and myo-inositol were identified by the GC-MS analysis to be important metabolites in the descrimination of a sample as untransformed NCo310 or transgenic. The reduced levels of p-coumaric acid suggests an alteration in the lignin levels of the transgenic sugarcane plants while the reduction in *myo*-inositol levels indicate that the MIOP is up-regulated to accomodate the repression of the SNOP as previously suggested (Bekker, 2007).

The analysis of the uronic acid content of the tissue cell walls indicated the increase observed in the previous report to still be present in the new generation of plants. This result together with the decrease in the *myo*-inositol levels supports the assumption that the MIOP is up-regulated in the UGD repressed transgenic sugarcane lines.

Chapter 4: Summary

Sugarcane, the primary crop of the South African sugar industry was transformed to down regulate UDP-Glucose dehydrogenase previously (Bekker, 2007). UGD repressed transgenic sugarcane lines had a significant sucrose content increase and an altered cell wall composition which made the bagasse more easily fermentable (unpublished results). These findings made the plants not only applicable to the sugar industry but also to the biofuel production sector. The altered cell wall of the UGD repressed plants was an unforeseen benefit as the aim was only to increase the sucrose content.

This study was tasked to delve deeper into the genetic and metabolic nature of the UGD repressed transgenic sugarcane lines in order to increase the current understanding of sugarcane metabolism.

The aim of this study was to generate transcriptional profiles of the wild type (NCo310) and UGD repressed transgenic sugarcane young and mature internode tissues *via* microarray technology from Affymetrix®. The second part of this study was to investigate the polar metabolome of the UGD repressed sugarcane lines. The aim of this section was to generate polar metabolome fingerprints for the untransformed NCo310 and transgenic lines. Nuclear Magnetic Resonance spectroscopy (NMR) and Gas chromatography linked mass spectroscopy (GC-MS) was employed for this analysis to determine if NMR could be applied to the high throughput fingerprinting of the polar metabolome.

The mature internodes (internode 10) of control (untransformed NCo310) and UGD repressed transgenic sugarcane plants growing under field trial conditions were harvested at SASRI and the transcriptional and metabolic profiles investigated. The transcriptional analysis identified several genes to be significantly differentially expressed between the wild type and transgenic lines which have previously been reported to be affected by increases in sucrose content in sugarcane (Watt *et al.*, 2005; Casu *et al.*, 2007). The transcriptional profiles of the two transgenic lines investigated differed in the number of transcripts differentially expressed which can be attributed to each of the UGD repressed transgenic lines being unique (Bekker, 2007). The metabolite profiles generated *via* NMR and GC-MS analysis supported each other as well as the transcriptional data. This analysis supports the results of the previous analysis on the UGD repressed sugarcane lines (Bekker, 2007).

Investigations of the UGD repressed sugarcane lines revealed significant increases in sucrose content which confirmed that the repression of UGD was still active. Transcript analysis suggests that the increase in sucrose is due to decreased breakdown of sucrose and/or increased synthesis (Table 2.2). Starch metabolism was observed to be altered due to transcripts involved in starch synthesis being down regulated and those of starch degradation being up-regulated. The accumulation of sucrose is presumed to be in part due to altered sugar signalling mediated by altered levels of the strong signalling compounds trehalose and trehalose-6-phosphate. This is supported by the finding that transcripts involved in trehalose metabolism were significantly differentially expressed in the UGD repressed transgenic lines. However, metabolite analysis was unable to confirm that significant changes in trehalose or trehalose-6-phosphate levels were present. Furthermore, several transcripts representing cellulose synthase are down regulated which suggest that the requirement for components for cell wall synthesis is reduced and available for sucrose synthesis.

The hypothesis is that the sugar nucleotide oxidation pathway (SNOP) of which UGD is a central part (Grof and Campbell, 2001; Grof *et al.*, 2007) is now bypassed by the *myo*-inositol oxidation pathway (MIOP) to supply the precursors (Loewus and Murthy, 2000; Karkonen, 2005; Bekker, 2007) needed for the synthesis of cell wall components as observed by increases in MIOP enzyme transcripts of MIPS (Table 2.2). The metabolite data supports this hypothesis as *myo*-inositol is identified by both GC-MS and NMR to have altered levels between the untransformed NCo310 and UGD repressed transgenic sugarcane lines. A significant decrease in *myo*-inositol levels (Table 2.2; Table 3.1; Figure 3.7) were observed in the UGD repressed transgenic sugarcane lines. Further investigation of the MIOP transcripts revealed that transcrips of glucuronokinase was however not significantly altered in the transgenic sugarcane lines and no other transcripts of the MIOP could be identified from the microarray probe sets. MIOX, another MIOP enzyme, was previously reported to be significantly up-regulated in the UGD repressed trasgenic lines (Bekker, 2007).

The switch from the SNOP to the MIOP is presumed to lead to a decrease in lignin content due to reduced availability of substrates. Several transcripts related to genes involved in flavonoid and lignin biosynthesis was significantly differentially expressed which forms the basis for this assumption. The increased fermentability of the bagasse and the reduction in p-coumaric acid levels support this assumption.

Future projects in this lab will focus on MIOP enzymes and consider them for future modification studies with the aim to increase sugarcane overall crop value. As stated by other studies on sugarcane improvement, the alteration of genes involved in sugar signalling are also potential future targets for genetic manipulation.

This work forms part of a larger project aimed at increasing sugarcane sucrose content. This work together with other is being compiled into a scientific article with the running title : Decreasing UDP-Glucose dehydrogenase activity in sugarcane leads to profound increases in sucrose accumulation.

References

Adjaye J, Herwig R, Herrmann D, Wruck W, BenKahla A, Brink TC, Nowak M, Carnwath JW, Hultschig C, Niemann H, Lehrach H (2004) Cross-species hybridisation of human bovine orthologous genes on high density cDNA microarrays. BMC Genomics **5**: 83-94

Amor Y, Haigler CH, Johnson S, Wainscott M, Delmer DP (1995) A membrane-associated form of sucrose synthase and its potential role in synthesis of cellulose and callose in plants. PNAS **92**: 9353-9357

Aitken A (1996) 14-3-3 and its possible role in co-ordinating multiple signalling pathways. Trends in CELL BIOLOGY **6**: 341-347

Akiyama K, Chikayama E, Yuasa H, ShimadaY, Tohge T, Shinozaki K, Hirai MY, Sakurai T, Kikuchi J, Saito K (2008) PRIMe: a Web site that assembles tools for metabolomics and transcriptomics. In Silico Biology **8**: 0027

Altschul SF, Gish, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J. Mol. Biol. **215**: 403-410

Avonce N, Leyman B, Thevelein J, Iturriaga G (2005) Trehalose metabolism and glucose sensing in plants. Biochemical Society Transactions **33**: 276-279

Avonce N, Mendoza-Vargas A, Morett E, Iturriaga G (2006) Insights on the evolution of trehalose biosynthesis. BMC Evolutionary Biology **6**: 109

Azumi Y, Watanabe A (1991) Evidence for a senescence-associated gene induced by darkness. Plant Physiology **95**: 577-583

Baafi E, Safo-Kantanka O (2008) Genotype-Environment Interaction (GXE) Effects on Some Major Rheological Properties of Cassava (Manihot esculenta, Crantz). American Journal of Food Technology **3**: 214-219

Barrière Y, Ralph J, Méchin V, Guillaumie S, Grabber JH, Argillier O, Chabbert B, Lapierre C (2004) Genetic and molecular basis of grass cell wall biosynthesis and degradability. II. Lessons from brown-midrib mutants. Comptes Rendus Biologies **327**: 847-860

Baxter CJ, Foyer CH, Turner J, Rolfe SA, Quick WP (2003) Elevated sucrose-phosphate synthase activity in transgenic tobacco sustains photosynthesis in older leaves and alters development. Journal of Experimental Botany **54**: 1813-1820

Bekker JPI (2007) Genetic Manipulation of the cell wall composition of sugarcane. PhD dissertation, Stellenbosch University, Stellenbosch, South Africa.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B **57**: 289-300

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. The Annals of Statistics **29**: 1165-1188

Bergmeyer HU, Bernt E (1974) Sucrose. In Methods of Enzymatic Analysis. 2nd edition Vol. 3: 1176-1179. Ed Bergmeyer HU. Verlag Chemie, Weinheim/Academic Press, New York

Besseau S, Hoffmann L, Geoffroy P, Lapierre C, Brigitte P, Legrand M (2007) Flavonoid accumulation in *Arabidopsis* repressed in lignin synthesis affects auxin transport and plant growth. The Plant Cell **19**: 148-162

Besset C, Chambert S, Queneau Y, Kerverdo S, Rolland H, Guilbot J (2008) Reactivity of melezitose and raffinose under Mitsunobu reaction conditions. Carbohydrate Research **343**: 929-935

Best M, Koenig K, McDonald K, Schueller M, Rogers A, Ferrieri RA (2011) Inhibition of trehalose breakdown increases new carbon partitioning into cellulosic biomass in *Nicotiana tabacum*. Carbohydrate Research **346**: 595-601

Bieleski RL (1960) The physiology of sugar-cane III. Characteristics of sugar uptake in slices of mature and immature storage tissue. Australian Journal of Biological Sciences **13**: 203 – 220

Bieleski RL (1964) The Problem of Halting Enzyme Action When Extracting Plant Tissues. Analytical Biochemistry **9**: 431-442

Bijlsma S, Bobeldijk I, Verheij ER, Ramaker R, Kochhar S, Macdonald IA, van Ommen B, Smilde AK (2006) Large-scale human metabolomics studies: a strategy for data (pre-) processing and validation. Analytical Chemistry **78**: 567-574

Birch RG (1997) Plant Transformation: problems and strategies for practical application. Annual Review of Plant Physiology and Plant Molecular Biology **48**: 297-326

Bjurhagen I, Olsson A, Zhang B, Gerber L, Kumar, Berglund, Burgert I, Sundberg B, Salmen L (2010) Ultrastructure and mechanical properties of *Populus* wood with reduced lignin content caused by transgenic down-regulation of cannamate 4-hydroxylase. Biomacromolecules **11**: 2359-2365

Bolstad BM, Irizarry RA, Åstrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics **19**: 185-193

Bosch S (2005) Trehalose and carbon partitioning in sugarcane. PhD dissertation, Stellenbosch University, Stellenbosch, South Africa.

Botha FC, Black KG (2000) Sucrose phosphate synthase and sucrose synthase activity during maturation of internodal tissue in sugarcane. Australian Journal of Plant Physiology **27**: 81–85

Bonin CP, Potter I, Vanzin GF, Reiter W (1997) The *MUR1* gene of *Arabidopsis thaliana* encodes an isoform of GDP-D-mannose-4,6-dehydratase, catalysing the first step in the *de novo* synthesis of GDP-L-fucose. PNAS **94**: 2085-2090

Breiman L (2001) Random Forests. Machine Learning **45**: 5-32

Broeckling CD, Reddy IR, Duran AL, Zhao X, Sumner LW (2006) MET-IDEA: Data Extraction Tool for Mass Spectrometry-Based Metabolomics. Analytical Chemistry **78**: 4334-4341

Brodsky L, Moussaieff A, Shahaf N, Aharoni A, Rogachev I (2010) Evaluation of peak picking quality in LC-MS metabolomics data. Analytical Chemistry **82**: 9177-9187

Brown SA, Neish AC (1955) Shikimic acid as a precursor in lignin biosynthesis. Nature **175**: 688-689

Brown AM, Jr. (1996) The biosynthesis of cellulose. Journal of Macromolecular Science 10: 1345-1373

Bull TA, Glasziou KT (1963) The evolutionary significance of sugar accumulation in *Saccharum*. Australian Journal of Biological Sciences **16**: 737-742

Burget EG, Reiter W (1999) The *mur4* mutant of Arabidopsis is partially defective in the de novo synthesis of uridine diphospho L-Arabinose. Plant Physiology **121**: 383-389

Burget EG, Verma R, Molhoj M, Reiter W (2003) The biosynthesis of L-arabinose in plants: molecular cloning and characterization of a Golgi-localized UDP-D-xylose 4-epimerase encoded by the the *MUR4* gene of Arabidopsis. The Plant Cell **15**: 523-531

Butaye KMJ, Cammue BPA, Delauré SL, De Bolle MFC (2005) Approaches to minimize varaition of transgene expression in plants. Molecular Breeding **16**: 79-91

Carmi N, Zhang G, Petreikov M, Gao Z, Eyal Y, Granot D, Schaffer AA (2003) Cloning and functional expression of alkaline α-galactosidase from melon fruit: similarity to plant SIP proteins uncovers a novel family of plant glycosyl hydrolases. The Plant Journal **33**: 97-106

Castleden CK, Aoki N, Gillespie VJ, MacRae EA, Quick WP, Buchner P, Foyer CH, Furbank RT, Lunn JE (2004) Evolution and function of the sucrose-phosphate synthase gene families in wheat and other grasses. Plant Physiology **135**: 1753-1764

Casu RE, Dimmock CM, Chapman SC, Grof CPL, McIntyre CL, Bonnett GD, Manners JM (2004) Identification of differentially expressed transcripts from maturing stem of sugarcane by in silico analysis of stem expressed sequence tags and gene expression profiling. Plant Molecular Biology **54**: 503-517

Casu RE, Jarmey JM, Bonnett GD, Manners JM (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 and Integrative Genomics **7**: 153-167

Català C, Rose JKC, Bennett AB (2000) Auxin-regulated encoding cell wall-modifying proteins are expressed during early tomato fruit growth. Plant Physiology **122**: 527-534

Charalampous FC, Lyras C (1957) Biochemical studies on inositol. IV. Conversion of inositol to glucuronic acid by rat kidney extracts. Journal of Biological Chemistry **228**: 1-13

Chen S, Hajirezaei M, Börnke F (2005) Differential expression of sucrose-phosphate synthase isoenzymes in Tobacco reflects their functional specialization during dark-governed starch mobilization in source leaves. Plant Physiology **139**: 1163-1174

Chen F, Dixon RA (2007) Lignin modification improves fermentable sugar yields for biofuel production. Nature Biotechnology **25**: 759-761

Chen C, Grennan K, Badner J, Zhang D, Gershon E, Jin L, Liu C (2011) Removing batch effects in analysis of expression microarray data: An evaluation of six batch adjustment methods. PLoS ONE **6**: 1-10

Cheng L, Zhou R, Reidel EJ, Sharkey TD, Dandekar AM (2005) Antisense inhibition of sorbitol synthesis leads to up-regulation of starch synthesis without altering CO_2 assimilation in apple leaves. Planta **220**: 767-776

Chikayama E, Suto M, Nishihara T, Shinozaki K, Kikuchi L (2008) Systematic NMR analysis of stable isotope labeled metabolite mixtures in plant and animal systems: coarse grained views of metabolic pathways. PLoS ONE **3**: e3805

Chikayama E, Sekiyama Y, Okamoto M, Nakanishi Y, Tsuboi Y, Akiyama K, Saito K, Shinozaki K, Kikuchi J (2010) Statistical indices for simultaneous large-scale metabolite detections for a single NMR spectrum. Analytical Chemistry **82**: 1653-1658

Colombo R, Lancas FM, Yariwake JH (2006) Determination of flavonoids in cultivated sugarcane leaves, bagasse, juise and in transgenic sugarcane by liquid chromatography-UV detection. Journal of Chromatography A **1103**: 118-124

Cosgrove DJ (2000) Expansive growth of plant cell walls. Plant Physiology and Biochemistry **38**: 109-124

Cosgrove DJ (2006) Growth of the plant cell wall. Nature Reviews **6**: 850-861

Cotelle V, Meek SEM, Provan F, Milne FC, Morrice N, MacKintosh C (2000) 14-3-3s regulate global cleavage of their diverse binding partners in sugar-starved *Arabidopsis* cells. The EMBO Journal. **19**: 2869-2876

Cui Q, Lewis IA, Hegeman AD, Anderson ME, Li J, Schulte CF, Westler WM, Eghbalnia HR, Sussman MR, Markley JL (2008) Metabolite identification *via* the Madison Metabolomics Consortium Database. Nature Biotechnology **26**: 162 – 164

Dallas PB, Gottardo NG, Firth MJ, Beesley AH, Hoffmann K, Terry PA, Freitas JR, Boag JM, Cummings AJ, Kees UR (2005) Gene expression levels assessed by oligonucleotide microarray analysis and quantitative real-time RT-PCR – how well do they correlate? BMC Genomics **6**: 59-69

Dauwe R, Morreel K, Goeminne G, Gielen B, Rohde A, Van Beeumen J, Ralph J, Boudet AM, Kopka J, Rochange SF, et al (2007) Molecular phenotyping of lignin-modified tobacco reveals associated changes in cell-wall metabolism, primary metabolism, stress metabolism and photorespiration. The Plant Journal **52**: 263-285

Dickonson DB (1982) Occurrence of glucuronokinase in various plant tissues and comparison of enzyme activity of seedlings and green plants. Phytochemistry **21**: 843-844

Doblin MS, Pettolino F, Bacic A (2010) Plant cell walls: the skeleton of the plant world. Functional Plant Biology **37**: 357–381

Doblin MS, Vergara CE, Read S, Newbigin E, Bacic A (2003) in: Rose JKC, The Plant Cell Wall. Annual Plant Reviews **8**: 183-222. Blackwell publishing, Oxford

Dolezal O, Cobbett C (1991) Arabinose kinase-deficient mutant of *Arabidopsis thaliana*. Plant Physiology **96**: 1255-1260

Dunstan WR, HENRY TA (1902) Cyanogenesisin plants II. The great millet Sorghum zullgare. Phil Trans, Royal Soc. (London) Ser. A. **199**: 399-410

Ehness R, Ecker M, Godt DE, Roitsch T (1997) Glucose and stress independently regulate source and sink metabolism and defense mechanisms *via* signal transduction pathways involving protein phosphorylation. The Plant Cell **9**: 1825-1841

Endres S, Tenhaken R (2011) Down-regulation of the *myo*-inositol oxygenase gene family has no effect on cell wall composition in Arabidopsis. Planta **157**: 157-169

Feingold DS (1982) Aldo (and keto) hexoses and uronic acids, in: Loewus FA and Tanner W, Plant Carbohydrates I, Encyclopedia of Plant Physiology, New Series, Vol. 13A, Springer Verlag, Berlin

Fly RD (2010) Approaches for the study of leaf carbohydrate metabolic compartmentation in *Arabidopsis thaliana*. MSc dissertation, Stellenbosch University, Stellenbosch, South Africa.

Friedman M, Jürgens HS (2000) Effect of pH on the stability of plant phenolic compounds. Journal of Agricaltural and Food Chemistry **48**: 2101-2110

Fujiki Y, Ito, M, Nishida I, Watanabe A (2000) Multiple signalling pathways in gene expression during sugar starvation. Pharmacological analysis of *din* gene expression in suspension-cultured cells of Arabidopsis. Plant Physiology **124**: 1139-1148

Fujiki Y, Yoshikawa Y, Sato T, Inada N, Ito M, Nishida I, Watanabe A (2001) Dark-inducible genes from *Arabidopsis thaliana* are associated with leaf senescence and repressed by sugars. Physiologia Plantarum **111**: 345-352

Garg AK, Kim J, Owens TG, Ranwala AP, Choi YD, Kochian LV, Wu RJ (2002) Trehalose accumulation in rice plants confers high tolerance levels to different abiotic stresses. PNAS **99**: 15898-15903

Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) affy-analysis of Affymetrix GeneChip data at the probe level. Bioinformatics **20**: 307-315

Geigenberger P, Langenberger S, Wilke I, Heineke D, Heldt HW, Stitt M (1993) Sucrose is metabolised by sucrose synthase and glycolysis within the phloem complex of *Ricinus communis* L. seedlings. Planta **190**: 446-453

Geigenberger P (2011) Regulation of starch biosynthesis in repsonse to a fluctuating environment. Plant Physiology **155**: 1566-1577

Georgi K, Boos KS (2006) Multidimensional on-line SPE for undisturbed LC-MS-MS analysis of basic drugs in biofluids. Chromatographia **63**: 523-531

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH,

Zhang J (2004) Bioconductor: Open software development for computational biology and bioinformatics. Genome Biology **5**: R80

Gibon Y, Blaesing OE, Hannemann J, Carillo P, Höhne M, Hendriks JHM, Palacios N, Cross J, Selbig J, Stitt M (2004) A robot-based platform to measure multiple enzyme activities in Arabidopsis using a set of cycling assays: comparison of changes of enzyme activities and transcript levels during diurnal cycles and in prolonged darkness. The Plant Cell **16**: 3304-3325

Gilbert RA, Shine JM, Miller JD, Rice RW, Rainbolt CR (2005) The effect of genotype, environment and time of harvest on sugarcane yields in Florida, USA. Field Crops Research **95**: 156–170

Glassop D, Roessner U, Bacic A, Bonnett GD (2007) Changes in the sugarcane metabolome with stem development. Are they related to sucrose accumulation? Plant Cell Physiology **48**: 573-584

Glasziou KT, Gayler KR (1972) Storage of sugars in stalks of sugar cane. Botanical Review **38**: 471-490

Göhlmann H, Talloen W (2009) Gene expression studies using Affymetrix microarrays. CRC Press, USA.

Gressel J, Zilberstein A (2003) Let them eat (GM) straw. TRENDS in Biotechnology **21**: 525-530

Gressel J (2008) Transgenics are imperative for biofuel crops. Plant Science **174**: 246-263

Grof CPL, Campbell JA (2001) Sugarcane sucrose metabolism: scope for molecular manipulation. Australian Journal of Plant Physiology **28**: 1-12

Grof CPL, So CTE, Perroux JM, Bonnett GD, Forrester RI (2006) The five families of sucrose-phosphate synthase genes in Saccharum spp. are differentially expressed in leaves and stem. Functional Plant Biology **33**: 605–610

Grof CPL, Albertson PL, Bursle J, Perroux JM, Bonnett GD, Manners M (2007) Sucrose-phosphate synthase, a biochemical marker of high sucrose accumulation in sugarcane. Crop Science **47**: 1530-1539

Hanson J, Hanssen M, Wiese1 A, Hendriks MMWB, Smeekens S (2008) The sucrose regulated transcription factor bZIP11 affects amino acid metabolism by regulating the expression of ASPARAGINE SYNTHETASE1 and PROLINE DEHYDROGENASE 2. The Plant Journal **53**: 935-949

Harper AD, Bar-Peled M (2002) Biosynthesis of UDP-Xylose. Cloning and characterization of a novel Arabidopsis gene family, *UXS*, encoding soluble and putative membrane-

bound UDP-Glucuronic acid decarboxylase isoforms. Plant Physiology **130**: 2188-2198

Hassan SF (2008) Development of sugar industry in Africa. Sugar Tech **10**: 197-203

Hays PA, Thompson RA (2009) A processing method enabling the use of peak height for accurate and precise proton NMR quantification. Magnetic Resonance in Chemistry **47**: 819-824

Hinterberg B, Klos C, Tenhaken R (2002) Recombinant Udp-glucosse dehydrogenase from soybean. Plant Physiology and Biochemistry. **40**: 1011-1017

Higuchi T (1990) Lignin biochemistry: Biosynthesis and biodegradation. Wood Science and Technology **24**: 23-63

Hoffmann L, Maury S, Martz F, Geoffroy P, Legrand M (2003) Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism/ The Journal of Biological Chemistry **278**: 95-103

Holton TA, Cornish EC (1995) Genetics and biochemistry of anthocyanin biosynthesis. The Plant Cell **7**: 1071-1083

Hu CG, Honda C, Kita M, Zhang Z, Tsuda T, Moriguchi T (2002) A simple protocol for RNA isolation from fruit trees containing high levels of polysaccharides and polyphenol compounds. Plant Molecular Biology Reporter **20**: 69a-69g

Huggett JH, Dheda K, Bustin S, Zumla A (2005) Real-time RT-PCR normalisation; strategies and considerations. Genes and Immunity **6**: 279-284

Inman-Bamber NG (1994) Effect of age and season on components of yield of sugarcane in South Africa. Proceedings of the Congress of the South African Sugar Technologists Association **68**: 23-27

Ishimaru K, Ono K, Kashiwagi T (2004) Identification of a new gene controlling plant height in rice using the candidate-gene strategy. Planta **218**: 388-395

Illovo Sugar Limited (2011) http://www.illovo.co.za/World_of_sugar/Sugar_Statistics/ International.aspx

Jacobsen KR, Fisher DG, Moore PH (1992) Developmental changes in the anatomy of the sugarcane stem in relation to phloem unloading and sucrose storage. Botanica Acta **105**:70–80

Johansson H, Sterky F, Amini B, Lundeberg J, Kleczkowski LA (2002) Molecular cloning and characterization of a cDNA encoding poplar UDP-glucose dehydrogenase, a key gene of hemicellulose/pectin formation. Biochimica et Biophysica Acta **1576**: 53-58

Kammerer B, Fischer K, Hilpert B, Schubert S, Gutensohn M, Weber A, Flügge U (1998) Molecular characterization of a carbon transporter in plastids from heterotrophic tissues: the glucose 6-phosphate/phosphate antiporter. Plant Cell **10**: 105-117

Kandel-Kfir M, Damari-Weissler H, German MA, Gidoni D, Mett A, Belausov E, Petreikov M, Adir N, Granot D (2006) Two newly identified membrane-associated and plastidic tomato HXKs: characteristics, predicted structure and intracellular localization. Planta **224**: 1341-1352

Kanehisa M, Goto S (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research **28**: 27-30

Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2009) KEGG for representing and analysis of molecular networks involving diseases and drugs. Nucleic Acids Research **38**: D355-D360

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2011) KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Research **40**: D109-D114

Kärkönen A (2005) Biosynthesis of UDP-GlcA: *via* UDPGDH or the myo-inositol oxidation pathway?. Plant Biosystems **139**: 46-49

Kärkönen A, Murigneux A, Martinant J, Pepey E, Tatout C, Dudley BJ, Fry SC (2005) UDP-glucose dehydrogenases of maize: a role in cell wall pentose biosynthesis. Biochemical Journal **391**: 409-415

Keller F, Pharr DM (1996) Metabolism of carbohydrates in sinks and sources: galactosyl-sucrose oligosaccharides. In *Photoassimilate Distribution in Plants and Crops* (Zamski E, Schaffer AA) eds. New York, pp. 115-184

Keller R, Brearley CA, Trethewey RN, Müller-Röber B (1998) Reduced inositol content and altered morphology in transgenic potato plants inhibited for 1D-*myo*-inositol-3-phosphate synthase. The Plant Journal **16**: 403-410

Kemsley EK, Le Gall G, Dainty JR, Watson AD, Harvey LJ, Tapp HS, Colquhoun IJ (2007) Multivariate techniques and their application in nutrition: a metabolomics case study. British Journal of Nutrition **98**: 1-14

Kendziorski C, Irizarry RA, Chen K, Haag JD, Gould MN (2005) On the utility of pooling biological samples in microarray experiments. PNAS **102**: 4252-4257

Klinghammer M, Tenhaken R (2007) Genome-wide analysis of the UDP-glucose dehydrogenase gene family in *Arabidopsis*, a key enzyme for matrix polysaccharides in cell walls. Journal of Experimental Botany **13**: 3609-3621

Kohli A, Gahakwa D, Vain P, Laurie DA, Christou P (1999) Transgene expression in rice engineered through particle bombardment: molecular factors controlling stable expression and transgene silencing. Planta **208**: 88-97

Kopka J, Schauer N, Krueger S, Birkemeyer C, Usadel B, Bergmüller E, Doermann P, Weckwerth W, Gibon Y, Stitt M, Willmitzer L, Fernie AR, Steinhauser D (2005) GMD@CSB.DB: The Golm Metabolome Database. Bioinformatics **21**: 16535-1638

Kotake T, Yamaguchi D, Ohzono H, Hojo S, Kaneko S, Ishida H, Tsumuraya Y (2004) UDP-sugar pyrophosphorylase with broad substrate specificity toward various monosaccharide 1-phosphates from pea sprouts. The Journal of Biological Chemistry **279**: 45728-45736

Kotake T, Hirosawa C, Ando Y, Tsumuraya Y (2010) Generation of nucleotide sugars for biomass formation in plants. Plant Biotechnology **27**: 231-236

Krishnan P, Kruger NJ, Ratcliffe RG (2004) Metabolite fingerprinting and profiling in plants using NMR. Journal of Experimental Botany **56**: 255-265

Kruger NJ, Troncoso-Ponce MA, Ratcliffe RG (2008) ^{1}H NMR metabolite fingerprinting and metabolic analysis of perchloric acid extracts from plant tissues. Nature Protocols **3**: 1001-1012

Kuo TM, VanMiddlesworth JF, Wolf WJ (1988) Content of raffinose oligosaccharides and sucrose in various plant seeds. Journal of Agricultural and Food Chemistry **36**: 32-36

Labate CA, Labate MTV, Bertolo L, Nascimento DD, Gutmanis G, Faraldo MIF (2008) Method for the genetic modulation of the biosinthesis of hemicellulose, cellulose and uronic acids in plant cells using gene expression cassettes. US Patent 20080277080 A1

Lakshmanan P, Geijskes RJ, Aitken KS, Grof CLP, Bonnett GD, Smith GR (2005) Sugarcane biotechnology: The challenges and opportunities. In Vitro Cellular and Developmental Biology **41**: 345-363

Lange C, Mittermayr L, Dohm JC, Holtgräwe D, Weishaar B, Himmelbauer H (2010) High-throughput identification of genetic markers using representational oligonucleotide microarray analysis. Theoretical and Applied Genetics **121**: 549-565

Lehner A, Chopera DR, Peters SW, Keller F, Mundree SG, Thomson JA, Farrant JM (2008) Protection mechanisms in the resurrection plant *Xerophyta viscosa*: cloning, expression, characterisation and role of *XvlNO1*, a gene coding for a *myo*-inositol 1-phosphate synthase. Functional Plant Biology **35**: 26-37

Lewis IA, Schommer SC, Hodis B, Robb KA, Tonelli M, Westler WM, Sussman MR, Markley JL (2007) Method for determining molar concentrations of metabolites in complex solutions from two-dimensional ^{1}H-^{13}C NMR spectra. Analytical Chemistry **79**: 9385-9390

Lewis IA, Schommer SC, Markley JL (2009) rNMR: open source software for identifying and quantifying metabolites in NMR spectra. Magnetic Resonance in Chemistry **47**: S123-S126

Li Y, Jones L, McQueen-Mason S (2003) Expansins and cell growth. Current Opinion in Plant Biology **6**: 603-610

Lin CY, Wu H, Tjeerdema RS, Viant MR (2007) Evaluation of metabolite extraction strategies from tissue samples using NMR metabolomics. Metabolomics **3**: 55-67

Li X, Weng JK, Chapple C (2008) Improvement of biomass through lignin modification. The Plant Journal **54**: 569–581

Li X, Bonawitz ND, Weng J, Chapple C (2010) The growth reduction associated with repressed lignin biosynthesis in *Arabidopsis thaliana* is independent of flavonoids. The Plant Cell **22**: 1620-1632

Li Q, Sun SSM, Yuan D, Yu H, Gu M, Liu Q (2010) Validation of candidate reference genes for the accurate normalization of Real-Time Quantitative RT-PCR data in rice during seed development. Plant Molecular Biology Reporter **28**: 49-57

Lingle SE (1989) Evidence for the Uptake of Sucrose Intact into Sugarcane Internodes. Plant Physiology **90**: 6-8

Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants. Nature Protocols **1**: 387-396

Loewus FA, Kelly S, Neufeld EF (1962) Metabolism of myo-inositol in plants: conversion to pectin, hemicelluloses, D-xylose, and sugar acids. Biochemistry **48**: 421-425

Loewus F (2005) Carbohydrate interconversions. Annual Review of Plant Physiology **22**: 337-364

Loewus FA, Loewus MW (1983) myo-Inositol: Its biosynthesis and metabolism. Annual Review of Plant Physiology **34**: 137-161

Loewus AL, Murthy PPN (2000) myo-Inositol metabolism in plants. Plant Science **150**: 1-19

Lommen A (2009) MetAlign: Interface-Driven, Versatile Metabolomics Tool for Hyphenated Full-Scan Mass Spectrometry Data Preprocessing. Analytical Chemistry **81**: 3079–3086

Ludwig C, Viant MR (2009) Two-dimensional J-resolved NMR spectroscopy: Review of a key methodology in the metabolomics toolbox. Phytochemical Analysis **21**: 21-32

Maretzki A, Thom M (1977) Characteristics of a galactose-adapted sugarcane cell line grown in suspension culture. Plant Physiology **61:** 544-548

Mayer AM, Staples RC (2002) Laccase: new functions for an old enzyme. Phytochemistry **60**: 551-565

McCormic AJ, Cramer MD, Watt DA (2008) Differential expression of genes in the leaves of sugarcane in response to sugar accumulation. Tropical Plant Biology. **1**: 142-158

McIntyre CL, Jackson M, Cordeiro GM, Amouyal O, Hermann S, Aitken KS, Eliott F, Henry RJ, Casu RE, Bonnett GD (2006) The identification and characterisation of alleles of sucrose phosphate synthase gene family III in sugarcane. Molecular Breeding **18**: 39-50

Meinzer FC, Moore PH (1988) Effect of Apoplastic Solutes on Water Potential in Elongating Sugarcane Leaves. Plant Physiology **86**: 873-879

Mevic B, Wehrens R (2007) The pls package: principal component and partial least squares regression in R. Journal of Statistical Software **18**: 1-18

Mhawech P (2005) 14-3-3 proteins-an update. Cell Research **15**: 228-236

Miller RG (1981) Simultaneous statistical inference. Springer-Verlang Inc., New York

Miller CJ (2005) simpleaffy: Very simple high level analysis of Affymetrix data. Bioinformatics **21**: 3683-3685

Moore BW, Perez VJ (1967) in Carlson FD, (Eds), Physiological and biochemical aspects of nervous integration. Prentice-Hall, Inc., Englewood Cliffs, N.J. (343-359)

Moore PH (1995) Temporal and spatial regulation of sucrose accumulation in the sugarcane stem. Australian Journal of Plant Physiology **22**: 6610679

Moore PH, Maretzki A (1996) Sugarcane. In Zamski E, Schaffer AA, (Eds), Photoassimilate Distribution in Plants and Crops: Source sink Relationships. (643-669) Marcel Dekker Inc., New York

Moorhead G, Douglas P, Cotelle V, Harthill J, Morrice N, Meek S, Deiting U, Stitt M, Scarabel M, Aitken A, MacKintosh C (1999) Phosphorylation-dependent interactions between enzymes of plant metabolism and 14-3-3 proteins. The Plant Journal **18**: 1-12

Moreau Y, Aerts S, De Moor B, De Strooper B, Dabrowski M (2003) Comparison and meta-analysis of microarray data: from the bench to the computer desk. TRENDS in Genetics **19**: 570-577

Naidoo S, Denby KJ, Berger DK (2005) Microarray experiments: considerations for experimental design. South African Journal of Science **101**: 347-354

Nakashima J, Chen F, Jackson L, Shadle G, Dixon RA (2008) Multi-site genetic modification of monolignol biosynthesis in alfalfa (*Medicago sativa*): effects on lignin composition in specific cell types. New Phytologist **179**: 738–750

Nielsen JS, Møller BL (2000) Cloning and expression of cytochrome P450 enzymes catalysing the conversion of tyrosine to *p*-hydroxyphenylacetaldoxime in the biosynthesis of cyanogenic glucosides in *Triglochin maritime*. Plant Physiology **122**: 1311-1321

Ness AS (2007) Microarray analysis: basic strategies for successful experiments. Molecular Biotechnology **36**: 205-219

Neufeld EF, Feingold DS, Hassid WZ (1959) Enzymic Phosphorylation of D-Glucuronic Acid by Extracts from Seedlings of *Phaseolus aureus*. Archives of Biochemistry and Biophysics **83**: 96-100

Nuss KJ, Brett PGC (1995) The release of variety NCo310 in 1945 and its impact on the sugar industry. Proceedings of the South African Sugar Technologists Association **69**: 3-8

Oka T, Jigami Y (2006) Reconstruction of *de novo* pathway for synthesis of UDP-glucuronic acid and UDP-xylose from intrinsic UDP-glucose in *Saccharomyces cerevisiae*. FEBS Journal **273**: 2645-2657

O'Neill MA, Eberhard S, Albershiem P, Darvill AG (2001) Requirement of borate cross-linking of cell wall rhamnogalacturonan II for *Arabidopsis* growth. SCIENCE **294**: 846-849

Osbourn AE (1996) Preformed antimicrobial compounds and plant defense against funga1 attack. The Plant Cell **8**: 1821-1831

Owen DF (1978) Why do aphids synthesize melezitose? Oikos **31**: 264-267

Panthee D, Cao C, Debenport S, Rodriguez G, Labate JA, Robertson LD, Breksa Iii AP, Van Der Knaap E, Mcspadden-Gardener B (2012) Magnitude of genotype x environment interactions affecting tomato fruit quality. HortScience **47**: 721-726

Papini-Terzi FS, Rocha FR, Vêncio RZN, Felix JM, Branco DS, Waclawovsky AJ, Del Bem LEV, Lembke CG, Costa MDL, Nishiyama Jr MY, Vicentini R, Vincentz GA, Ulian EC, Menossi M, Souza GM (2009) Sugarcane genes associated with sucrose content. BMC Genomics **10**: 120

Parman C, Halling C, Gentleman R (2005) affyQCReport: QC Report generation for affyBatch objects. R package version 1.17.0

Patrick JW (1997) Phloem unloading: Sieve element unloading and post-sieve element transport. Annu. Rev. Plant Physiology Plant Molecular Biology **48**: 191-222

Paul M, Pellny T, Goddijn O (2001) Enhancing photosynthesis with sugar signals. TRENDS in Plant Science **6**: 197-200

Paul M (2007) Trehalose 6-phosphate. Current Opinion in Plant Biology **10**: 303-309

Pauly M, Eberhard S, Albersheim P, Darvill A, York WS (2001) Effects of the *mur1* mutation on xyloglucans produced by suspension-cultured *Arabidopsis thaliana* cells. Planta **214**: 67-74

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

Pear JR, Kawagoe Y, Schreckengost WE, Delmer DP, Stalker DM (1996) Higher plants contain homologs of the bacterial *celA* genes encoding the catalytic subunit of cellulose synthase. PNAS **93**: 12637-12642

Pearson K (1901) On lines and planes of closest fit to systems of points in space. Philosophical Magazine **2**: 559-572

Pellny TK, Ghannoum O, Conroy JP, Schluepmann H, Smeekens S, Andralojc J, Krause KP, Goddijn O, Paul MJ (2004) Genetic modification of photosynthesis with *E.coli* genes for trehalose synthesis. Plant Biotechnology Journal **2**: 71-82

Peng X, Wood CL, Blalock EM, Chen KC, Landfield PW, Stromberg AJ (2002) Statistical implications of pooling RNA samples for microarray experiments. BMC Bioinformatics **4**: 26-44

Perneger TV (1998) What's wrong with Bonferroni adjustments. British Medical Journal **316**: 1236-1238

Peters S, Mundree SG, Thomson JA, Farrant JM, Keller F (2007) Protection mechanisms in the resurrection plant *Xerophyta viscosa* (Baker): both sucrose and raffinose family oligosaccharides (RFOs) accumulate in leaves in response to water deficit. Journal of Experimental Botany **58**: 1947-1956

Peters S, Keller F (2009) Frost tolerance in excised leaves of the common bugle (*Ajuga reptans* L.) correlates positively with the concentrations of raffinose family oligosaccharides (RFOs). Plant, Cell and Environment **32**: 1099-1107

Peters S, Egert A, Stieger B, Keller F (2010) Functional identification of Arabidopsis *ATSIP2* (At3g55720) as an alkaline galactosidase with a substrate specificity for

raffinose and an apparent sink-specific expression pattern. Plant and Cell Physiology **51**: 1815-1819

Pieslinger AM, Hoepflinger MC, Tenhaken R (2009) Nonradioactive enzyme measurement by high-performance liquid chromatography of partially purified sugar-1-kinase (glucuronokinase) from pollen of Lilium longiflorum. Analytical Biochemistry **388**: 254-2

Pieslinger AM, Hoepflinger MC, Tenhaken R (2010) Cloning of Glucuronokinase from Arabidopsis thaliana, the Last Missing Enzyme of the myo-Inositol Oxygenase Pathway to Nucleotide Sugars. The Journal of Biological Chemistry **285**: 2902-2910

Pilon-Smits EAH, Terry N, Sears T, Hyeong K, Zayed A, Seongbin H, van Dun K, Voogd E, Verwoerd TC, Krutwagen RWHH, Goddijn OJM (1998) Trehalose-producing transgenic tobacco plants show improved growth performance under drought stress. Journal of Plant Physiology **152**: 525-523

Price J, Laxmi A, Martin SK, Jang J (2004) Global transcription profiling reveals multiple sugar signal transduction mechanisms in Arabidopsis. The Plant Cell **16**: 2128-2150

R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

Rae AL, Grof CPL, Casu RE, Bonnett GD (2005) Sucrose accumulation in the sugarcane stem: pathways and control points for transport and compartmentation. Field Crops Research **92**: 159-168

Ramon M, Rolland F (2007) Plant development: introducing trehalose metabolism. TRENDS in Plant Science **12**:185-188

Reboul R, Geserick C, Pabst M, Frey B, Wittmann D, Lütz-Meindl U, Léonard R, Tenhaken R (2011) Down-regulation of UDP-glucuronic acid biosynthesis leads to swollen plant cell walls and severe developmental defects associated with changes in pectic polysaccharides. The Journal of Biological Chemistry **286**: 39982-39992

Reiter W, Chapple CCS, Somerville CR (1993) Altered growth and cell walls in a fucose-deficient mutant of *Arabidopsis*. SCIENCE **261**: 1032-1035

Reiter W, Chapple CCS, Somerville CR (1997) Mutants of *Arabidopsis thaliana* with altered cell wall polysaccharide composition. The Plant Journal **12**: 335-458

Reiter W, Vanzin GF (2001) Molecular genetics of nucleotide sugar interconversion pathways in plants. Plant molecular Biology **47**: 95-113

Reiter WD (2002) Biosynthesis and properties of the plant cell wall. Current Opinion in Plant Biology **5**: 536-542

Richmond TA, Somerville CR (2000) The Cellulose Synthase Superfamily. Plant Physiology **124**: 495-498

Robertson D, Smith C, Bolwell GP (1996) Inducible UDP-glucose dehydrogenase from French bean (*Phaseolus vulgaris* L.) locates to vascular tissue and has alcohol dehydrogenase activity. Biochemical Journal **313**: 311-317

Rockett JC, Hellmann GM (2004) Confirming microarray data-is it really necessary? Genomics **83**: 541-549

Roelofsen A (1958) Cell wall structure as related to surface growth. Acta botanica neerlandica **7**: 77-89

Roessner U, Wagner C, Kopka J, Trethewey RN, Willmitzer L (2000) Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. The Plant Journal **23**: 131–142

Roessner U, Willmitzer L, Fernie AR (2001) High-Resolution Metabolic Phenotyping of Genetically and Environmentally Diverse Potato Tuber Systems. Identification of Phenocopies. Plant Physiology **127**: 749–764

Rohde A, Morreel K, Ralph J, De Rycke R, Kushnir S, Van Doorsselaere J, Goeminne G, Joseleau JP, Vuylsteke M, Van Driessche G, Van Beeumen J, Messens E, Boerjan W (2004) Molecular phenotyping of the pal1 and pal2 mutants of Arabidopsis thaliana reveals far-reaching consequences on phenylpropanoid, amino acid, and carbohydrate metabolism. The Plant Cell **16**: 2749-2771

Rolland F, Baena-Gonzalez E, Sheen J (2006) Sugar Sensing and Signaling in Plants: Conserved and Novel Mechanisms. Annual Review of Plant Biology **57**: 675-709

Rossouw D, Kossmann J, Botha FC, Groenewald J (2010) Reduced neutral invertase activity in the culm tissues of transgenic sugarcane plants results in a decrease in respiration and sucrose cycling and an increase in the sucrose to hexose ratio. Functional Plant Biology **37:** 22–31

Rubin EM (2008) Genomics of cellulosic biofuels. Nature Reviews **454**: 841-845

Sachdeva M, Bhatia S, Batta SK (2011) Sucrose accumulation in sugarcane: a potential target for drop improvement. Acta Physiologiae Plantarum **33**: 1571-1583

Samac DA, Litterer L, Temple G, Jung HG, Somers DA (2004) Expression of UDP-glucose dehydrogenase reduces cell-wall polysaccharide concentration and increases xylose content in alfalfa stems. Applied Biochemistry and Biotechnology **11**: 1167-1182

Sampedro J, Cosgrove DJ (2005) The expansin superfamily. Genome Biology **6**: 242

Schäfer WE, Rohwer JM, Botha FC (2004) Protein expression and localization of sucrose synthase in the sugarcane culm. Physiologia Plantarum **121**: 187-195

Schauer N, Fernie AR (2006) Plant metabolomics: towards biological function and mechanism. TRENDS in Plant Science **11**: 508-516

Schijlen EGWM, de Vos RCH, van Tunen AJ, Bovy AG (2004) Modification of flavonoid biosynthesis in crop plants. Phytochemistry **65**: 2631-2648

Schilmiller AL, Stout J, Weng J, Humpherys J, Ruegger MO, Chapple C (2009) Mutations in the *cinnamate 4-hydroxylase* gene impact metabolism, growth and development in Arabidopsis. The Plant Journal **60**: 771-782

Schluepmann H, Pellny T, van Dijken A, Smeekens S, Paul M (2003) Trehalsoe 6-phosphate is indispensable for carbohydrate utilization and growth in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences **100**: 6849-6854

Schwald W, Chan M, Breuil C, Saddler JN (1988) Comparison of HPLC and colorimetric methods for measuring cellulolytic activity. Applied Microbiology and Biotechnology **28**: 398-403

Scott IM, Vermeer CP, Liakata M, Corol DI, Ward JL, Lin W, Johnson HE, Whitehead L, Kular B, Baker JM, Walsh S, Dave A, Larson TR, Graham IA, Wang TL, King RD, Draper J, Beale MH (2010) Enhancement of plant metabolite fingerprinting by machine learning. Plant Physiology **153**: 1506-1520

Seifert GJ (2004) Nucleotide sugar interconversions and cell wall biosynthesis: how to bring the inside to the outside. Current Opinion in Plant Biology **7**: 277-284

Seitz B, Klos C, Wurm M, Tenhaken R (2000) Matrix polysaccharide precursors in Arabidopsis cell walls are synthesized by alternate pathways with organ-specific expression patterns. The Plant Journal **21**: 537-546

Shih JH, Michalowska AM, Dobbin K, Ye Y, Qiu TH, Green JE (2004) Effects of pooling mRNA in microarray class comparisons. Bioinformatics **20**: 3318-3325

Sibout R, Eudes A, Mouille G, Pollet B, Lapierre C, Jouanin L, Se´guin 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**: 2059-2076

Sims REH, Hastings A, Schlamadinger B, Taylor G, Smith P (2006) Energy crops: current status and future prospects. Global Change Biology **12**: 2054-2076

Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology **3**: Article 3

Smyth GK (2005). Limma: linear models for microarray data. In: Bioinformatics and Computational Biology Solutions using R and Bioconductor. Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W (eds), Springer, New York, p 397-420

Snyman SJ, Baker C, Huckett BI, McFarlane SA, van Antwerpen T, Berry S, Omarjee J, Rutherford RS, Watt DA (2008) South African Sugarcane Research Institute: Embracing biotechnology for crop improvement research. Sugar Tech **10**: 1-13

Stewart DC, Copeland L (1998) Uridine 5'-diphosphate-glucose dehydrogenase from soybean nodules. Plant Physiology **116**: 349-355

Sticklen MB (2008) Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol. Nature Reviews **9**: 433-443

Streb S, Egli B, Eicke S, Zeeman SC (2009) The debate on the pathway of starch synthesis: a closer look at low-starch mutants lacking plastidial phosphoglucomutase supports the chloroplast-localized pathway. Plant Physiology **151**: 1769-1772

Sumner LW, Mendes P, Dixon RA (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochemistry **62**: 817–836

Sun JX, Sun XF, Sun RC, Su YQ (2004) Fractional extraction and structural characterization of sugarcane bagasse hemicelluloses. Carbohydrate Polymers **56**: 195-204

Suprasanna P, Patade VY, Desai NS, Devarumath RM, Kawar PG, Pagariya MC, Ganapathi A, Manickavasagam M, Babu KH (2011) Biotechnological developments in sugarcane improvement: an overview. Sugar Tech **13**: 322-335

Tang HR, Wang YL, Nicholson JK, Lindon JC (2004) Use of relaxation-edited one-dimensional and two dimensional nuclear magnetic resonance spectroscopy to improve detection of small metabolites in blood plasma. Analytical Biochemistry **325**: 260-272

Tattersall GN, Bak S, Jones PR, Olsen CE, Nielsen JK, Hansen ML, Høj BL (2001) Resistance to an herbivore through engineering cyanoganic glucoside synthesis. SCIENCE **293**: 1826-1828

Taylor NG (2008) Cellulose biosynthesis and deposition in higher plants. New Phytologist **178**: 239-252

Tikunov Y, Lommen A, de Vos CHR, Verhoeven HA, Bino RJ, Hall RD, Bovy AG (2005) A Novel Approach for Nontargeted Data Analysis for Metabolomics. Large-Scale Profiling of Tomato Fruit Volatiles. Plant Physiology **139**: 1125–1137

Tsao L, Dou K, Sun G, Lyu Y (2001) Synthesis of melezitose derivatives. Chemistry of Natural Compounds **37**: 397- 401

Turlapati PV, Kim K, Davin LB, Lewis NG (2011) The laccase multigene family in *Arabidopsis thaliana*: towards adressing the mystery of their gene function(s). Planta **233**: 439-470

Turner W, Botha FC (2002) Purification and kinetic properties of UDP-glucose dehydrogenase from sugarcane. Archives of Biochemistry and Biophysics **407**: 209–216

Unda F, Canam T, Preston L, Mansfield SD (2011) Isolation and characterization of galactinol synthases from hybrid poplar. Journal of Experimental Botany: doi:10.1093/jxb/err411, 1-11

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM (2007) Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Research **35**: W71-W74

van Bel AJE (2003) The phloem, a miracle of ingenuity. Plant, Cell and Environment **26**: 125-149

van den Berg RA, Hoefsloot HCJ, Westerhuis JA, Smilde AK, van der Werf M (2006) Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics **7**: 142-155

van der Merwe J (2005) The characterization of the UDP-glucose dehydrogenase gene and regulatory sequences of sugarcane. PhD Thesis, Stellenbosch University

van der Merwe J (2006) Isolation and evaluation of the sugarcane UDP-glucose dehydrogenase gene and promoter. PhD Thesis, Stellenbosch University

Vanholme R, Morreel K, Ralph J, Boerjan W (2008) Lignin engineering. Current Opinion in Plant Biology **11**: 278–285

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

Wang M, Goldstein C, Su W, Moore PH, Albert HH (2005) Production of biologically active GM-CSF in sugarcane: a secure biofactory. Transgenic Research **14**: 167-178

Ward JL, Baker JM, Beale MH (2006) Recent applications of NMR spectroscopy in plant metabolomics. FEBS Journal **274**: 1126-1131

Watt DA, McCormick AJ, Govender C, Carson DL, Cramer MD, Huckett BI, Botha FC (2005) Increasing the utility of genomics in unravelling sucrose accumulation. Field Crops Research **92**: 149-158

Wei H, Wang M, Moore PH, Albert HH (2003) Comparative expression analysis of two sugarcane polyubiquitin promoters and flanking sequences in transgenic plants. Journal of Plant Physiology **160**: 1241-1251

Weng J, Chapple C (2010) The origin and evolution of lignin biosynthesis. New Phytologist **187**: 273-285

Wettenhall JM, Simpson KM, Satterley K, Smyth GK (2006) affylmGUI: a graphical user interface for linear modeling of single channel microarray data. Bioinformatics **22**: 897-899

Whetten R, Sederoff R (1995) Lignin Biosynthesis. Plant Cell **7**: 1001-1013

White EJ, Venter M, Hiten NF, Burger JT (2008) Modified cetyltrimethylammonium bromide method improves robustness and versatility: the benchmark for plant RNA extraction. Biotechnology Journal **3**: 1424-1428

Whittaker A, Botha FC (1997) Carbon Partitioning during Sucrose Accumulation in Sugarcane lnternodal Tissue. Plant Physiology **115**: 1651-1 659

Wilson K, Walker J (2005) Principles and Techniques of Biochemistry and Molecular Biology (6th   edition) Cambridge University Press, Cambridge, New York

Wishart DS (2008) Quantitative metabolomics using NMR. Trends in Analytical Chemistry. **27**: 228-237

Welbaum GE, Meinzer FC (1990) Compartmentation of solutes and water in developing sugarcane stalk tissue. Plant Physiology **93**: 1147-1153

Wold H (1975) Path models with latent variables: The NIPALS approach. In Blalock HM, Aganbegian A, Borodkin FM, Boudon R, Capecchi V (Eds.) Quantitative sociology: International perspectives on mathematical and statistical modeling (307-357) New York: Academic.

Wold S, Martens H, Wold H (1983) The multivariate calibration problem in chemistry solved by the PLS method. Lecture Notes in Mathematics **973**: 286-293

Yamamoto R, Inouhe M, Masuda Y (1988) Galactose inhibition of auxin-induced growth of mono- and dicotyledonous plants. Plant Physiology **86**: 1223-1227

Yang Q, Trinh HX, Imai S, Ishihara A, Zhang L, Nakayashiki H, Tosa Y, Mayama S (2003) Analysis of the involvement of hydroxyanthranilate hydroxycinnamoyltransferase and

caffeoyl-CoA 3-O-methyltransferase in phytoalexin biosynthesis in oat. Molecular Plant-Microbe Interactions **17**: 81-89

Zhang S, Gant TW (2005) Effect of pooling samples on the efficiency of comparative studies using microarrays. Bioinformatics **21**: 4378-4383

Zhou R, Sicher RC, Quebedeaux B (2002) Apple leaf sucrose-phosphate synthase is inhibited by sorbitol-6-phosphate. Functional Plant Biology **29**: 569-574

Zhu YJ, Komor E, Moore PH (1997) Sucrose accumulation in the sugarcane stem is regulated by the difference between the activities of soluble acid invertase and sucrose phosphate synthase. Plant Physiology **115**: 609-616

Xia Z, Akim LG, Argyropoulos DS (2001) Quantitative ^{13}C NMR analysis of lignins with internal standards. Journal of Agricultural and Food Chemistry. **49**: 3573-3578

Xia J, Psychogios N, Young N, Wishart DS (2009) MetaboAnalyst: a web server for metabolomic data analysis and interpretation. Nucleic Acids Research **37**: W652-W660

Xia J, Wishart DS (2011) Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. Nature Protocols **6**: 743-60

Xu P, Chen F, Mannas JP, Feldman T, Sumner LW, Roossinck (2008) Virus infection improves drought tolerance. New Phytologist *doi*: 10.1111/j.1469-8137.2008.02627.x

Appendix

Table A.1: Probe sets found to be significantly differentially expressed in UGD 7 and UGD 18 when compared to the untransformed NCo310 line. Fold change values are in $\log_2$ space. A p-value threshold of $p < 0.05$ was applied. Only probe sets for which annotation was available at the time of writing and which had a fold change higher or lower than 1.5 are displayed.

UGD 7

Probe set	FC	Annotation
Sof.965.1.S1_at	5.76	Lipid transfer protein
Sof.3563.1.S1_at	4.47	Peroxidase
Sof.4551.1.S1_a_at	4.17	Nucleotide metabolism
Sof.4551.1.S1_at	4.12	Nucleotide metabolism
Sof.1984.1.S1_at	3.86	beta-glucosidase isozyme 2 precursor
Sof.5334.1.A1_a_at	3.13	Histone
Sof.5232.2.S1_x_at	2.85	Histone
Sof.1705.1.S1_s_at	2.82	Serine protease
Sof.5232.1.S1_a_at	2.78	Histone
Sof.1072.1.A1_at	2.18	Fbox protein
Sof.83.1.S1_at	2.11	Cytochrome P450
Sof.4111.1.S1_at	2.07	Histone
Sof.4025.2.S1_at	2.02	Histone
Sof.4025.2.S1_x_at	2.01	Histone
Sof.3597.1.S1_at	1.94	Lipid transfer protein
Sof.3427.1.S1_at	1.93	MIPS
Sof.4028.1.S1_a_at	1.88	Histone
Sof.3576.1.S1_x_at	1.80	Asparagine synthetase
Sof.3576.1.S1_at	1.78	Asparagine synthetase
Sof.5334.2.A1_x_at	1.75	Histone
Sof.4078.1.S1_at	1.71	Lipid transfer protein
Sof.721.1.A1_at	1.67	Major intrinsic protein
SofAffx.699.1.S1_at	1.66	MIPS
Sof.4541.1.S1_at	1.63	Cell organisation
Sof.1003.1.S1_at	1.63	Cytochrome P450
SofAffx.699.1.S1_s_at	1.59	MIPS
Sof.2403.1.A1_a_at	1.57	Transport
Sof.122.1.S1_at	1.54	HCT
Sof.3869.1.S1_at	1.51	Nucleotide metabolism
Sof.410.1.A1_at	1.51	Asparagine synthetase
Sof.1382.1.S1_at	-1.51	Psudo ARR transcription factor family
Sof.79.2.A1_at	-1.51	Protein phosphatase
Sof.2799.2.S1_at	-1.54	Protein postranslational modification
Sof.2571.1.S1_at	-1.58	calcium-dependent protein
Sof.2106.3.S1_a_at	-1.58	Lysine-specific histone demethylase

Sof.583.1.S1_at	-1.59	IAA14 - auxin-responsive aux iaa family member
Sof.2023.2.S1_a_at	-1.61	GTP cyclohydrolase II
Sof.1768.2.S1_x_at	-1.61	MAP kinase
Sof.79.1.S1_at	-1.62	Protein phosphatase
Sof.1768.3.S1_x_at	-1.64	MAP kinase
Sof.79.1.S1_a_at	-1.67	Protein phosphatase
Sof.2161.1.S1_at	-1.68	Protein phosphatase
Sof.1173.1.S1_at	-1.69	Alanine-glyoxylate aminotransferase
Sof.3562.2.S1_at	-1.72	bZIP transcription factor family
Sof.560.1.A1_at	-1.75	Metalloendoproteinase
Sof.2752.1.A1_at	-1.77	Kinase
Sof.1049.1.A1_at	-1.78	Protein degradation
SofAffx.548.1.S1_at	-1.81	Cytokinin oxidase
Sof.1768.3.S1_a_at	-1.81	MAP kinase
SofAffx.542.1.S1_at	-1.87	DIN10
SofAffx.1338.1.S1_at	-1.89	Receptor protein kinase
Sof.2971.1.S1_at	-1.98	Mitochondrial substrate carrier family protein
Sof.1007.1.A1_at	-1.99	Protein postranslational modification
Sof.116.1.S1_at	-2.14	Hyoscyamine 6-dioxygenase
Sof.1768.1.S1_at	-2.16	MAP kinase
Sof.3653.1.S1_at	-2.20	Trehalose-6-phosphate phosphatase
Sof.3531.1.S1_at	-2.21	Calcium depedant signalling
Sof.196.2.S1_at	-2.22	Calcium depedant signalling
Sof.5118.1.S1_at	-2.35	Cellulose synthase
Sof.161.1.S1_at	-2.38	Glucose-6-phosphate phosphate translocator
Sof.2106.1.A1_at	-2.38	Oxidase
Sof.3931.2.A1_s_at	-2.39	Glucose-6-phosphate phosphate-translocator
Sof.569.2.S1_a_at	-3.03	Aspartate protease
Sof.1325.1.S1_at	-3.07	Ethylene-responsive element binding protein family
Sof.3105.2.A1_at	-3.99	Trehalose-6-phosphate phosphatase

UGD 18

Probe set	FC	Annotation
Sof.3264.1.S1_at	6.31	Thaumatin-Like Protein
Sof.3315.1.S1_at	5.73	Stress related
Sof.1454.1.S1_at	4.84	Stress related
Sof.1713.1.A1_at	4.54	Cytochrome P450
Sof.1462.1.S1_at	4.21	Thaumatin-Like Protein
Sof.3550.1.S1_at	3.97	Beta Glucanase
Sof.478.1.A1_at	3.69	Xylanase Inhibitor
Sof.3101.1.S1_at	3.66	Chalcone Synthase
Sof.1072.1.A1_at	3.59	Fbox
Sof.3338.1.S1_at	3.51	Cell Wall Modification
Sof.1705.1.S1_s_at	3.48	Serine Carboxypeptidase Ii

Sof.1705.1.S1_at	3.47	Serine Carboxypeptidase Precursor
Sof.504.1.S1_at	3.25	Stress related
Sof.4551.1.S1_at	3.15	Nucleotide Metabolism
Sof.4551.1.S1_a_at	3.09	Nucleotide Metabolism.Degradation
Sof.1727.1.S1_s_at	2.98	Chitinase
Sof.3597.1.S1_at	2.94	Lipid Transfer Protein (Ltp)
Sof.2770.2.S1_at	2.92	Fbox
Sof.2770.2.S1_x_at	2.88	Fbox
Sof.4799.1.S1_at	2.75	Beta 1,3 Glucan Hydrolase
Sof.1408.1.S1_at	2.74	Chitinase
Sof.5334.1.A1_a_at	2.67	Histone
Sof.1218.1.S1_at	2.65	Mee14 (Maternal Effect Embryo Arrest 14)
Sof.2770.1.S1_at	2.61	Fbox
SofAffx.40.2.S1_at	2.60	Stach degradation
Sof.83.1.S1_at	2.51	Cytochrome P450
Sof.1984.1.S1_at	2.49	Beta-Glucosidase Isozyme 2 Precursor
Sof.4258.1.S1_x_at	2.38	Xylanase Inhibitor
Sof.122.1.S1_at	2.36	HCT
Sof.3289.1.S1_at	2.35	Stress related
Sof.478.2.S1_at	2.24	Stress related
Sof.2060.2.S1_at	2.14	Hydroxymethylglutaryl- Lyase
Sof.1710.1.S1_at	2.12	Postranslational Modification
Sof.2962.1.S1_a_at	2.01	Protein degradation
Sof.3576.1.S1_at	2.01	Asparagine Synthetase
SofAffx.1973.1.S1_at	1.94	Invertase Cell Wall4
Sof.3576.1.S1_x_at	1.93	Asparagine Synthetase
Sof.2965.1.S1_at	1.90	Protochlorophyllide Reductase
Sof.4111.1.S1_at	1.88	Histone
Sof.724.1.A1_at	1.87	Amino Acid Metabolism
Sof.4240.1.S1_at	1.85	Stress related
Sof.2157.1.A1_s_at	1.79	Serine Threonine Protein Kinase
Sof.48.1.S1_at	1.76	Receptor Like Cytoplasmatic Kinase Vii Magnesium-Protoporphyrin Ix Monomethyl Ester (Oxidative)
Sof.1537.2.S1_at	1.73	Cyclase
Sof.411.1.A1_at	1.72	Serine Carboxypeptidase Precursor
Sof.823.1.A1_at	1.72	Leucoanthocyanidin Reductase Magnesium-Protoporphyrin Ix Monomethyl Ester (Oxidative)
Sof.1537.1.S1_at	1.71	Cyclase
Sof.1003.1.S1_at	1.68	Cytochrome P450
Sof.2338.1.A1_at	1.64	Methylcrotonoyl- Carboxylase Subunit Alpha
Sof.1713.2.S1_at	1.64	Cytochrome P450
Sof.4252.2.S1_a_at	1.63	Alanine Aminotransferase
Sof.2899.2.S1_at	1.63	Major Intrinsic Protein
Sof.2060.1.A1_at	1.62	Hydroxymethylglutaryl- Lyase
Sof.5334.2.A1_x_at	1.62	Histone
Sof.3217.1.S1_a_at	1.59	Trehalose-6-Phosphate Synthase

Sof.4410.1.S1_at	1.58	Cysteine Synthase
Sof.1865.1.A1_at	1.57	Mitochondrial Deoxynucleotide
Sof.4236.1.S1_at	1.54	Dismutases And Catalases
Sof.3267.1.S1_at	1.52	Glutathione S Transferases
Sof.572.1.S1_at	1.51	Leucoanthocyanidin Reductase
Sof.1028.1.S1_at	-1.50	Oxidase
Sof.970.2.S1_at	-1.51	Rhamnose synthesis
Sof.4673.1.S1_a_at	-1.53	Amino Acid Metabolism
Sof.258.1.S1_at	-1.54	Receptor Like Cytoplasmatic Kinase Vii
SofAffx.409.1.S1_at	-1.54	HCT
Sof.3922.3.S1_a_at	-1.54	Calcium dependant signalling
Sof.1122.2.S1_a_at	-1.54	Caffeoyl- O-Methyltransferase 1
SofAffx.1338.1.S1_at	-1.54	Receptor Kinase
Sof.1082.1.S1_at	-1.55	Peptide and Oligopeptide transporter
Sof.2881.1.S1_at	-1.55	Protein-Tyrosine Phosphatase
Sof.3882.1.A1_at	-1.56	Fbox
Sof.3531.1.S1_at	-1.57	Calcium dependant signalling
Sof.3722.2.S1_at	-1.57	Calcium dependant signalling
Sof.1478.1.A1_at	-1.58	P- And V-Atpases
Sof.863.1.S1_at	-1.59	Major Intrinsic Protein
Sof.5198.1.S1_a_at	-1.60	CCOAOMT
Sof.2161.1.S1_at	-1.63	Protein Phosphatase
Sof.3286.1.S1_s_at	-1.66	MYB-Related Transcription Factor
Sof.3444.2.S1_a_at	-1.68	Alpha-L-Fucosidase
Sof.3905.1.S1_at	-1.68	Chalcone synthesis
Sof.5198.2.S1_a_at	-1.69	Caffeoyl- 3-O-Methyltransferase
Sof.2571.1.S1_at	-1.72	Calcium dependant signalling
Sof.5220.1.S1_at	-1.73	Ethylene-Responsive Transcription Factor
Sof.583.1.S1_at	-1.74	AUX/IAA transcriptional regulation
Sof.1007.1.A1_at	-1.74	Postranslational Modification
Sof.3822.1.S1_at	-1.77	Cellulose Synthase
Sof.2878.1.S1_at	-1.78	ACP Thioesterase
Sof.2854.1.S1_at	-1.80	Alpha-Glucan-Protein Synthase
Sof.970.1.S1_a_at	-1.80	Rhamnose Biosynthetic Enzyme Expressed
Sof.3783.1.S1_at	-1.80	Triacylglycerol Lipase
Sof.3562.2.S1_at	-1.81	bZIP Transcription Factor
Sof.116.1.S1_at	-1.83	Hyoscyamine 6-Dioxygenase
Sof.210.1.S1_at	-1.91	MAP Kinase
Sof.3569.3.S1_a_at	-1.93	Xyloglucan Endotransglycosylase Hydrolase
Sof.196.2.S1_at	-1.94	Calcium dependant signalling
Sof.2657.1.S1_at	-1.96	Leaf Senescence Related
Sof.1261.1.S1_at	-1.99	Stress related
Sof.5235.1.A1_a_at	-2.03	Xyloglucan Endotransglycosylase
Sof.512.1.S1_at	-2.04	Pyrophosphate-Fructose-6-P Phosphotransferase
Sof.2007.1.S1_at	-2.05	Potassium transporter
Sof.2971.1.S1_at	-2.20	Metabolite Transporter

Sof.4682.1.S1_at	-2.23	Phenol metabolism
Sof.2084.1.S1_at	-2.27	Cytochrome P450 Reductase
Sof.5118.1.S1_at	-2.43	Cellulose Synthase
Sof.1050.1.S1_at	-2.51	Oxidase
Sof.702.1.S1_at	-2.53	Chalcone Synthase
Sof.5217.1.S1_at	-2.71	Xyloglucan Endotransglycosylase
Sof.1126.1.A1_at	-2.76	Embryogenesis Transmembrane
Sof.513.1.S1_at	-2.76	Lipid Transfer Protein (Ltp)
Sof.1126.2.S1_at	-2.91	Embryogenesis Transmembrane
Sof.3105.2.A1_at	-2.91	Trehalose-6-Phosphate Phosphatase
Sof.1325.1.S1_at	-3.11	Ethylene-Responsive Element Binding Protein
Sof.3931.2.A1_s_at	-3.71	Glucose-6-Phosphate Phosphate-Translocator
Sof.161.1.S1_at	-3.83	Glucose-6-Phosphate Phosphate Translocator 2
Sof.3875.1.S1_s_at	-4.18	Lipid Transfer Protein
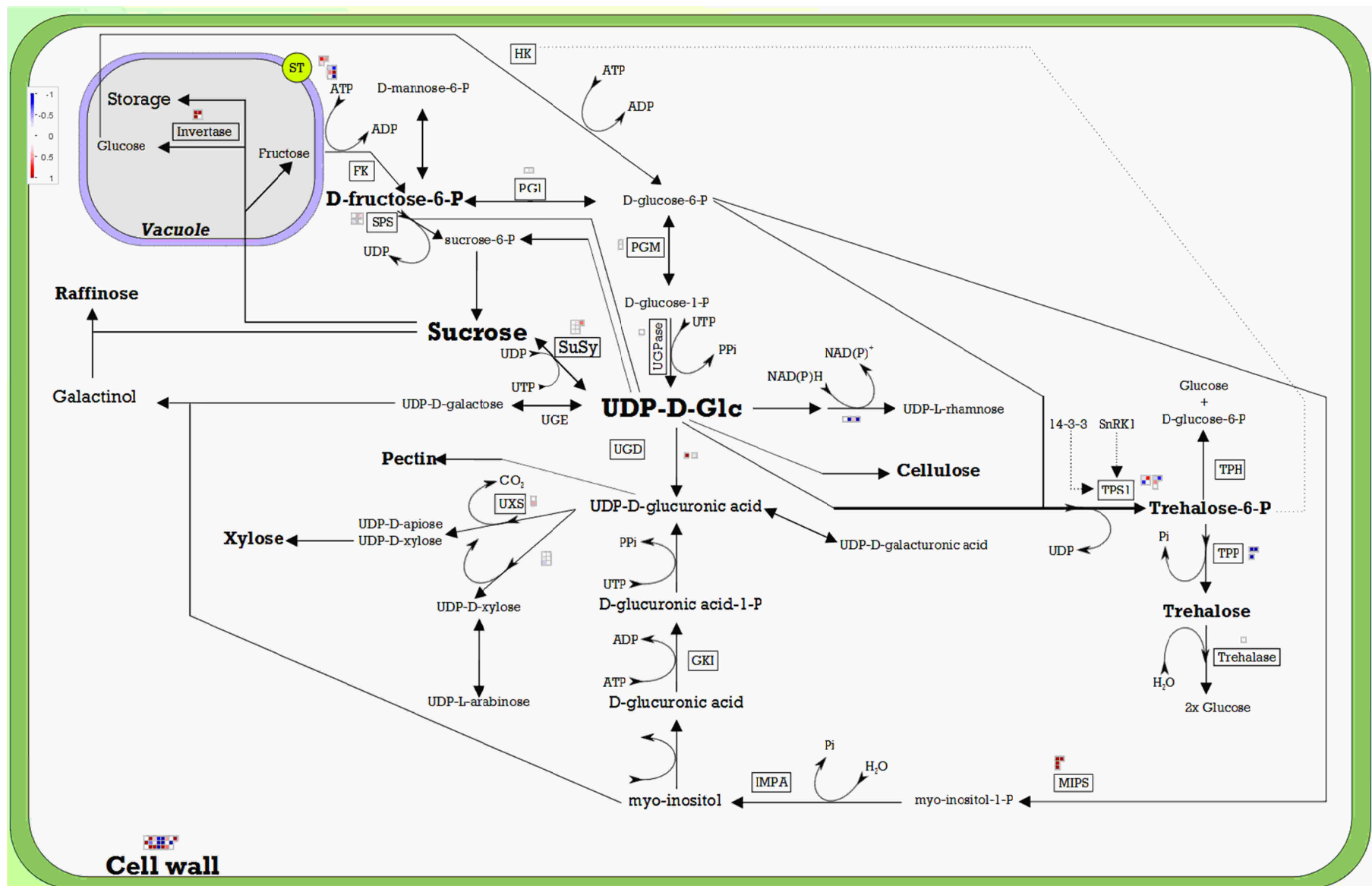
Figure A1: UGD 7 Fold change values of transgenic vs untransformed NCo310 plants displayed graphically via MapMan figures. Blue indicates a transcript is down regulated while red indicates it is up-regulated. White indicates no or very little difference in transcript abundance between the WT and the transgenic line under investigation. The intensity of the colour is a measure of the transcript abundance. A colour scale representing transcript abundance is present at the top left of each figure
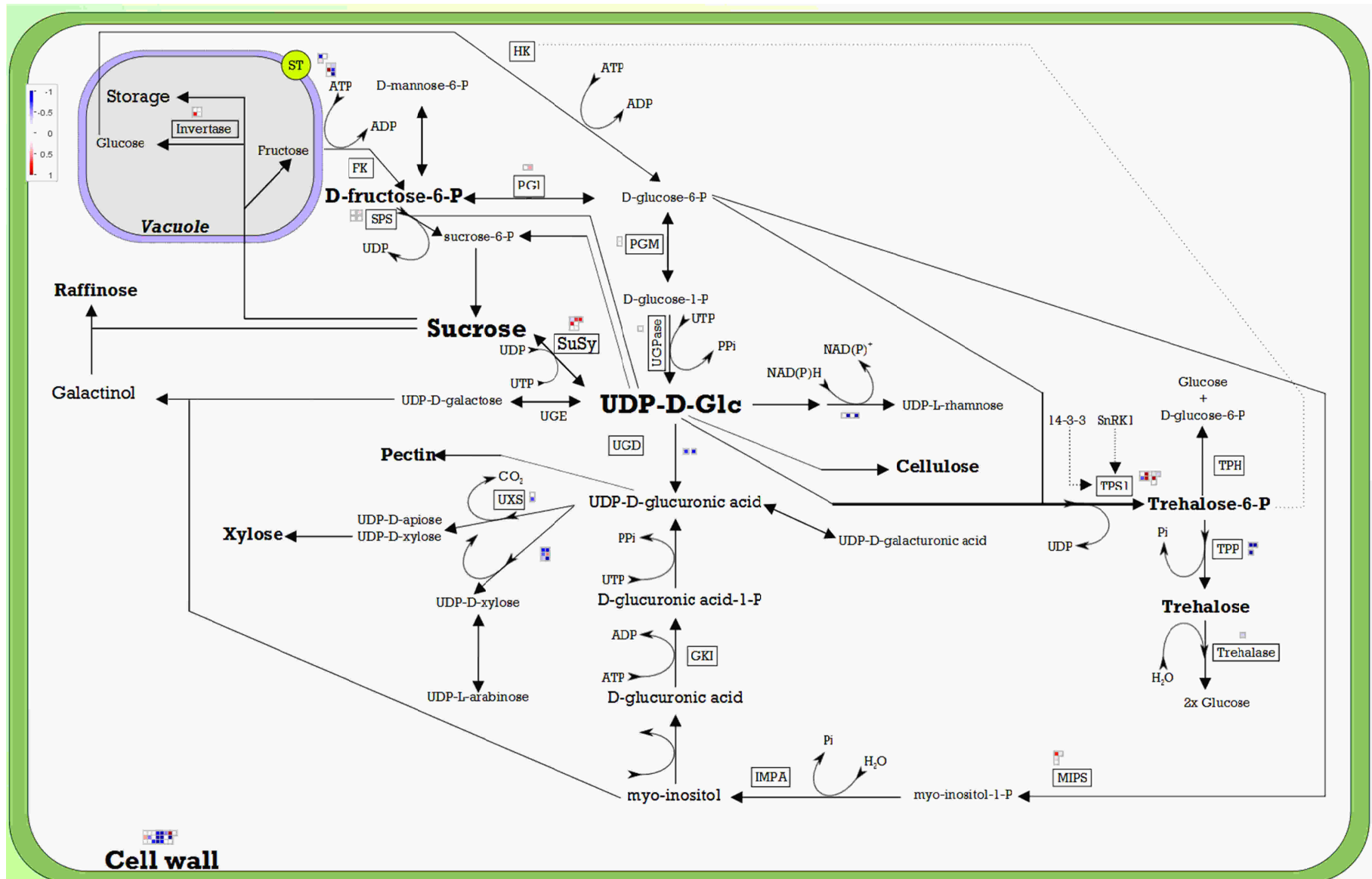
Figure A2: UGD 18 Fold change values of transgenic vs untransformed NCo310 plants displayed graphically via MapMan figures. Blue indicates a transcript is down regulated while red indicates it is up-regulated. White indicates no or very little difference in transcript abundance between the WT and the transgenic line under investigation. The intensity of the colour is a measure of the transcript abundance. A colour scale representing transcript abundance is present at the top left of each figure

```
Alignment: DIN10 sequences


           ....|....| ....|....| ....|....| ....|....| ....|....|
                10         20         30         40         50
XP_0036179 MTVGAGISVG DGNLMVLGKK VLSQVHENVL VTPASGGSLI NG-AFIGVSS
NP_0011057 MTVGAGIAVQ DGSLLALGAK VLREVRGNVL VTPAAGGGLT NG-AFLGVRS
NP_0011057 MTVTPRITVS DGRLTVRGRT VLTGVPDNVS AAHAAGAGLV DG-AFVGAHA
NP_0011475 MTVTPRITVS DGRLTVRGRT VLTGVPDNVS AAHAAGAGLF DG-AFVGAHA
AAM75140.1 MTVTPKISVN DGNLVVHGKT ILTGVPDNIV LTPGSGLGLV AG-AFIGATA
ACG27589.1 MTVTPRITVS DGRLTVRGRT VLTGVPDNVS AAHAAGAGLV DG-AFVGAHA
AAA32975.1 MTVTPQITVG DGRLAVRGRT VLSGVPDNVT AAHAAGAGLV DG-AFVGATA
BAI66422.1 MTVTPQITVS DGRLAVRGRT VLSGVPDNVT AAHASGAGLV DG-AFVGATA
AAT77910.1 MTVTPRITVA EGRLVAHGRT ILTGVADNIA LTHASGAGLV DG-AFVGATA
BAH19983.1 MTIKPAVRIS DGNLIIKNRT ILTGVPDNVI TTSASEAGPV EG-VFVGAVF
XP_0028781 MTITSNISVQ NDNLVVQGKT ILTKIPDNII LTPVTGNGFV SG-AFIGATF


           ....|....| ....|....| ....|....| ....|....| ....|....|
                60         70         80         90        100
XP_0036179 DQKGSRRVFP IGKLEELRFM SLFRFKMWWM TQRMGNCGQE IPFETQFLLI
NP_0011057 APAASRSIFP VGKLRDQRFV CTFRFKMWWM TQRMGSASRD IPSETQFLLV
NP_0011057 GEAKSHHVFT FGTLRDCRFL CLFRFKLWWM TQRMGVSGRD VPLETQFMLV
NP_0011475 GEAKSHHVFT FGTLRDCRFL CLFRFKLWWM TQRMGVSGRD VPLETQFMLV
AAM75140.1 SNSKSLHVFP VGVLEGTRFL CCFRFKLWWM TQRMGTSGRD IPFETQFLLM
ACG27589.1 GEAKSHHVFT FGTLRDCRFL CLFRFKLWWM TQRMGVSGRD VPLETQFMLV
AAA32975.1 AEAKSHHVFT FGTLRDCRFM CLFRFKLWWM TQRMGTSGRD VPLETQFILI
BAI66422.1 GEAKSHHVFT FGTLRDCRFM CLFRFKLWWM TQRMGTSGRD VPLETQFILI
AAT77910.1 DEPKSLHVFT FGTLRDLRFM CCFRFKLWWM TQRMGTSGRD VPLETQFMLL
BAH19983.1 NKEESKHIVP IGTLRNSRFM SCFRFKLWWM AQRMGEMGRD IPYETQFLLV
XP_0028781 EQSKSLHVFP VGVLEGLRFM CCFRFKLWWM TQRMGSCGKD IPLETQFMLL


           ....|....| ....|....| ....|....| ....|....| ....|....|
                110        120        130        140        150
XP_0036179 EAHKGCDIEG GIDNGEQDQD ---------- -GSTYAVLLP LLEGDFRAVL
NP_0011057 ERSGG----- ----GEQ--- ---------- -PVVYTVFLP VLEGSFRAVL
NP_0011057 EVPASDG--- ---DGDD--- ---------- -APAYVVMLP LLEGQFRAAL
NP_0011475 EVPASDG--- ---DGDD--- ---------- -APAYVVMLP LLEGQFRAAL
AAM75140.1 ESKGNDGE-- ---DPDNS-- ---------- -STIYTVFLP LLEGQFRAAL
ACG27589.1 EVPASDG--- ---DGDD--- ---------- -APAYVVMLP LLEGQFRAAL
AAA32975.1 EVPAAAGN-- --DDGDSS-- DGDS------ -EPVYLVMLP LLEGQFRTVL
BAI66422.1 EVPAAAGN-- --DDGDS--- ---------- -EPVYLVMLP LLEGQFRTVL
AAT77910.1 ESRDGGG--- ---GGE---- ---------- --AVYVVMLP LLEGQFRAAL
BAH19983.1 ESNDGSHLES DGANGVECN- ---------- -QKVYTVFLP LIEGSFRSCL
XP_0028781 ESKDEVEG-- ---NGDDA-- ---------- -PTIYTVFLP LLEGQFRAVL


           ....|....| ....|....| ....|....| ....|....| ....|....|
                160        170        180        190        200
XP_0036179 QGNDQ-NEIE ICVESGCPDV EEFDGTHLVF IGAGSDPYKV ITNAVKTVEK
NP_0011057 QGNAA-DELE ICLESGDPDV ESFQGSHLVF VGAGSDPFEV ITSSVKAVER
NP_0011057 QGNDR-DELQ ICIESGDKAV QTDQAAHMVY LHAGDNPFDT VTAAVKAVEK
NP_0011475 QGNDR-DELQ ICIESGDKAV QTDQAAHMVY LHAGDNPFDT VTAAVKAVEK
AAM75140.1 QGNEK-NEME ICLESGDNTV ETNQGLSLVY MHAGTNPFEV ITQAVKAVEK
ACG27589.1 QGNDR-DELQ ICIESGDKAV QTDQAAHMVY LHAGDNPFDT VTAAVKAVEK
AAA32975.1 QGNDQ-DELQ ICIESGDKAV ETEQGMNNVY VHAGTNPFDT ITQAVKAVEK
BAI66422.1 QGNDQ-DQLH ICIESGDKAV QTEQGMNSLY IHAGTNPFDT ITQAVKAVEK
AAT77910.1 QGNDR-DELE ICIESGDKAV QTAQGTYMVY VHAGANPFDT ITQAVKVVER
BAH19983.1 QGNVN-DEVE LCLESGDVDT KRSSFTHSLY IHAGTDPFQT ITDAIRTVKL
```

```
XP_0028781    QGNEK-NEIE ICLESGDKAV ETSQGTHLVY VHAGTNPFEV IRQSVKAVER


              ....|....| ....|....| ....|....| ....|....| ....|....|
                    210        220        230        240        250
XP_0036179    HLKTFCHRER KKMPDMLNWF GWCTWDAFYT NVTSENVKEG LQSFEEGGIP
NP_0011057    HLQTFSHREK KKMPDILNWF GWCTWDAFYT NVNAQGVKQG LQSLEKGGVS
NP_0011057    HLQTFHHRDK KKLPSFLDWF GWCTWDAFYT DVTADGVKNG LQSLSKGGAP
NP_0011475    HLQTFHHRDK KKLPSFLDWF GWCTWDAFYT DVTADGVKHG LQSLSKGGAP
AAM75140.1    HTQTFLHREK KKLPSFLDWF GWCTWDAFYT DATAEGVVEG LKSLSEGGAP
ACG27589.1    HLQTFHHRDK KKLPSFLDWF GWCTWDAFYT DVTADGVKNG LQSLSKGGAP
AAA32975.1    HTQTFHHREK KTVPSFVDWF GWCTWDAFYT DVTADGVKQG LRSLAEGGAP
BAI66422.1    HMQTFHHREK KKLPSFVDWF GWCTWDAFYT DVTADGVKQG LRSLAEGGVP
AAT77910.1    HLQTFHHREK KKLPSFLDWF GWCTWDAFYT DVTADGVKQG LQSLAEGGTP
BAH19983.1    HLNSFRQRHE KKLPGIVDYF GWCTWDAFYQ EVTQEGVEAG LKSLAAGGTP
XP_0028781    HMQTFHHREK KKLPSFLDWF GWCTWDAFYT DVTAEGVDEG LKSLSGGGTP


              ....|....| ....|....| ....|....| ....|....| ....|....|
                    260        270        280        290        300
XP_0036179    AKFVIIDDGW QSVSMDPNGV E--WKHDCAA NFANRLTHIK ENHKFQKDGK
NP_0011057    PRFVIIDDGW QSVAMDPVGI A--CLSDNSA NFANRLTHIR ENHKFQKNGR
NP_0011057    PRFLIIDDGW QQIASENKPD P-NVAVQEGA QFASRLTGIK ENTKFQTKPD
NP_0011475    PRFLIIDDGW QQIASENKPD P-NVAVQEGA QFASRLTGIK ENTKFQTKPD
AAM75140.1    PKFLIIDDGW QQIEAKPKDA D--CVVQEGA QFASRLSGIK ENHKFQKNGN
ACG27589.1    PRFLIIDDGW QQIASENKPD P-NVAVQEGA QFASRLTGIK ENTKFQTKPD
AAA32975.1    PRFLIIDDGW QQIGSENKDD P-GVAVQEGA QFASRLTGIR ENTKFQSEHN
BAI66422.1    PRFLIIDDGW QQIGSENKED P-GVAVQEGA QFASRLTGIK ENTKFQSEHD
AAT77910.1    PRFLIIDDGW QQIGSENKED AGNAVVQEGA QFASRLIGIK ENTKFQKTTT
BAH19983.1    PKFVIIDDGW QSVERDATVE A---GDEKKE SPIFRLTGIK ENEKFKKKDD
XP_0028781    PKFLIIDDGW QQIENKEKDE N--CVVQEGA QFATRLVGIK ENAKFQKSDQ


              ....|....| ....|....| ....|....| ....|....| ....|....|
                    310        320        330        340        350
XP_0036179    EGQRI----E DPAMGLHHIT NEIKKEHAIK HVYVWHAITG YWGGVKPGI-
NP_0011057    EGHRE----D DPAKGLAHVV NEIKGKHQLK YVYVWHAITG YWGGVRPGA-
NP_0011057    GDGDG----E QAAGGLKRLV AETKDAHGVK QVYVWHAMAG YWGGVTPTAG
NP_0011475    GDGDG----E QAPGGLKRLV AETKDAHGVK QVYVWHAMAG YWGGVTPTAG
AAM75140.1    --------NY DQVPGLKVVV DDAKKQHKVK FVYAWHALAG YWGGVKPAS-
ACG27589.1    GDGDG----E QAAGGLKRLV AETKDAHGVK QVYVWHAMAG YWSGVTPTAG
AAA32975.1    ---------Q EETPGLKRLV DETKKEHGVK SVYVWHAMAG YWGGVKPSA-
BAI66422.1    ---------Q DDTPGLKRLV EETKKGHGVK SVYVWHAMAG YWGGVKPSA-
AAT77910.1    TAMADGGETA ASAAGLKALV EEAKKEHGVK YVYVWHAMAG YWGGVKPAA-
BAH19983.1    ---------- -PNVGIKNIV KIAKEKHGLR YVYVWHAITG YWGGVRPGE-
XP_0028781    --------KD TQVSGLKSVV DNAKQRHNVK QVYAWHALAG YWGGVKPAA-


              ....|....| ....|....| ....|....| ....|....| ....|....|
                    360        370        380        390        400
XP_0036179    SGMEHYESKM AFPISSPGVK SNQPDEALDT IAINGLGLVN PEKVFHFYDE
NP_0011057    AGMEHYGSKM QRPVPSPGVQ KNERCDALDS MTANGLGLVN PDRAFSFYDE
NP_0011057    TAMERYEPAL AYPVQSPGVT GNQPDIVMDS LSVLGLGLVH PRRVRDFYGE
NP_0011475    TAMERYEPAL AYPVQSPGVT GNQPDIVMDS LSVLGLGLVH PRRVRDFYGE
AAM75140.1    PGMEHYDSAL AYPVQSPGML GNQPDIVVDS LAVHGIGLVH PKKVFNFYNE
ACG27589.1    TAMERYEPAL AYPVQSPGVT GNQPDIVMDS LSVLGLGLVH PRRVRDFYGE
AAA32975.1    AGMEHYEPAL AYPVQSPGVT GNQPDIVMDS LSVLGLGLVH PRRVHRFYDE
BAI66422.1    AGMEHYESAL AYPVQSPGVT GNQPDIVMDS LSVLGLGLVH PRKVYNFYDE
AAT77910.1    EGMEHYESAV AFPVQSPGVM GNQPDIVMDS LSVLGLGLVH PRMALAFYGE
BAH19983.1    ----EYGSVM KYPNMSKGVV ENDPTWKTDV MTLQGLGLVS PKKVYKFYNE
XP_0028781    SGMEHYDSAL AYPVQSPGVL GNQPDIVMDS LAVHGLGLVN PKKVFNFYNE
```

```
           ....|....| ....|....| ....|....| ....|....| ....|....|
                410        420        430        440        450
XP_0036179 LHSYLASAGI DGVKVDVQNI LETLGAGHGG RVKLARKYHQ ALEASISRNF
NP_0011057 LHSYLASAGI DGVKVDVQNV LETLGAGHGG RVMLARKYQQ ALEASVARNF
NP_0011057 LHAYLASCGV DGVKVDVQNI IETLGAGHGG RVAITRAYHR ALEASVAHSF
NP_0011475 LHAYLASCGV DGVKVDVQNI IETLGAGHGG RVAITRAYHR ALEASVARSF
AAM75140.1 LHSYLASCGI DGVKVDVQNI IETLGAGHGG RVTLTRSYHQ ALEASIARNF
ACG27589.1 LHAYLASCGV DGVKVDVQNI IETLGAGHGG RVAITRAYHR ALEASVARSF
AAA32975.1 LHAYLAACGV DGVKVDVQNI VETLGAGHGG RVALTRAYHR ALEASVARNF
BAI66422.1 LHAYLAACGV DGVKVDVQNI VETLGAGHGG RVALTRAYHR ALEASVARNF
AAT77910.1 LHAYLASCGV DGVKVDAQNI IETLGAGHGG RVSLTRAFHR ALEASVARSF
BAH19983.1 LHSYLADAGV DGVKVDVQCV LETLGGGLGG RVELTRQFHQ ALDSSVAKNF
XP_0028781 LHSYLASCGV DGVKVDVQNI IETLGAGLGG RVSLTRSYHQ ALEASIARNF


           ....|....| ....|....| ....|....| ....|....| ....|....|
                460        470        480        490        500
XP_0036179 PDNGIICCMS HNTDGLYSSK RSAVIRASDD FWPRDPASHT IHIASVAYNT
NP_0011057 PDNGIISCMS HSTDNLYSSK RSAVIRASDD FWPRDPASHT IHVASVAYNT
NP_0011057 PDNGCISCMC HNSDMLYSAR QTAVVRASDD FYPRDPASHT VHVASVAYNT
NP_0011475 PDNGCISCMC HNSDMLYSAR QTAVVRASDD FYPRDPASHT VHVASVAYNT
AAM75140.1 SDNGCIACMC HNTDSLYSAK QTAVVRASDD YYPRDPTSHT IHISSVAYNS
ACG27589.1 PDNGCISCMC HNSDMLYSAR QTAVVRASDD FYPRDPASHT VHVASVAYNT
AAA32975.1 PDNGCISCMC HNTDMLYSAK QTAVVRASDD FYPRDPASHT VHISSVAYNT
BAI66422.1 PDNGCISCMC HNTDMLYSAK QTAVVRASDD FYPRDPASHT VHISSVAYNT
AAT77910.1 PDNGCISCMC HNTDMLYSAR QTAVVRASDD FYPLDPASHT IHISSVAYNT
BAH19983.1 PDNGCIACMS HNTDALYCSK QAAVIRASDD FYPRDPVSHT IHIASVAYNS
XP_0028781 TDNGCISCMC HNTDGLYSAK QTAIVRASDD FYPRDPASHT IHIASVAYNS


           ....|....| ....|....| ....|....| ....|....| ....|....|
                510        520        530        540        550
XP_0036179 IFLGEFMQPD WDMFHSLHPM AEYHAAARAV GGCPIYVSDK PGHHDFNLLK
NP_0011057 VFLGEFMQPD WDMFHSVHPM AEYHAAARAV GGCAIYVSDK PGSHDFNLLK
NP_0011057 VFLGEFMQPD WDMFHSLHPA AEYHGAARAI GGCPIYVSDK PGNHNFELLR
NP_0011475 VFLGEFMQPD WDMFHSLHPA AEYHGAARAI GGCPIYVSDK PGNHNFELLR
AAM75140.1 LFLGEFMQPD WDMFHSLHPT AEYHGAARAI GGCAIYVSDK PGNHNFDLLK
ACG27589.1 VFLGEFMQPD WDMFHSLHPA AEYHGAARAI GGCPIYVSDK PGNHNFELLR
AAA32975.1 LFLGEFMQPD WDMFHSLHPA AEYHGAARAI GGCPIYVSDK PGNHNFDLLR
BAI66422.1 LFLGEFMQPD WDMFHSLHPA AEYHGAARAI GGCPIYVSDK PGNHNFDLLK
AAT77910.1 LFLGEFMQPD WDMFHSLHPA AEYHGAARAI GGCPIYVSDK PGNHNFELLK
BAH19983.1 VFLGEFMQPD WDMFHSVHPA AEYHASARAI SGGPLYVSDS PGKHNFELLR
XP_0028781 LFLGEFMQPD WDMFHSLHPT AEYHAAARAV GGCAIYVSDK PGNHNFDLLR


           ....|....| ....|....| ....|....| ....|....| ....|....|
                560        570        580        590        600
XP_0036179 KLVLPDGSIL RAKLPGRPTK DCLFSDPARD GKSLLKIWNM NDYSGVVGVF
NP_0011057 KLVLPDGSIL RAKLPGRPTR DCLFSDPARD GKSVLKIWNL NEHSGVVGAF
NP_0011057 KLVLPDGSVL RAQLPGRPAR DCLFSDPARD GASLLKIWNL NKCGGVVGVF
NP_0011475 KLVLPDGTVL RAQLPGRPTR DCLFSDPARD GASLLKIWNL NKCGGVVGVF
AAM75140.1 KLVLPDGSVL RAQLPGRPTR DSLFNDPARD GISLLKIWNM NKCSGVVGVF
ACG27589.1 KLVLPDGSVL RAQLPGRPTR DCLFSDPARD GASLLKIWNL NKCGGVVGVF
AAA32975.1 KLVLPDGSVL RAQLPGRPTR DCLFSDPARD GASLLKIWNM NKCAGVVGVF
BAI66422.1 KLVLPDGSVL RAQLPGRPTR DCLFSDPARD GASLLKIWNM NKCAGVVGVF
AAT77910.1 KLVLPDGSVL RARLPGRPTR DCLFVDPARD GASLLKIWNV NKCTGVVGVF
BAH19983.1 KLVLPDGSIL RARLPGRPTR DCLFADPARD GVSLLKIWNM NKYTGVLGVY
XP_0028781 KLVLPDGSVL RAKLPGRPTR DCLFADPARD GISLLKIWNM NKFTGIVGVF
```

```
                ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                     610         620         630         640         650
   XP_0036179  NCQGAGWCKV GKKNLIHDEN PGTVTDIIRA KDIDHLSTVA D--DKWTGDA
   NP_0011057  NCQGAGWCRV AKKNLIHDQQ PGTLSGVIRA QDVEHLGRVA D--HGWNGDV
   NP_0011057  NCQGAGWCRV TKRTRVHDAS PGTLTGTVRA DDVDAIARVA SDGGGWDGET
   NP_0011475  NCQGAGWCRV TKRTRVHDAS PGTLTGTVRA DDVDAIARVA GDGGGWDGET
   AAM75140.1  NCQGAGWCRI TKKTRIHDES PGTLTTSVRA ADVDAISQVA G--ADWKGDT
   ACG27589.1  NCQGAGWCRV TKRTRVHDAS PGTLTGTVRA DDVDAIARVA SDGGGWDGET
   AAA32975.1  NCQGAGWCRV AKKTRIHDEA PGTLTGSVRA EDVEAIAQAA G-TGDWGGEA
   BAI66422.1  NCQGAGWCRV VKKTRIHDEA PGTLTGSVRA EDVEGITQAT G-TDDCTGDA
   AAT77910.1  NCQGAGWCRI TKKTRVHDAA PGTLTGSVRA DDVDAIADVA G--TGWTGDA
   BAH19983.1  NCQGAAWSST ERKNIFHQTK TDSLTGSIRG RDVHSISEAS TDPTTWNGDC
   XP_0028781  NCQGAGWCKE TKKNQIHDTS PGTLTGSVCA DDADQISQVA G--EDWSGDS


                ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                     660         670         680         690         700
   XP_0036179  VIFSHLRGEV VYLPKDVSIP ITMKSREYEL FTIVPVKELP NG-----VKF
   NP_0011057  VVYLHVGGEV VYLPKNALLP VTLRSREYEV FTVVPLKHLP NG-----TSF
   NP_0011057  VVYAHRTREL VRLPRGVALP VTLGPLQYEV FHVCPLRAVV PG-----VSF
   NP_0011475  VVYAHRTREL VRLPRGVALP VTLGPLQYEV FHVCPLRAVV PG-----FSF
   AAM75140.1  IVYAYRSGDL IRLPKGASVP VTLKVLEYDL LHISPLKDIA SN-----ISF
   ACG27589.1  VVYAHRTREL VRLPRGVALP VTLGPLQYEV FHVCPLRAVV PG-----VSF
   AAA32975.1  VVYAHRAGEL VRLPRGATLP VTLKRLEYEL FHVCPVRAVA PG-----VSF
   BAI66422.1  VVYTHRAGEL VRLPRGATLP VTLKRLEYEL FHVCPVRAVA PD-----ISF
   AAT77910.1  VVYAHRSGEL IRLPKGATLP VTLKVLEFEL FHVCPVMTVA PGGGGGGVTF
   BAH19983.1  AVYSQSRGEL IVMPYNVSLP VSLKIREHEI FTVSPISHLV DG-----VSF
   XP_0028781  IVYAYRSGEV VRLPKGASIP LTLKVLEYEL FHISPLKEIT AN-----ISF


                ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                     710         720         730         740         750
   XP_0036179  APIGLIKMFN SGGAVKEFSS GFNG-V---- ---------- ----------
   NP_0011057  AAIGLLGMFN SGGAVRELRF GGED------ ---------- ----------
   NP_0011057  APVGLLDMFN AGGAVEECDV ISDAGG---- ---------- ----------
   NP_0011475  APVGLLDMFN AGGAVEECDV ISNVGG---- ---------- ----------
   AAM75140.1  APIGLLDMFN TGGAVEQVNV QVVEP--IPE FDGEVAS-EL TCSLPNDRPP
   ACG27589.1  APVGLLDMFN AGGAVEECDV ISDAGG---- ---------- ----------
   AAA32975.1  APIGLLHMFN AGGAVEECTV ETGEDGN--- ---------- ----------
   BAI66422.1  APIGLLHMFN AGGAVEECVV RTNEDDK--- ---------- ----------
   AAT77910.1  APIGLLDMFN SGGAVEECDV VRALDAAGEA EA-------- -------EAE
   BAH19983.1  APIGLVNMYN SGGAIEGLRY EAEK------ ---------- ----------
   XP_0028781  APIGLLDMFN SSGAIESMDI NPVTDKKHEL FDGEVSS--A SPALSDNRSP


                ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                     760         770         780         790         800
   XP_0036179  -ANVSMKVRG CGLFGAYSSA Q-PKLITVDS EEVEFSYEEE SGLVTIDLS-
   NP_0011057  -ADVELRVRG SGTVGAYSST K-PTCVAVDS KAVGFSYDAT CGLISFELG-
   NP_0011057  -KAMALRVRG CGRFGAYCSR E-PARCLLDS AEVEFSYDID TGLVSVDLL-
   NP_0011475  -KAMALRVRG CGRFGAYCSR E-PARCLLDS AEVEFSYDAD TGLVSVDLP-
   AAM75140.1  TATITMKARG CRRFGLYSSQ R-PLKCSVDK VDVDFVYDEV TGLVTFEIP-
   ACG27589.1  -KAMALRVRG CGRFGAYCSR E-PARCLLDS AEVEFSYDID TGLVSVDLL-
   AAA32975.1  -AVVGLRVRG CGRFGAYCSR R-PAKCSVDS ADVEFTYDSD TGLVTADVP-
   BAI66422.1  -AVVALRVRG CGRFGAYCSR R-PAKCSLDS ADVEFGYDAD TGLVTVDVP-
   AAT77910.1  AAVVRLRARG CGRFGAYSSR R-PARCALDA VEVEFSYDAD TGLVALDVP-
   BAH19983.1  -MKVVMEVKG CGKFGSYSSV K-PKRCVVES NEIAFEYDSS SGLVTFELDK
   XP_0028781  TALISLSVRG CGRFGAYSSQ R-PLKCAVGS TETDFTYDAE VGLVTLNLP-
```

```
                    ....|....| ....|
                          810
XP_0036179    VPEKELYQWN ISIDL
NP_0011057    IPDQEMYLWT VTVEY
NP_0011057    VPEQELYLWT LEIMI
NP_0011475    VPEQELYRWT LEIMV
AAM75140.1    IPTEEMYRWD IEIQV
ACG27589.1    VPEQELYLWT LEIMI
AAA32975.1    VPEKEMYRCA LEIRV
BAI66422.1    VPEEEMYRWT LEIRV
AAT77910.1    VPAHELYKWT VEIQV
BAH19983.1    MPIENKRFHL IQVEL
XP_0028781    VTSEEMFRWH VEILI
```

Alignment A1: Alignment of DIN10 protein sequence with similar sequences.