A toolkit for the rapid improvement of bioenergy crops: The cell wall fraction

A thesis submitted for the doctorate of philosophy

Antony Martin

B Math/B Sc (Hons)

March, 2016

Declarations

Statement of originality

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying subject to the provisions of the Copyright Act 1968.

Statement of collaboration

I hereby certify that the work embodied in this thesis has been done in collaboration with other researchers, or carried out in other institutions. I have included as part of the thesis a statement clearly outlining the extent of collaboration, with whom and under what auspices.

Statement of authorship

I hereby certify that the work embodied in this thesis contains published papers of which I am a joint author. I have included as part of the thesis a written statement, endorsed by my supervisor, attesting to my contribution to the joint publications.

Thesis by publication

"I hereby certify that this thesis is submitted in the form of a series of published papers of which I am a joint author. I have included as part of the thesis a written statement from each co-author; and endorsed by the Faculty Assistant Dean (Research Training), attesting to my contribution to the joint publications."

_____

Antony P. Martin

28/03/2016

_____

Date

Abstract

With the realisation of the impending impact of climate change, the scientific community has been working towards generating viable renewable energy sources that do not contribute to the greenhouse effect. Bioethanol is one form of renewable energy that could replace the majority of transportation fuels, however, its production efficiency and commercial viability relies on optimisation of agricultural quality and yields, biomass conversion to sugars, and fermentation of these sugars into ethanol. The focus of this dissertation will be the optimisation of agricultural quality and yields of bioenergy crops.

Most food crops have been domesticated for thousands of years providing selection pressure to increase harvest index (*i.e.* a high ratio of fruit/seed to vegetative biomass), among other traits. Bioenergy crops, however, are grown for their vegetative biomass more than fruit or seed yields, thus farmers seek a very different plant. To meet the demands of this growing industry, rapid improvements in agricultural yields of both soluble sugars and the cell wall fraction of bioenergy crops, driven by biomass increases, are required. Improvements through traditional breeding strategies, including screening of mutagenized populations, is limited by the availability of high-throughput methods for assessing phenotypes within large populations. Improvements through genetic modification are limited by a fundamental understanding of the biogenesis of vegetative biomass, the biosynthesis and composition of cell walls, and the accumulation of soluble sugars.

Here I present a toolkit for the rapid improvement of bioenergy crops. A holistic high-throughput screening strategy was developed for the bioenergy crop, *Sorghum bicolor*, which rapidly quantifies biomass, cell wall composition and soluble sugar concentrations of stalks, and it was subsequently applied to a large mutagenesis population. *Setaria viridis* was established as a model C_4 grass for the study of stalk development (the major source of bioenergy crop biomass). A developing internode was identified as an ideal experimental system for the study of the biogenesis of stalk tissue, the biosynthesis of cell walls and the import of

soluble sugars for storage. RNA sequencing was performed in this system as a resource for gene discovery in relation to these three processes. To complement this resource for gene discovery a technique for 3D visualisation of proteins and tissue structure, PEA-CLARITY was developed and Raman confocal imaging was applied to *S. viridis* stems for *in situ* determination of cell wall composition.

The high-throughput screening method will provide a tool to drive rapid improvements in commercial bioenergy crops and allow rapid assessment of phenotypes in a field setting. The use of *S. viridis* as a transformable model C_4 grass, along with genetic resources for gene discovery and cutting edge imaging techniques, will allow rapid improvements in understanding of the processes that underpin bioenergy crop yields and will undoubtedly uncover opportunities for crop improvement through genetic manipulation. This toolkit will be integral for the improvement of agricultural yields from bioenergy crops and, therefore, the commercial viability of the biofuel industry.

Acknowledgements

The work presented in this dissertation is the result of the combined effort of many people, many experiences which have been made available to me, and unconditional support from those around me.

There are too many people to individually thank, but I would like to especially acknowledge the wonderful team of people who have directly contributed, day in and day out, to the work presented here. Will for the daily grind and *everything* from the last 4 years, this is as much yours as it is mine, Chris G and Bob for unconditional support, Chris L during my times in Brisbane and China constantly providing thought provoking discussions, John Lunn for the opportunities in Germany, and Wayne, Nicole, Neil and Alex for providing a high quality field site in Gatton.

To my colleagues in the lab from Newcastle, Canberra, Gölm, Versailles, and Brisbane, my family and friends, the sugarcane farmers, the guys at Ethtec, and anybody else who has helped me along the way, I am very grateful for everything you have done that has contributed to the completion of this piece of work. Without all of you I would not have had the wonderful experiences I have had during this time, and would not have produced the quality of work that is presented here.

Table of Contents

CHAPTER ONE

Introduction

Thesis overview

This thesis by publication outlines the body of work that I have contributed through a series of scientific publications towards the development of a toolkit for the rapid improvement of bioenergy crops, with specific focus on improving the cell wall fraction of these crops for the production of bioethanol.

Chapter one gives an overview of the rationale for the work that was performed. It sets the scientific context, as of 2012 when the project was embarked upon, and how this research fits within it. The aims of the work and the resulting publication list are outlined.

Chapter two is a literature review outlining the current understanding of the production of biomolecules by C_4 grasses and their use as a biofuel feedstock. It specifically focuses on the composition and deconstruction of the cell wall fraction of C_4 grasses. This review of the literature identifies a need for the development of tools and resources for rapid improvement of bioenergy crops to meet the demands of a growing industry. Some of the research and work produced in this chapter were published as a book chapter.

Chapter three consists of one published journal article in '*Biotechnology for Biofuels*', which outlines a holistic high-throughput screening strategy for *Sorghum bicolor*, and one submitted manuscript for which '*Biotechnology for Biofuels*' reviewers have requested additional work, which remains underway, before re-submitting. This publication shows the use of the high-throughput screening strategy in identifying cell wall variants in a population of mutagenised *Sorghum bicolor* plants.

Chapter four consists of one published journal article in '*Biotechnology for Biofuels*'. It outlines the use of *Setaria viridis* as a model C_4 grass and a developing internode as an ideal experimental system for the study of biogenesis of grass stems, soluble sugar accumulation, and importantly for this body of work, cell wall biosynthesis.

RNA sequencing in this publication has provided a resource for gene discovery within this system.

Chapter five consists of one published journal article in '*Scientific Reports*' and some unpublished data generated by myself during the period of this thesis. Both are imaging techniques designed to enhance the study of genes identified in chapter four. The '*Scientific Reports*' publication outlines an imaging technique, PEA-CLARITY, which allows immunolabeling and visualisation of fluorescent proteins in whole intact tissues in 3D, and the unpublished work shows the use of Raman confocal microscopy for *in situ* imaging of cell wall composition in *S. viridis* stems.

Chapter six is a general discussion, which draws on each chapter to state the contribution this work has made to the scientific approaches used for the improvement of bioenergy crops. It provides lists of gene candidates for future manipulation of the cell wall fraction of biomass and outlines the future prospects of this work.

Research context

In 2012, response to human induced climate change was beginning to unfold and a large research effort toward developing bioenergy crops, and biotechnologies for efficiently deconstructing cell walls for their use as a biofuel was established. At the same time, rapid advances in genome sequencing technologies were driving a push away from *Arabidopsis thaliana* as the sole plant model organism. *Brachypodium dystachion* was established as a C_3 grass model organism, and as part of the C_4 rice consortium, funded by the Bill and Melinda Gates foundation, *Setaria viridis* was established as a C_4 grass model system. After producing genome sequences in 2010 and 2012 respectively, most of the work on these two newly established model organisms centred around the collection of natural diversity, the production of recombinant and mutant populations, development of transformation and crossing techniques and, in *S. viridis*, screening for phenotypes related to photosynthesis.

Within this context, we identified the major bottleneck to rapid improvements in commercial bioenergy crop yields as being the availability of robust, high-throughput phenotypic screening techniques for traits associated with high yielding bioenergy crops. We also identified *S. viridis* as an ideal organism for generating a fundamental understanding of stalk growth, cell wall biosynthesis and composition, and soluble sugar accumulation, since it is closely related to the major proposed bioenergy crops, *Sorghum bicolor*, *Zea mays*, *Miscanthusxgiganteus,* switchgrass and sugarcane with similar physiology and genetics. We hoped that moving this fundamental research into *S. viridis*, a plant with a small stature, small genome, and rapid life cycle, along with transformation capabilities and the availability of low cost short read transcriptome sequencing, would accelerate fundamental research in these C_4 grasses.

In addition, in 2012, the neuroscience world was beginning to move cellular imaging from 2D to 3D space to enable neuron tracings and visualisation of complex interactions in 3D. Adapting these neuroscience developments to plant cellular imaging allowed us to look at structure-function relationships in a new light, in 3D. The combination of these novel imaging techniques, including *in situ* imaging of cell wall components, genetic manipulations now possible in the C_4 model plant *S viridis* (not described in this thesis), and transcriptomic maps designed for gene discovery, will vastly improve our understanding of the molecular and structural workings of these bioenergy plants.

Research aims

Sweet *Sorghum bicolor* has been suggested as a high yielding crop species with low water and supplementary nutrient requirements that is capable of producing two harvests per year from a single planting [1]. Very little has been done, however, to select for bioenergy traits such as high biomass, cell wall saccharification efficiency and high soluble sugar storage in *S. bicolor*. Such trait optimisation requires the utilisation of segregating populations and, importantly, an effective

screening strategy to identify the highest quality, highest yielding plants. Considering these requirements, the first two aims of the project were:

Aim 1: Develop FTIR prediction models for a holistic high-throughput screening framework

Aim 2: Screen *S. bicolor* mutagenesis populations for cell wall variants

In addition, underpinning accelerated breeding strategies, a fundamental understanding of the cell wall of C_4 grasses, derived through a model system used for rapidly genetically manipulating cell wall characteristics will be integral to fine-tuning biomass for increased yields and availability of substrates. *Setaria viridis* is an emerging C_4 grass model species [2] that is closely related to *S. bicolor* sharing a common ancestor ~20 mya [3], compared with monocot/dicot divergence that occurred ~200 mya [4]. It is also both transiently and stably transformable [5]. Thus, its small stature and 8-week life cycle make it an ideal choice for fundamental research on C_4 grass cell walls.

Like all monocot grasses, *S. viridis* has a stem consisting of internodes, which represent a repeating unit that contains the full developmental progression between nodes. This developmental gradient provides an ideal system for investigating the molecular changes that drive cell wall deposition through the different developmental stages and therefore the third aim of the project is:

Aim 3: Provide a comprehensive map of molecular changes that underpin development of a grass internode and generate an RNA transcriptome resource using this experimental system in the C_4 grass model species *S. viridis*.

To investigate genes identified in **Aim 3** a classical approach would be to generate antibodies for the gene of interest to localise the protein and manipulate the gene function by RNAi knockdown or overexpression using transformation systems and study the resulting phenotype. To understand changes to cell walls as a result of

genetic manipulation, imaging of cell walls and proteins involved in biosynthesis will be required at the cellular level. Recent advances in 3D protein localisation in animal tissues has shown the utility of imaging proteins in whole tissues in 3D, therefore, in this project we aimed to:

Aim 4: Develop a method for 3D localisation of proteins in plant tissues.

Aim 5: Develop a method for imaging plant cell wall composition *in situ* in *Setaria viridis* stems.

Publications

Martin AP, Brown CW, Nguyen DQ, Byrt CS, Lambrides CJ, Grof CPL **Cell wall development in an elongating internode of '**Genetics and Genomics of *Setaria'*, Springer (2016). – *in print*

Martin AP, Palmer MP, Byrt CS, Furbank RT, Grof CPL **A holistic high-throughput screening framework for biofuel feedstock assessment that characterises variations in soluble sugars and cell wall composition in *Sorghum bicolor*** *Biotechnology for Biofuels* **6**, 1-13, doi:10.1186/1754-6834-6-186 (2013).

Martin AP, Palmer WM, Brown C, Byrt CS, Furbank RT, Grof CPL **Detection and characterisation of *Sorghum bicolor* cell wall variants using FTIR spectra generated from a high-throughput screening framework** *Biotechnology for Biofuels* Currently addressing reviewer comments before re-submission (2016)

Martin AP, Palmer WM, Brown C, Furbank RT, Lunn JE, Grof CPL **A developing *Setaria viridis* internode: an experimental system for the study of biomass generation in a C_4 model species** *Biotechnology for Biofuels* **9**:45 doi:10.1186/s13068-016-0457-6 (2016).

Palmer WM, Martin AP, Flynn JR, Reed SL, White RG, Furbank RT, Grof CPL **PEA-CLARITY: 3D molecular imaging of whole plant organs** *Nature: Scientific Reports* **5**, Article number: 13492 doi:10.1038/srep13492 (2015).

Conference presentations

Palmer WM, Martin AP, Flynn JR, Furbank RT, Grof CP **PEA-CLARITY: 3D immunohistochemistry of whole mount plant tissues** *American Society of Plant Biologists, Plant Biology, Minneapolis, USA* (2015) - Poster

Martin AP, Palmer WM, Mouille G, Lunn JE, Stitt M, Furbank RT, Grof CPL **Molecular characterisation of a developing *S. viridis* internode** *ComBio, Canberra, Australia* (2014) - Poster

Flynn JR, Martin AP, Palmer WM, Rank MM, Callister RJ, Smith KM **Whole tissue clearing using the CLARITY technique** *ComBio, Canberra, Australia* (2014) - Poster

Palmer WM, Martin AP, Furbank RT, Grof CP **PEA-CLARITY: 3D immunohistochemistry of whole mount plant tissues** *ComBio, Canberra, Australia* (2014) – Talk delivered by Palmer WM

Martin AP, Palmer WM, Mouille G, Lunn JE, Stitt M, Furbank RT, Grof CPL **Molecular characterisation of a developing *S. viridis* internode** *1st International Setaria Genetics Conference, Beijing, China* (2014) - Poster

Martin AP, Palmer WM, Furbank RT, Grof CPL **A holistic high-throughput screening framework for biofuel feedstock assessment that characterises variation in soluble sugar and cell wall composition in *Sorghum bicolor*** *7th European Plant Science organisation, Porto heli, Greece* (2013) - Poster

Martin AP, Byrt CS, Furbank RT, Grof CPL **FTIR spectroscopy for characterisation of cell wall variants** *18th International Botanical Congress, Melbourne, Australia* (2011) – talk delivered by Grof CPL

Additional publications

Li Z, Palmer WM, Martin AP, Wang R, Rainsford F, Jin Y, Patrick JW, Yang Y, Ruan YL **High invertase activity in tomato reproductive organs correlates with enhanced sucrose import into, and heat tolerance of, young fruit** *Journal of Experimental Botany*, doi:10.1093/jxb/err329 (2011).

McCurdy DW, Dibley S, Cahyanegara R, Martin AP, Patrick JW **Functional characterisation and RNAi-mediated suppression reveals roles for hexose transporters in sugar accumulation by tomato fruit** *Molecular plant* **3**, 1049-1063, doi:http://dx.doi.org/10.1093/mp/ssq050 (2010).

Edwards J, Martin AP, Andriunas F, Offler CE, Patrick JW, McCurdy DW **GIGANTEA is a component of a regulatory pathway determining wall ingrowth deposition in phloem parenchyma transfer cells of *Arabidopsis thaliana*** *The Plant Journal* **63**, 651-661, doi:10.1111/ (2010)

CHAPTER TWO

Literature review

2.1 Summary

The literature review presented in Chapter 2 of this dissertation forms the basis of the book chapter entitled '*Cell wall development in an elongating internode of Setaria*'. The book chapter acts as a supplement for more in-depth discussion of, and reasoning for the use of *Setaria viridis* as a model organism for the study of bioenergy. A PDF of the final book chapter is presented first in 2.2 and that is followed by the literature review in 2.3.

2.2 Publication 1: Cell wall development in an elongating internode of *Setaria*

Martin AP, Brown CW, Nguyen DQ, Byrt CS, Lambrides CJ, Grof CPL **Cell wall development in an elongating internode of *Setaria*** 'Genetics and Genomics of Setaria', Springer (2016) – *in print*

Statements of contribution of others: Publication 1

*I attest that Research higher degree candidate **Antony Martin** contributed extensively to the development of this publication as first author. Research and text from Antony's literature review presented in this chapter was re-appropriated and expanded in this book chapter. RNAseq data generated by Antony was used for cell wall gene analysis. Antony contributed intellectually to the generation of FTIR predictive models for cell wall components of Setaria presented in this chapter. The book chapter was reviewed by CPLG and CJL, figures 4-8, & 9 were prepared by CWB, and CJL based on data analysed by Antony and the chapter was review by CSB.*

_____

Antony Martin

_____

_ Christopher Brown

_____

Dac Nguyen

_____

Caitlin Byrt

_____

Christopher Lambrides

_____

Christopher Grof

Cell wall development in an elongating internode of *Setaria*.

Martin AP, Brown CW, Nguyen DQ, Byrt CS, Lambrides CJ, Grof CPL

Corresponding author:

Prof Christopher Grof
University of Newcastle
chris.grof@newcastle.edu.au

2

Contents

4

Introduction

The study of photoassimilate partitioning in plants has been a fundamental research area over many decades. The major focus of these studies has been to understand the allocation of carbon to some economic component of the plant, for example, grain in cereal crops, starch in root and tuber crops, and sucrose in 'sweet' crops. The goal has been to increase harvest index, the proportion of the economic component relative to the total plant biomass. These studies are ongoing, but in recent years researchers have directed greater attention to other questions of carbon allocation. The economic importance of biomass has changed as plant cell walls are increasingly used as a raw material for biofuel and biochemical production. This has influenced researchers to consider how plant cell walls are synthesised and what carbon resources are tied up in their production. It is not clear how much carbon could be reallocated from the synthesis of plant cell walls to other parts of the plant. To date, the majority of research on plant cell walls has focussed on grain crops such as rice, wheat and maize where the influence of grain cell wall composition is important for human nutrition [1]. There is less information about the biology of plant cell walls in vegetative tissues such as the developing culm.

Herein the genetic control of plant cell wall development in the culm is briefly reviewed and preliminary research investigating internode elongation of *Setaria*, a new model grass belonging to the Panicoideae, is introduced. An analysis of cell wall components of more than 200 *Setaria* germplasm lines has also been undertaken and is likely to provide the platform for more detailed investigation of plant cell walls in the developing culm of C_4 grasses. *Setaria* possesses the efficient C_4 pathway of photosynthesis and is therefore primed for high productivity under conditions of elevated light and temperature, similar to closely related species such as sugarcane (*Saccharum* spp.), Sorghum (*Sorghum*

bicolor (L.) Moench), switchgrass (*Panicum virgatum* L.) and Miscanthus spp. Hence, in addition to its role as a model for C_4 photosynthesis [2], *Setaria* can also serve as a model system for photoassimilate partitioning in the developing culm.

Partitioning of photoassimilate

Sucrose, the product of photosynthesis, is transported over long distances through the sieve elements of the phloem within the vascular system. Having reached growing sink organs, sucrose is hydrolysed by the enzyme sucrose synthase (SuSy) to produce the key molecule, UDP-Glucose. Photoassimilate partitioning viewed from the position of the metabolite UDP-Glucose can be considered in terms of three principal demands (Figure 1). Demand 1 is carbon directed towards the structural components of cellulose and callose, catalysed by the enzymes cellulose and callose synthase respectively, and non-cellulosic polysaccharides/pectin polymers catalysed by UDP-Glucose dehydrogenase. To fulfil Demand 1 and complete the synthesis of the growing cell walls, UDP-glucose can be interconverted to a range of five and six carbon nucleotide sugars through a series of complex enzymatic pathways [3]. Demand 2 is carbon directed towards respiration (glycolysis), lipid and starch biosynthesis, with the first step being catalysed by UDP-Glucose pyrophosphorylase. Demand 3 is carbon directed towards sucrose and catalysed primarily by sucrose phosphate synthase (SPS), sucrose phosphate phosphatase (SPP) and SuSy. Trehalose phosphate synthase (TPS) catalyses the synthesis of the key sensor metabolite Trehalose 6-phosphate (T6P); T6P is also reliant on UDP-Glucose as a substrate, however, the concentration of T6P is small and hence the demand upon UDP–Glucose is minor. Pyrophosphate: fructose 6-phosphate 1-phosphotransferase catalyses the reversible conversion of fructose 6-phosphate (Fru 6-P) and pyrophosphate (PPi) to fructose 1, 6-bisphosphate (Fru 1,6-BP) and inorganic phosphate (Pi).

6

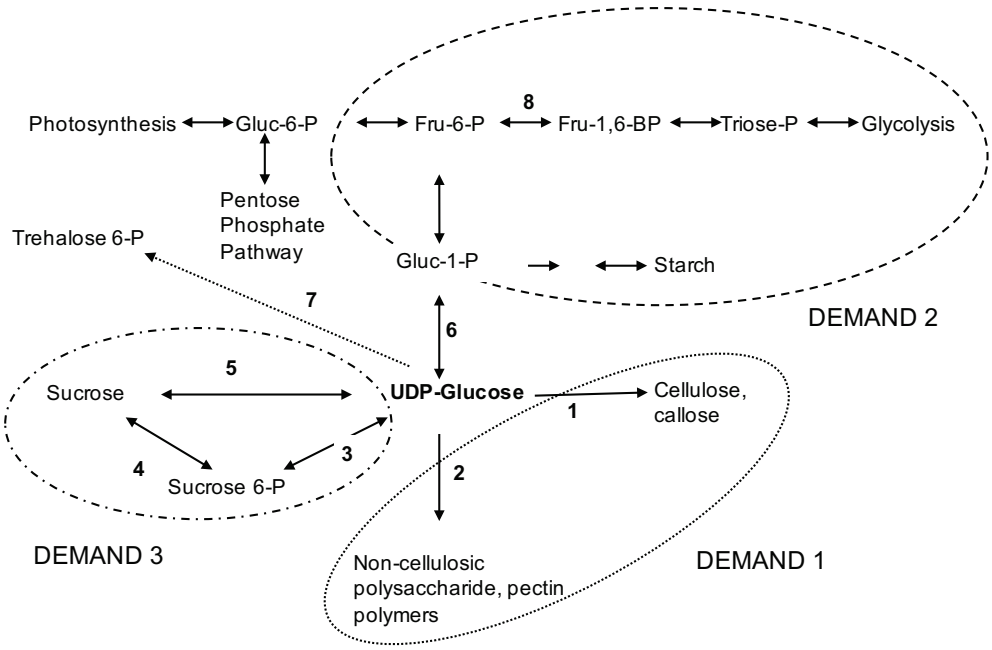


Figure 1. Photoassimilate partitioning in the growing sink organ, the culm, of *Setaria* and other monocot grasses. Photoassimilate partitioning, viewed from the central position occupied by the key metabolite UDP-Glucose, considered in terms of three principal demands. Demand 1 is carbon directed towards structural components including cellulose/callose catalysed by cellulose/callose synthase (**1**), and non-cellulosic polysaccharide catalysed by UDP-Glucose dehydrogenase (**2**). Demand 2 is carbon directed towards respiration, protein, lipid and starch synthesis, the first step being catalysed by UDP-glucose pyrophosphorylase (**6**). Demand 3 is carbon directed towards sucrose and catalysed primarily by the enzymes sucrose-phosphate synthase (**3**), sucrose-phosphate phosphatase (**4**) and sucrose synthase (**5**). Trehalose-phosphate synthase catalyses the synthesis of the sensor metabolite trehalose 6-phosphate (T6P) (**7**); Pyrophosphate:fructose 6-phosphate 1-phosphotransferase catalyses the reversible conversion of fructose 6-phosphate (Fru 6-P) and pyrophosphate (PPi) to fructose 1,6-bisphosphate (Fru 1,6-BP) and inorganic phosphate (Pi) (**8**).

Cell wall development in an elongating internode

The culm, the central axis of the grass shoot, is made up of multiple repeating phytomeric units, comprised of a node and an internode, where each node bears

a leaf. Within each growing internode, the intercalary meristem resides at the base of the internode, immediately above the node below. Acropetally above the meristematic region, cells undergo expansion, maturing towards the top of the internode. Primary cell walls are synthesised while the cell is growing whereas the secondary cell walls are produced as cell expansion ceases. The secondary cell wall is deposited inside the primary cell wall and characteristically is made up of three distinct layers (S1-S3), although in some fibre cells such as those of bamboo, many more layers have been reported [4].

The plant body is made up of more than 30 different cell types, each of which is likely to vary in cell wall composition and structure [5]. The complex nature of this extracellular layer is reflected in the number of genes, more than a thousand, that modulate cell wall construction and metabolism in maize [6]. In broad terms, the cell wall is composed of cellulose intertwined in a complex matrix with non-cellulosic polysaccharides and pectin. The cellulose microfibrils are synthesised by a large protein complex embedded in the plasma membrane whilst the production of many non-cellulosic polysaccharides involves partial synthesis in the Golgi apparatus and delivery in secretory vessels to the surface of the cell for assembly. Non-cellulosic polysaccharides interact with cellulose to form the strong resilient structure that is the cell wall [7]. In grasses, approximately 85% of the primary cell wall is composed of (1,4)-β-glucans (cellulose), and non-cellulosic polysaccharides: arabinoxylan (AX), glucurono(arabino)xylan (GAX), xyloglucan (XyG), and (1,3;1,4)-β-glucans (mixed linkage glucans; MLG). The remaining components are small amounts of pectin, structural proteins, unpolymerised phenolics and silica. The secondary cell wall is structurally rigid and is comprised of cellulose, AX, GAX, silica and heavy deposits of the phenolic polymer lignin, which together make up approximately 95% of secondary cell wall dry weight [8]. There are also small amounts of pectin and

unpolymerised phenolics such as ferulic and ρ-coumaric acid (See Table 1; adapted from [8]).

Table 1. The approximate composition of primary and secondary cell walls in typical eudicots and grasses expressed as percentage of dry weight (% Dry Wt.).

Components	Primary wall (% Dry Wt.)		Secondary wall (% Dry Wt.)	
	Grass	**Eudicot**	**Grass**	**Eudicot**
Cellulose	20-30	15-30	35-45	45-50
Non-cellulosic polysaccharide				
XyG	1-5	20-25	Minor	Minor
Xylans	20-40	5	40-50	20-30
MLG	10-30	Absent	Minor	Absent
Mannans & Glucomannans	Minor	5-10	Minor	3-5
Pectins	5	20-35	0.1	0.1
Structural proteins	1	10	Minor	Minor
Phenolics				
Ferulic & *ρ*-coumaric acids	1-5	Minor	0.5-1.5	Minor[a]
Lignin	Minor	Minor	20	7-10
Silica			5-15	Variable

After Vogel (2008).

[a] Except order Caryophyllales.

Cellulose

Cellulose, the most abundant polymer on earth, is an un-branched, β-1,4 linked chain of D-glucose, successively inverted 180° and ranging in size from 2 000–25 000 glucose residues [5, 9, 10]. The repeating unit in a single cellulose chain is the dimer, cellobiose. The polymerisation of cellulose is catalysed by cellu-

lose synthase (CesA) enzymes which belong to the Glycosyltransferase 2 superfamily [11]. Cellulose synthase-like (Csl) proteins encoded by the *Csl* genes also belong to this superfamily [12].

Immunogold labelling of CesA enzymes has revealed that they are embedded in the plasma membrane and form a particle rosette consisting of six rosette subunits [13]. Each subunit is composed of six CesA enzymes which may be catalytically active as single polypeptide or dimer subunits [14, 15]. The eight highly conserved transmembrane domains of the proteins encoded by the CesA gene family defines the barrel shape of the rosette and the cytosolic orientation of the enzyme active site such that glucan chains are produced and secreted from the centre of the barrel into the extracellular space [10, 16]. This cellulose synthase enzyme complex (CSC) produces multiple glucan chains that weave and bind together spontaneously via hydrogen bonds, to produce cellulose microfibrils, made up of between 12 and 36 interwoven glucan chains [17–19]. Recent NMR data and computational simulations of cellulose molecular dynamics indicate that 18- and 24- chain models are consistent with scattering and diffraction data; and spectroscopic measurements of Arabidopsis cellulose microfibrils indicated that the cellulose synthase complex is a hexamer of equimolar stoichiometry that synthesizes an 18-glucan chain microfibril [19–21]. The cellulose synthase enzyme complex uses UDP-glucose as the substrate and the growing glucan chain is guided through the plasma membrane by cortical microtubules which ensure organised orientation of the cellulose microfibrils. This facilitates anisotropic cell wall expansion perpendicular to cellulose microfibril orientation and maximises strength in secondary cell walls [22].

A complement of 10 genes encoding CesA has been identified in *A. thaliana* [12, 23], 10 in rice [24], 12 in sorghum [25] and 12 to 14 in maize [25, 26]. Mutant analysis has demonstrated that at least three non-redundant CesAs form a functional cellulose synthase enzyme complex (CSC) and in rice, OsCesA 4,

10

7 and 9 are required for secondary cell wall synthesis [24]. Expression analysis in maize has suggested that ZmCesA1 to 9 are involved in primary cell wall synthesis whilst ZmCesA10 to 12 are involved in secondary cell wall cellulose deposition [26, 27]. In *A. thaliana*, CesAs have been shown to be co-expressed in a range of tissues and demonstrate protein-protein interactions.

A number of other proteins have been proposed to play a role in cellulose synthesis, at least in Arabidopsis. Within the plasma membrane, CSI1 contributes to the mediation of CSC-microtubule alignment and in conjunction with KORRIGAN, a putative β-1,4 glucanase, may affect the motility of the CSC [19]. Extracellular proteins, such as COBRA and CTL1, glycosylphosphatidylinositol (GPI)–anchored proteins, may also impact upon the velocity of the glucan chain production driven by the CSC [19]. However, the precise mechanism of action of many of these ancillary proteins remains elusive. Mutations affecting primary cell wall cellulose synthesis tend to result in swollen cells, most likely due to isotropic cell expansion and inability to regulate cell turgor in an expanding cell [28]. In the grasses barley, maize and rice, mutations in secondary cell wall cellulose synthesis genes consistently result in a brittle stem phenotype [24, 29–31], whereas in *A. thaliana* disruption of secondary cell wall cellulose synthesis resulted in collapsed xylem elements [23].

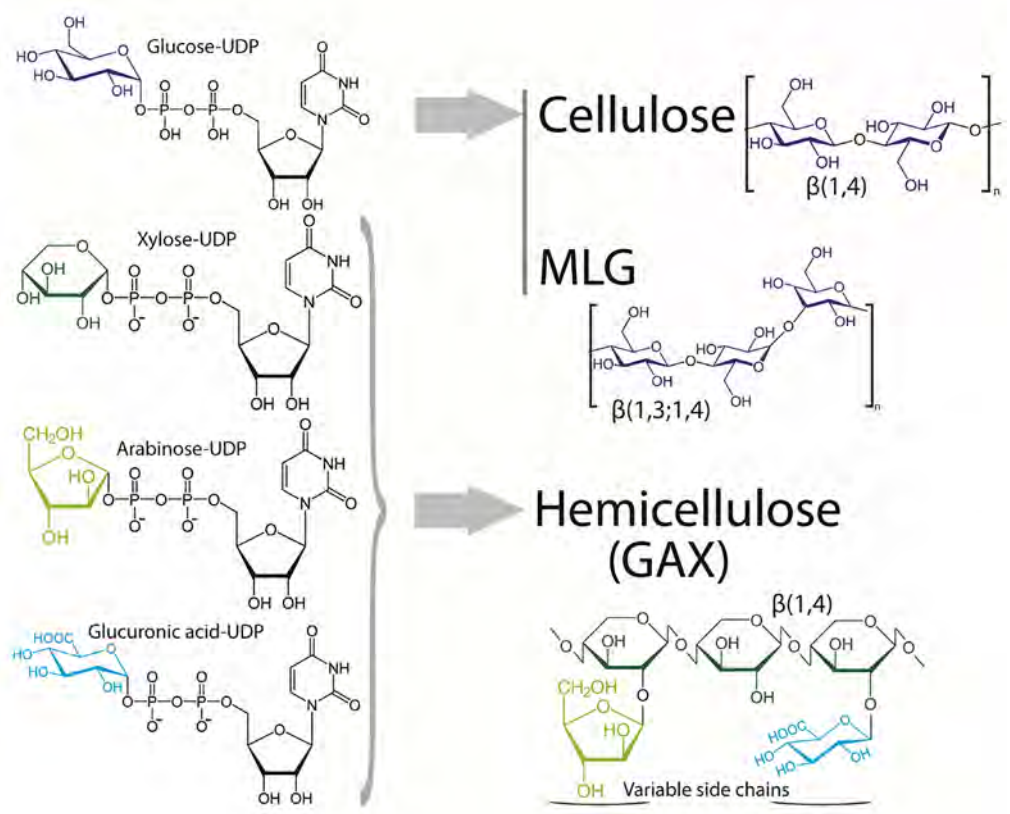


Figure 2. Major nucleotide sugar substrates required to build the cell wall polysaccharides of grasses (Demand 1). MLG; β(1,3;1,4)-glucans or mixed linkage glucans, G(AX); glucurono(arabino)xylan. Six carbon monomers are coloured in shades of blue and five carbon monomers are coloured in shades of green whilst the nucleotide functional groups are black.

Non-cellulosic polysaccharides

The major non-cellulosic polysaccharides found in grasses, MLG, AX and XyG, are polymers of β-(1,4) linked sugars, often with a glucan backbone, making them structurally similar to cellulose. Non-cellulosic polysaccharides are often composed of a mixture of six and five carbon sugars and their backbone is often highly substituted with monosaccharide or disaccharide side chains [23]. In Type II commelinoid grasses, as for plants possessing Type I cell wall

composition such as Arabidopsis, the widely accepted model describes non-cellulosic polysaccharides coating cellulose microfibrils, linking them together via hydrogen bonds and forming the cellulose-non-cellulosic polysaccharide network that confers mechanical strength to a cell wall [32]. For Type I cell walls, the microfibrils are tethered with xyloglucans and embedded in various pectins, whereas Type II cell walls are low in pectin.

Xylans

Glucurono(arabino)xylan (GAX) and arabinoxylan (AX), the predominant non-cellulosic polysaccharide in grass cell walls (Table 1; [8]), are polymers of β-1,4 linked D-xylose with variable non-repeating arabinose and glucuronic acid side chains for GAX, and arabinose substitutions for AX (Figure 2; [33]). The synthesis of xylans occurs in the Golgi using UDP-xylose as the substrate for the xylan backbone [34]. The GAX xylose backbone is highly substituted with arabinose and glucuronic acid side chains during synthesis in the Golgi, however, these side chains are removed upon secretion to the wall and ferulic acid is often linked to the remaining arabinose side chains [8]. This mode of synthesis has implications for cell wall molecular architecture. The arabinose and xylose side chains are usually attached at the O-2 and O-3 position [8], which blocks efficient hydrogen bonding to cellulose microfibrils and other GAX [35, 36]. Regulation and modulation of side chain substitution may play a pivotal role in cell wall expansion of type II primary cell walls.

Xylan biosynthesis requires initiation, elongation and termination of the xylan backbone and, in the case of GAX, addition of arabinose and glucuronic acid side chains [23]. In *A. thaliana, GUX1* and *GUX2* genes encoding glucuronyltransferases have been localised to the Golgi and when transcription is repressed resulted in reduced addition of side chains to the xylan backbone, however no reduction in synthesis of the xylan backbone was observed [37]. This observa-

tion indicates that xylan backbone synthesis occurs via an independent mechanism to side chain substitution. In grasses, a number of glycosyltransferase (GT) families have been implicated in both xylan backbone synthesis and side chain substitution of xylans [38]. Based upon a broad scale bioinformatics approach [39], the GT43 gene family encoding β-1,4-xylan synthases, have been implicated in xylan backbone synthesis whereas the GT47 gene family, encoding xylan α-1,2- or α-1,3-arabinosyl transferases are proposed to direct arabinose substitutions of the xylan backbone. The heterologous expression of wheat and rice genes belonging to the GT61 family, *TaXAT2*, *OsXAT2*, and *OsXAT3* in Arabidopsis *GUX1/GUX2* double mutants produced arabinosylated xylan, definitively demonstrating gain-of-function. The GT75 gene family is proposed to encode a glucuronosyltransferase required for the side chain addition of glucuronic acid [40].

Mixed linkage glucans

Mixed linkage glucans are a polymer of glucose with mixed β-1,3 and β-1,4 glycosidic linkages without any side chains [8]. The mixed linkages between glucose monomers result in a polymer very similar to cellulose but with regular bends along the chain. They are present in high proportions (10-30% cell wall dry weight) in expanding primary cell walls of grasses with maximum levels occurring at the peak of cell expansion [41], suggesting mixed linkage glucans play a major role in this developmental process.

Since MLGs are mostly specific to grasses, less is known about the molecular mechanisms of their synthesis. Identification of a quantitative trait locus for MLG content in barley grain led to the discovery of six rice cellulose synthase-like F (*CslF*) genes in the syntenic region of the sequenced rice genome [42]. These *CslF* genes were heterologously expressed in *A. thaliana*, which does not produce MLG naturally and small amounts of MLG were detected [42]. In addition, the *CslF* family is one of two grass specific *Csl* gene families supporting

the argument that *CslF* genes are responsible for biosynthesis of MLG, which is specific to grasses. Similarly, the *CslH* gene family, also specific to grasses has been shown to produce small amounts of MLG when heterologously expressed in *A. thaliana* [43]. Expression studies suggest that *CslF* genes are required in expanding primary cell walls, whilst *CslH* genes are required in seed endosperms or secondary cell walls where MLG is used as a carbon energy store [23]. Both gene families are thought to only introduce β-1,4 glycosidic bonds into MLGs and since heterologous expression in *A. thaliana* only produces minute quantities of MLG it is considered likely that another unidentified gene family is responsible for producing the β-1,3 linkages in the MLG polymer [23]. Although there is a substantial body of immunochemical and biochemical data that supports the synthesis of full length mixed linkage β-glucans in the Golgi apparatus [7], preparation of grass tissues with high pressure cryofixation to preserve cellular ultrastructure and antigenicity revealed the presence of CSLF6 in the plasma membrane and intracellular membranes [44], challenging the dogma that all non-cellulosic polysaccharides are synthesised and assembled in the Golgi complex.

Xyloglucan (XyG)

Xyloglucan (XyG), a major component of eudicot cell walls, is only present in minor amounts in grasses (Table 1). It is a polymer of β-1,4 linked D-glucose with xylose monomeric and polymeric (xylose, galactose, fucose) side chains at the O-6 position in grasses. Three xylosyl residues are substituted in repeating units of four glucose backbone monomers [45]. Following synthesis in the Golgi complex, directed by synthases encoded by the *CslC* gene clade, xyloglucan is transported to the cell membrane. The number of CslC enzymes contributing to biosynthesis and the mechanism for side chain addition remains largely undescribed in grasses [32, 45, 46].

Phenolics

Phenolics, secondary metabolites which are both extremely abundant and widely distributed in the plant kingdom, play important functional roles in pigmentation, growth, reproduction and pathogen resistance [47, 48]. They possess one or more aromatic rings bearing one or more hydroxyl branches and are derived from the Shikimate-Phenypropanoid pathway [47, 49]. In the cell walls of grasses, the significant phenolic compounds identified include the hydroxycinnamates, ferulic acid and ρ-coumaric acid, and the polyphenol lignin which is a complex polymer that varies in monomer composition between species and cell types.

Lignin

Lignin is the third most abundant heteropolymer (approximately 20% dry weight) in secondary cell walls of grasses, predominantly associated with vascular bundles. It is a complex aromatic polymer generated by irregular linkage of three main phenylalanine derived monolignols, Hydroxyphenyl (H), Guaicyl (G) and Syringyl (S) [50], with the resultant polymer being highly branched due to a variety of possible linkages (β-O-4, β-β, β-5) between monomers. The incorporation of lignin in plant cell walls through covalent binding to non-cellulosic polysaccharides provides rigidity, strength and a barrier against external physical forces and pathogen attack. Lignin also renders the xylem impermeable to water and solutes, thereby facilitating water and nutrient transport.

H lignin is uniquely present in grass cell walls with a typical grass cell wall containing ~35-49% G units, 40-61% S units, and 4-15% H units [8, 51]. The ratio of S:G lignin has been implicated in defining cell wall characteristics since G lignin is more highly branched than S lignin and therefore has greater opportunity to cross-link to cell wall polysaccharides. A lower S:G ratio therefore, typically results in a more rigid and less digestible cell wall [52][53].

16

The highly controlled process of lignification can be considered in terms of three principal stages 1) biosynthesis of monolignols in the cytosol; 2) transportation of these precursor monolignols across the plasma membrane and 3) lignification of the monolignols by oxidative polymerisation [54]. Stage 1 occurs on the cytosolic side of the ER, where chloroplast derived phenylalanine is deaminated by the enzyme phenylalanine ammonia lyase (PAL) to form cinnamic acid [50]. Successive hydroxylation and methylation reactions form ρ-coumaroyl CoA, an important metabolite and branching point of monolignol or flavonoid biosynthesis pathways. At this point monolignol biosynthesis also diverges to produce ρ-coumarate aldehyde and ρ-coumaroyl shikimic acid through the catalytic activity of hydroxycinnamoyl CoA reductase (CCR) and hydroxycinnamoyl CoA shikimate:quinate hydroxycinnamoyl transferase (HCT), respectively. The aldehyde is then converted into the ρ-coumaryl alcohol, catalysed by hydroxycinnamyl alcohol dehydrogenase (CAD) and subsequently converted into the H monomeric lignin subunit [55]. In the alternate arm of the pathway, ρ-coumaroyl shikimic acid undergoes serial reduction reactions by ρ-coumaroyl shikimate 3′-hydroxylase (C3H), ρ-hydroxycinnamoyl CoA shikimate:quinate ρ-hydroxycinnamoyl transferase; (HCT/CST), Caffeoyl CoA O-methyl transferase (CCoAOMT) and (hydroxy)cinnamoyl CoA reductase (CCR) to produce coniferaldehyde. Synapaldehyde is produced by the catalytic activity of ferulic acid/coniferaldehyde/coniferyl alcohol 5-hydroxylase (F5H) and caffeic acid/5-hydroxyferulic acid O-methyl transferase (COMT) in series upon coniferaldehyde. Synapy and coniferyl alcohol, produced by the activity of (hydroxyl) cinnamyl alcohol dehydrogenase (CAD), are transported through the plasma membrane to the apoplasm and oxidised to produce the remaining two monomeric lignin subunits, Syringyl and Guaicyl lignin [56, 57].

Three possible mechanisms have been proposed for the transport of the monolignols to the apoplasm for polymerisation. The hydrophobic monolignols may

diffuse passively through the plasma membrane thereby preventing their toxic accumulation in cells [58]. The widely accepted mechanism of vesicular secretion of the monolignols into the extracellular space by the Golgi complex has recently been challenged. TEM (transmission electron microscopy) autoradiographs of radiolabelled monolignols in cells undergoing lignification in Lodgepole pine (*Pinus contorta* Dougl. Ex Loud.) tracheids, have allowed monolignol transport to be visualized, revealing that they are not translocated via the Golgi [59]. An alternative mechanism of lignin subunit transport has been proposed, involving ATP-binding (ABC) transporters, initially based upon transcript profiling, which shows tight co-expression of some ABC transporters with monolignol biosynthesis genes [60]. ABC transporters belong to a gene superfamily with 130 putative members reported in Arabidopsis [61] and more than 130 in rice [62]. They are considered to be responsible for the transport of metabolites, signalling molecules lipids and proteins across cell membranes, with some members demonstrated to be phenolic/polyphenolic transporters [63]. The putative ABC transporter AtABC29, highly expressed in Arabidopsis roots and anthers, has been identified as a monolignol transporter by both microarray and experimental analysis [64]. Mutant Arabidopsis knockout plants of AtABC29, demonstrate increased levels of ρ-coumaryl alcohol in the cytosol and reduced lignification [64]. Six closely related ABC transport members that belong to the pleiotropic drug resistance (PDR) subfamily, have been proposed as potential transporters responsible for trafficking of the two remaining monolignols through the plasma membrane [63].

Once delivered to the apoplasm, the monolignols are primed for polymer assembly by dehydrogenation (reduction of the HCA-CoA esters) catalysed by several enzymes including peroxidases, laccases and polyphenol oxidases [56, 63]. The monolignol radicals are randomly cross-coupled to one another or to a

growing polymer complex via endwise coupling predominantly by β-O-4 linkages, to form the three dimensional, branched, interlocking lignin networks [51, 55, 56, 65]. Based upon micro-autoradiography, lignification is initiated at Ca^{2+} rich nucleation sites, as calcium is required for the polymerisation of coniferyl alcohol, where lignin is covalently linked with the non-cellulosic polysaccharide and pectin constituents of the primary cell wall [55]. During the early stages of cell wall lignification, H lignin is believed to be deposited and functions as a scaffold to determine cell shape prior to deposition of G and S lignin [66]. In the grasses sugarcane and rice, lignification begins in the cell walls of the protoxylem vessels, progresses to the middle lamellae of fibre cells, the secondary wall of metaxylem vessels and finally to the secondary wall of the fibres [67][5].

Hydroxycinnamates

The hydroxycinnamates, ferulic acid and ρ-coumaric acid, are intermediates in the monolignol biosynthesis pathway and constitute up to 4% and 3% of cell wall dry weight respectively [8]. A unique feature of grass cell walls is the covalent ester linkage of ferulic acid to arabinose side chains of arabinoxylan or glucoarabinoxylan [68]. The arabinoxylan linked ferulates may form dimers, trimers and tetramers that cross link adjacent arabinoxylans during cell wall deposition and lignification [69]. Furthermore, ferulic acid is able to provide the bridge to form lignin-ferulate-polysaccharide complexes through ester-ether linkages [5, 70] thereby filling the functional role of structural proteins in Type I cell walls by cross linking non-cellulosic polysaccharide, cellulose and lignin constituents [8]. It has also been suggested that ferulic acids act as initiation sites for lignin polymerisation within the secondary cell wall [71] and genetically engineering monolignol ferulate conjugates specifically into poplar xylem significantly increased cell wall digestibility [72].

Pectins

Pectin is a family of complex galacturonic acid-rich polysaccharides comprising ~35% and 5% of eudicot and grass cell walls, respectively. They consist of a galacturonan backbone that can be substituted at various positions with simple sugars including xylose, rhamnose and apiose, as well as complex side chains [73]. Four types of pectin polysaccharides have been described namely, homogalacturonan (HG), xylogalacturonan (XGA), apiogalacturonan, rhamnogalacturonan I (RGI) and rhamnogalacturonan II (RGII) with HG accounting for approximately 65% of total pectin in plants. Pectins, particularly HG and RGII, play a principal role in shaping and strengthening the cell wall by acting as adhesive agents in the middle lamella [73].

Growth of the cell wall

The widely accepted architectural model of primary cell walls of both eudicots and monocots consists of cellulose microfibrils coated in non-cellulosic polysaccharides and tethered together covalently by hydrogen bonds to create a strong matrix network. Growth of the cell involves irreversible expansion of the cell wall in conjunction with an influx of water into the cell. The growth process, culminating in the synthesis, secretion and intercalation of new moieties into the developing cell wall requires vacuolar enlargement and solute uptake to maintain osmotic potential, hence turgor pressure.

Cell growth is physically constrained by the cell wall and dogma dictates that enzymes act specifically upon the non-cellulosic polysaccharide tethers (MLGs in grasses) to loosen their links to the cellulose microfibrils, hence promoting slippage or ‘wall loosening’. However, plant endoglucanases and endotransglycosylase/hydrolases do not significantly influence cell wall relaxation primarily because the xyloglucans accessible to these enzymes are not ‘load bearing’ [74]. Furthermore, solid-state Nuclear Magnetic Resonance data generated from

mung bean cell walls [75], indicated that less than 8% of cellulose surfaces are coated with xyloglucan.

Empirical evidence indicates that the widely accepted model of primary cell wall architecture needs to be revisited and the tethering role of xyloglucans re-considered. A 'biomechanical hotspot' hypothesis has been proposed [32], whereby limited cellulose-cellulose contact takes place at mechanical junctions possibly mediated by a xyloglucan monolayer binding the hydrophobic surfaces together. Cell wall extensibility may be promoted by expansin activity, through some undescribed mechanism, at these sites. Pectins are also proposed to play a significant role, as contact between pectins and the cellulose microfibrils is estimated to be greater than that of xyloglucan.

Molecular characterisation of the elongating *Setaria* internode.

One way to investigate the genes involved in internode development is to analyse the gene transcript levels. Although highly complex, cell walls are constructed from a limited number of components, predominantly cellulose microfibrils, non-cellulosic polysaccharides, lignin and pectin; and many of the genes involved in making these components are known. To begin unravelling this complexity in *Setaria,* gene expression in internodes of *Setaria* undergoing development from cellular division, cellular expansion of the primary cell wall and culminating in secondary cell wall deposition have recently been analysed (Figure 3; [76]). Similar strategies have been used for other grass species including rice [77], maize [27, 78] and sugarcane [79]. Previous maize (Bosch et al. 2011) and sugarcane studies (Casu et al. 2007), undertook microarray transcriptomic analysis comparing entire elongating and fully elongated culm internodes. In previous studies of elongating internodes differences in gene expression relative to mature fully elongated internodes are observed, however some

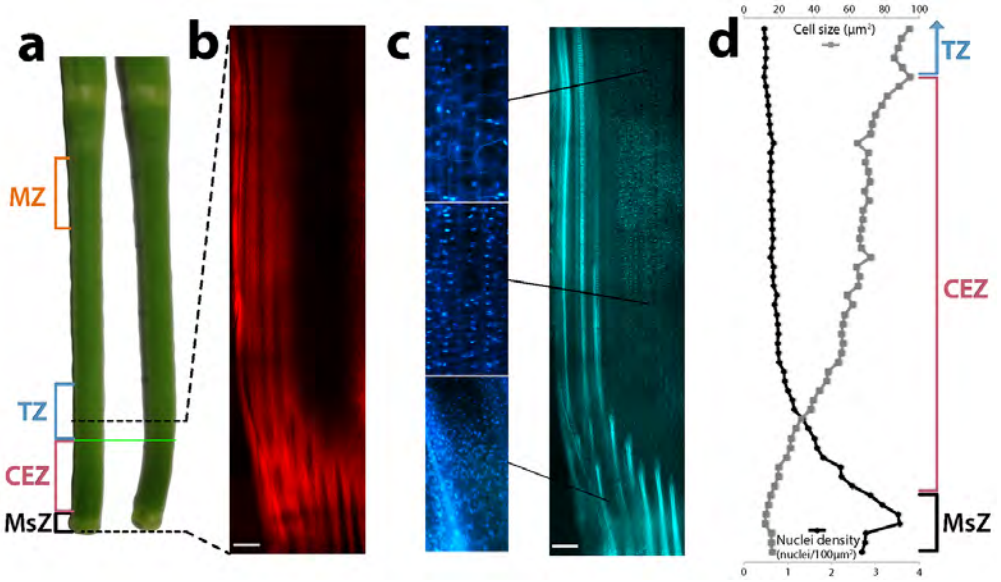


Figure 3. Regions of *Setaria* internode 5 selected for RNA isolation and sequencing. a) Internode 5 harvested at the 'half head emergence' developmental stage with its leaf sheath stripped, alongside an equivalent internode where the lower, flexible zone has been bent to display the difference between the upper rigid and the lower more flexible zone. The green line indicates the interface between the flexible and rigid zones of the internode. The harvested meristematic (MsZ - black), cell expansion (CEZ - pink), transitional (TZ - blue), and the mature (MZ - orange) zones are indicated. **b**) and **c)**, Longitudinal, vibratome cut, 50 µm thick section of the lower region of the internode (indicated by dotted black lines), stained with DAPI and viewed under UV illumination with **b)** the red chlorophyll emissions isolated, and **c**) the blue DAPI emissions isolated with enlarged regions offset (white scale bars are 50 µm). **e**) Nuclei density (black circles; nuclei/100 μm^2) and cell size (grey squares; μm^2) measured using image J were plotted at intervals along the lower region of the internode. This figure was adapted from Martin *et al.*, 2016.

changes are likely to be dampened by homogenisation of the different developmental regions within the internode during sample preparation. Recently a more refined approach was taken with an elongating maize internode, where sections 10 cm in length were divided into 1 cm sections prior to detailed analysis [27].

By studying 1 cm sections the authors observed that the expression of genes encoding key enzymes in the biosynthesis and modification of cellulose, non-cellulosic polysaccharide (predominantly glucuronoarabinoxylan) and lignin, peaked in the transitional region of internode development prior to completion of the deposition of secondary cell walls.

Setaria CesA and *Csl* genes

Synthases involved in producing the cell wall 'backbone' are encoded by members of the *CesA/Csl* superfamily of genes [15]. Cellulose is synthesised by members of the *CesA* clade; mannans and glucomannans by members of the *CslA* clade; xyloglucans, by members of *CslC* clade, and mixed-linkage β-glucans, by members of the *CslF* and *CslH* clades [7]. Ten of the 13 *Setaria CesA* genes identified here exhibited a greater than 1 Log_2 fold change in the region of cell elongation as compared to the meristematic region (Figure 4). Most striking is the fourfold Log_2 change in three of the *CesA* genes (*CesA 4, 10 and 12*) in the transitional and maturing regions of the internode. Phylogenetically, these three genes are most similar to *CesA4 (*At5g44030*), 7* (At5g17420) and *8* (At4g18780) from Arabidopsis and *CesA10, 11, 12(13)* from maize, which have been implicated in secondary cell wall cellulose synthesis [26]. This sustained elevated expression is at odds with the report of ongoing but reduced expression of a broad cohort of genes in the maturation region of the maize internode [27].

Genes belonging to the *CslA* clade were highly expressed in *Setaria* (with one exception, *CslA11*). *CslA* expression was particularly high in the meristematic and elongation zones of the internode (Figure 5). In maize, the cell walls within the elongation zone were measured to contain 3% mannan and then decreased acropetally in the internode [27], matching the *Setaria CslA* gene expression reported here. The genes belonging to the *CslC* subgroup putatively encode the

xyloglucan synthases and these exhibited an elliptical profile, with highest expression in the elongation zone and relatively lower expression in the transitional and maturing zones of the internode.

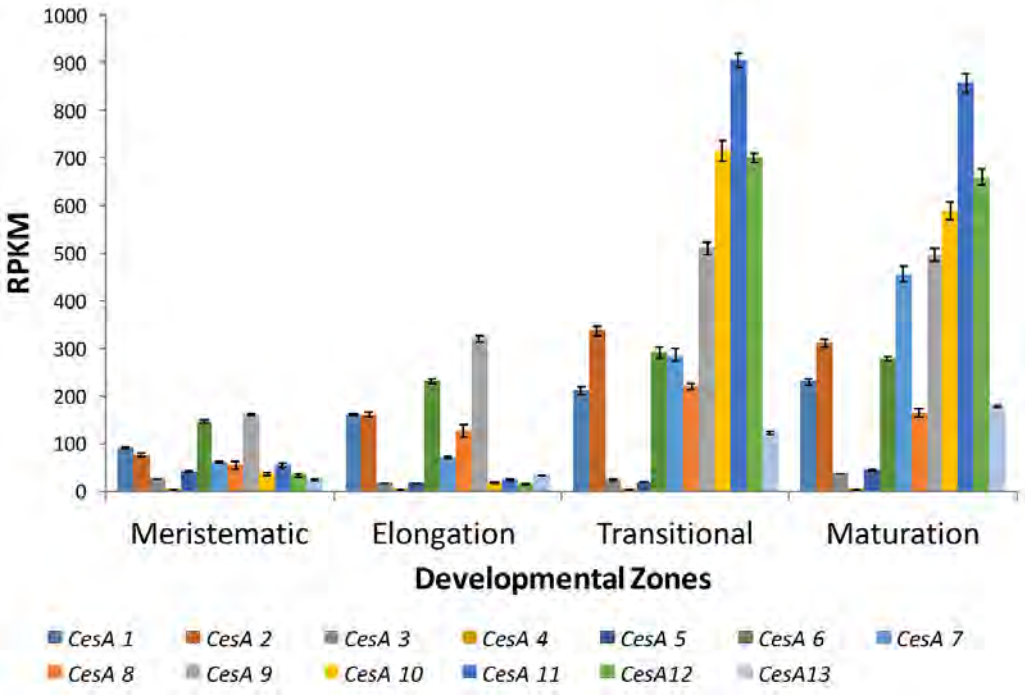


Figure 4. Expression profile of *CesA* genes in an elongating internode of *Setaria*. RNAseq reads per kilobase per million reads (RPKM) expression levels of all identified cellulose synthase A (*CesA*) transcripts in *S. viridis* showing maximal *CesA* expression during the transitional stage of internode development. An average of 4 biological replicates ± SE are displayed in the meristematic, cell elongation, transitional, and mature zones of the internode. Data was obtained from (76).

CslF6, a member of the *CslF* subgroup exhibited greater expression in the transitional zone relative to the meristematic zone and this higher expression was also observed in the maturing region of the internode (Figure 5). The expression of *CslH2* was highest in the elongation zone and close to zero in the transitional and maturation regions of the internode. The expression of *CslH1* increased

with the maturity of the internode although the level of expression was extremely low.

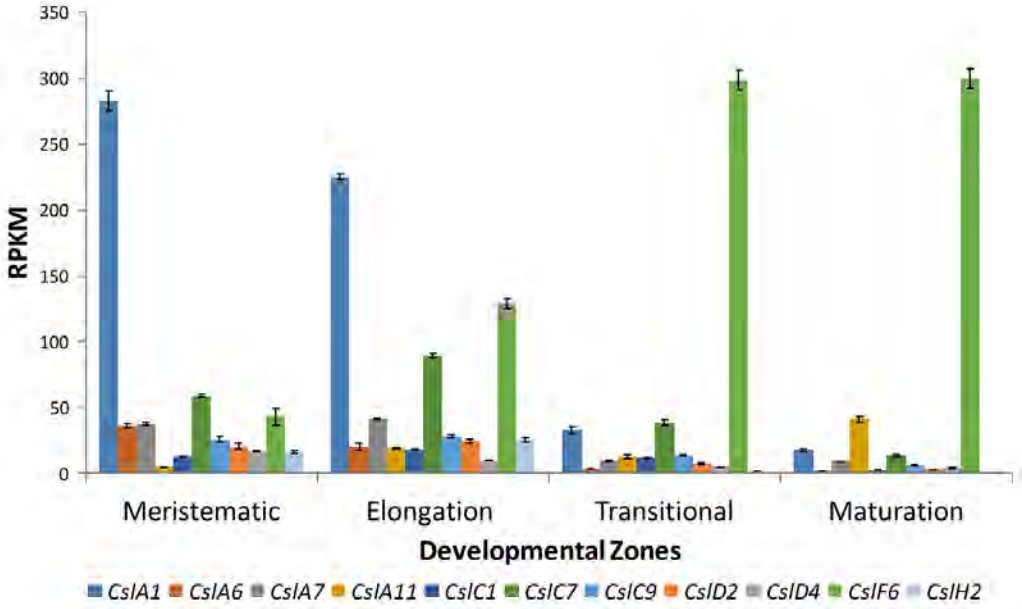


Figure 5. Expression profile of *Csl* genes in an elongating internode of *Setaria*. RNAseq reads per kilobase per million reads (RPKM) expression levels of all identified cellulose synthase-like (*Csl*) transcripts in *S. viridis* showing differential expression of *Csl* genes through internode development. An average of 4 biological replicates ± SE are displayed in the meristematic, cell elongation, transitional, and mature zones of the internode. Data was obtained from (76).

Setaria heteroxylan synthase genes

Xylan, represented in grasses as AX or GAX, makes up between 20 and 40% (dry weight) of the primary cell wall and between 40 and 50% of the secondary cell wall composition (Table 1; [8]). In maize, GAX makes up more than 30% of the primary cell wall and close to 50% in the secondary cell walls [27]. The glycosyltransferases contributing to the synthesis of heteroxylans are encoded by members belonging to a large superfamily of genes [80]. *Setaria* genes belonging in the GT family 47 (Figure 6) exhibit high levels of expression and

substantial Log_2 fold changes in the transitional and maturing regions of the internode, consistent with these glycosylases being involved in elongation of the xylan backbone. Similarly, a number of members of the GT family 8 increase strongly in expression in these same regions, consistent with their purported role in xylan backbone substitution [81].

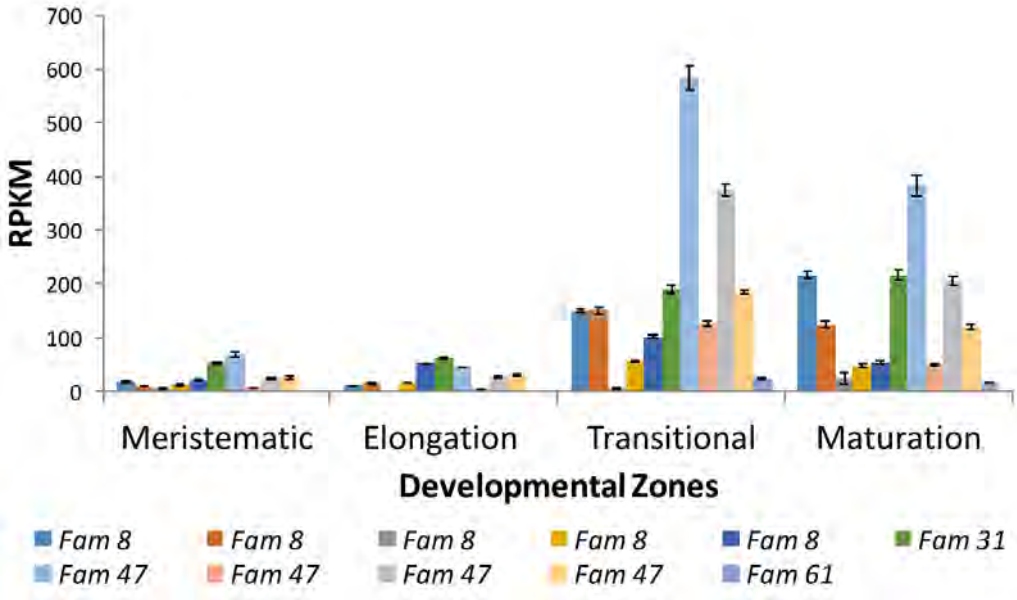


Figure 6. Expression profile of Glycosyltransferase (GT) genes in an elongating internode of *Setaria*. RNAseq reads per kilobase per million reads (RPKM) expression levels of identified glycosyltransferase (*GT*) transcripts in *S. viridis* showing maximal expression of *GT* genes in the transitional zone of the developing internode. An average of 4 biological replicates ± SE are displayed in the meristematic, cell elongation, transitional, and mature zones of the internode. Data was obtained from (76).

Setaria lignin biosynthesis genes

The genes encoding the ten enzymes participating in catalytic steps of lignin biosynthesis (See Figure 7) are well known [51]. The most highly expressed representatives of each group in *Setaria* are illustrated in Figure 8. All of the *Setaria* lignin biosynthesis genes analysed here are considered to be 'highly enriched in the internode' in that the Log_2 fold expression is many times higher

than expression of the same gene in a ‘whole plant’ transcriptome analysis encompassing multiple tissues across three developmental stages of growth, namely seed germination, vegetative growth and reproduction [82].

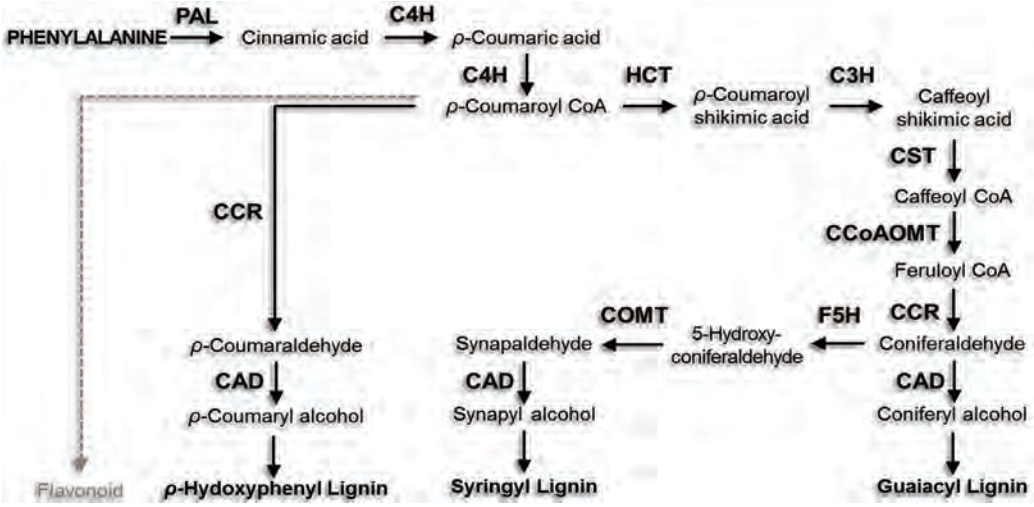


Figure 7. The lignin biosynthesis pathway. This model depicts the ten enzymatic steps leading to monolignol synthesis. The hydroxycinnamyl alcohols produced, Guaiacyl, ρ-Hydroxyphenyl and Syringyl polymerise to form lignin. The enzymes and abbreviations are: phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H), 4-coumarate-CoA ligase (4CL), hydroxycinnamoyl CoA:shikimate transferase (HCT), p-coumarate 3-hydroxylase (C3H), caffeoyl CoA O-methyltransferase (CCoAOMT), cinnamyl CoA reductase (CRR), ferulate 5-hydroxylase (F5H), caffeic acid O-methyltransferase (COMT), and cinnamyl alcohol dehydrogenase (CAD). Modified from (90).

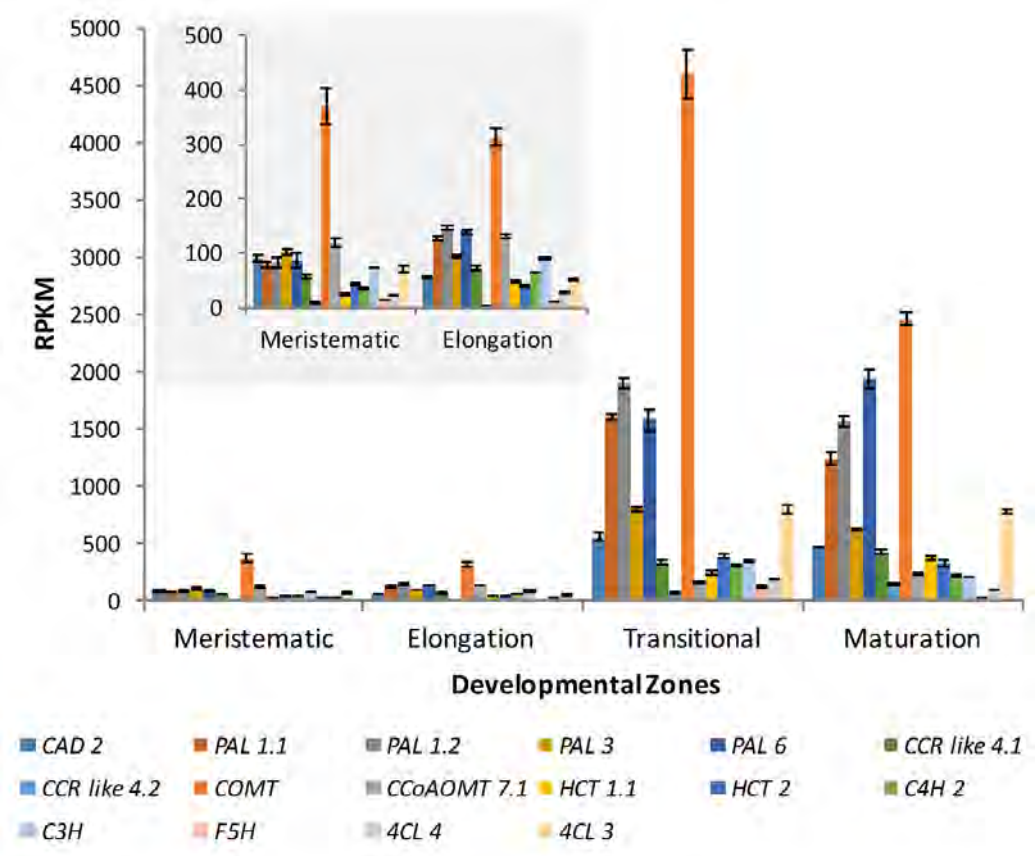


Figure 8. Expression profile of genes encoding enzymes involved in lignin biosynthesis in an elongating internode of *Setaria*. RNAseq reads per kilobase per million reads (RPKM) expression levels of identified lignin biosynthesis enzyme transcripts in *S. viridis* showing maximal expression in the transitional zone of the developing internode. An average of 4 biological replicates ± SE are displayed in the meristematic, cell elongation, transitional, and mature zones of the internode. Data was obtained from (76).

Transcript profiling of carbon demands

An overview of the levels of cell wall gene transcripts and other genes associated with the partitioning of photoassimilate (see Figure 1) within the context of a developing *S. viridis* internode was recently published [76].

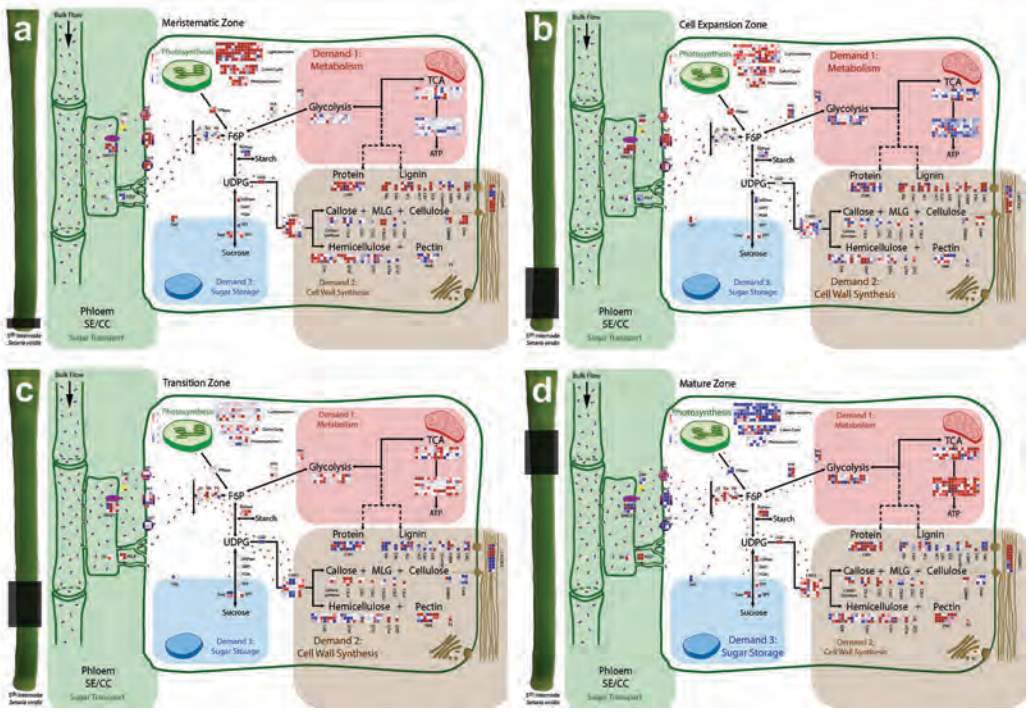


Figure 9. Schematic model of carbon demands during internode development based on transcriptomic analysis. A schematic pathway of carbon import into the developing *S. viridis* fifth internode, showing gene expression relating to the three major demands on carbon supply. Genes were assigned into categories (as labelled on the diagram) and their expression values were displayed as squared $\log_2$ fold change from the squared average of the four internode zones for each gene. Up-regulated genes are shown in *blue*, whilst down-regulated genes are in *red*. *MLG* mixed linkage glucans, *RS* raffinose synthase, *PDLP* plasmodesmata localised proteins, *SWEET* sugars will eventually be exported transporters, *CWI* cell wall invertase, *HT* hexose transporters, *PLT* polyol transporters, *SUT* sucrose transporters, *CI* cytosolic invertases, *HK* hexokinase, *FK* fructokinase, *FPBase* fructose-1,6-bisphosphatase, *F6P* fructose-6-phosphate, *PFK* phosphofructokinase, *AGPase* glucose-1-phosphate adenylyltransferase, *UGD*UDP-glucose 6-dehydrogenase, *UGPase* UDP-glucose phosphorylase, *G6PI* glucose-6-phosphate isomerase, *PGM* phosphoglucomutase, *SPP* sucrose phosphate phosphatase, *SPS* sucrose phosphate synthase, *Susy* sucrose synthase, *TMT* tonoplast membrane transporter, *CWPS* cell wall precursor synthesis, *CWP* cell wall protein, *PAL* phenylalanine ammonia-lyase,*C4H* cinnamate-4-hydroxylase,

4CL 4-hydroxycinnamate CoA ligase, *HCT* hydroxycinnamoyl transferase, *C3H* coumarate 3-hydroxylase, *CCR* cinnamoyl-CoA reductase, *CCoAOMT* caffeoyl CoA 3-*O*-methyltransferase, *COMT* caffeic acid *O*-methyltransferase, *F5H* ferulate 5-hydroxylase, *CAD* cinnamyl alcohol dehydrogenase, *CSLA-H* cellulose synthase-like A-H, *CesA* cellulose synthase complex, *GT* glycosyltransferase, *PME* pectin methylesterase, *PS* pectin synthesis, *SE* sieve element, *CC* companion cell. *Blue* and *purple dots* are representative of sugar flow. Reproduced from (76).

As expected, within the developing internode, the ‘meristematic zone’ and ‘cell expansion zone’ showed gene expression related to metabolism/energy production, some aspects of cell wall synthesis, namely callose synthases, and hemicellulose synthesis via GT families, and AGPases implying starch accumulation. Plasmodesmatal proteins were active, which indirectly implied that the carbon source was predominantly sugars imported through symplasmic connections from sieve elements containing sugars delivered to the growing sink tissue by bulk flow (Figure 9).

The major carbon demand in the ‘transitional zone’ of the internode shifted to cell wall synthesis, specifically lignin, cellulose, cell wall proteins, and some GT families (hemicellulose synthesis). UDP-glucose 6-dehydrogenase (UGP) and cell wall precursor synthesis (CWPS) proteins were highly expressed showing how carbon from UDP-glucose pools is directed into cell wall synthesis at this stage of internode development where cell expansion ceases and extensive deposition of secondary cell walls occurs (Figure 9).

In the ‘mature zone’ of the developing internode, the cell walls are now thickened secondary walls, and the internode can realise its capacity for sugar accumulation. This is indicated by the expression of sugar transporter genes from the polyol transporter (PLT), hexose transporter (HT), sucrose effluxer (SWEET) and sucrose transporter (SUT) families. Photosynthesis genes are also

active in this mature region, indicating an additional carbon source that does not require long distance transport of sugars. Tonoplast monosaccharide transporters (TMTs) also show increased expression in this zone indicating their role in sugar storage within vacuoles in mature internodes (Figure 9).

This recently published overview of carbon partitioning within a developing internode provides a valuable platform for gene discovery; not only identifying gene families that are active at different stages of internode development, but also allowing identification of the dominant isoforms from each family for genetic manipulation and further experimental study of internode development in a C_4 grass.

Germplasm to support the molecular dissection of plant cell wall components in *Setaria*

We assembled 214 ecotypes of *Setaria* from around the world and evaluated them in a glasshouse experiment for agronomic traits and percent lignin. The plants were grown under controlled conditions 24°C/20°C day/night temperature in 2013. There was large variation for all traits measured (Table 2) including height, maturity (anthesis), number of internodes, biomass, stem width, tiller number, number of leaves and percent lignin. Not surprisingly, biomass and other growth attributes such as stem width and number of internodes were positively correlated to height and maturity (see PCA in Figure 10; a narrow angle between vectors is indicative of high correlation). To a lesser extent percent lignin was also positively correlated to height and maturity while tiller number and leaf number traits in this germplasm set were independent of height and maturity.

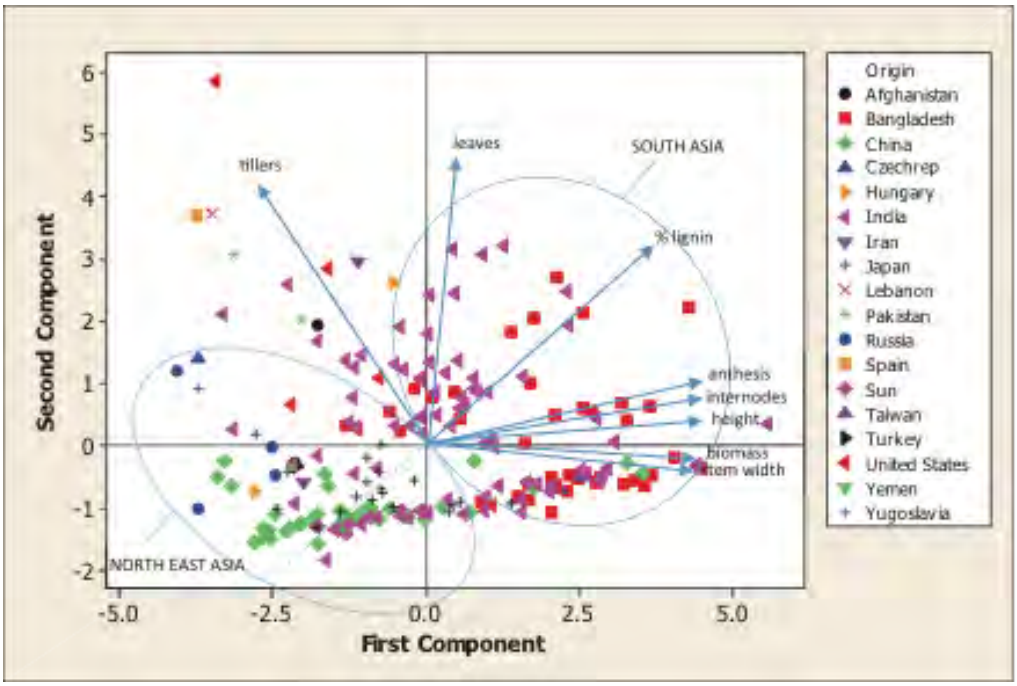


Figure 10. **Geographical relationship to physiological attributes of *Setaria* genotypes** Principal component analysis (PCA) of physiological attributes, including lignin composition of stem tissue categorised by geographical origin of 214 *Setaria italica* germplasm lines collected from around the world. The trial was conducted at the University of Newcastle, NSW Australia under controlled temperature of 24/20 day/night and natural day-length (10.5-14.5 hrs)

With respect to percent lignin the germplasm could be broadly divided into two groups, South Asia and North East Asia. Germplasm from South Asia (India, Bangladesh) was generally later maturing and had higher percent lignin compared to the group from North East Asia (China, Taiwan, Japan, Russia) that were earlier maturing and had lower percent lignin. Nevertheless, germplasm within the same maturity group varied greatly for percent lignin and this material could be exploited to study lignin deposition.

32

Table 2. Mean, minimum and maximum values for several attributes measured in 214 *Setaria italica* germplasm lines grown in a controlled environment glasshouse at The University of Newcastle, NSW Australia in 2013.

Attribute	Mean	Minimum	Maximum
Height (mm)	1276.0	110.0	2673.0
Maturity (days to anthesis)	69.2	30.0	201.0
Number of internodes	10.3	3.0	19.0
Biomass (g dry weight)	3.8	0.04	15.9
Stem width (mm)	10.7	3.0	19.0
Number of tillers	2.5	1.0	21.0
Number of leaves	22.3	5.0	72.0
Percent lignin	20.3	16.6	23.8

Techniques for the analysis of *Setaria* cell wall traits

To characterise the genetic factors that contribute to variation in cell wall traits, robust phenotyping systems are needed. Identification of traits that enhance the conversion potential of plant biomass into liquid fuels requires screening of thousands of genetic variants. Phenotyping systems that are accurate, low cost and high throughput are needed. Examples of phenotyping systems include enzymatic assays to evaluate saccharification efficiency of biomass, chromatographic tools to identify differences in composition of cell walls and spectroscopic methods such as near-infrared (NIR) and Fourier-transform infrared (FTIR) and pyrolysis molecular beam mass spectroscopy [83]. These techniques are combined with conventional chemical assays to analyse a set of samples and build calibration models to link compositional information with spectral variation [84]. These models and screening tools are then used to efficiently predict the biochemical properties of unknown samples based on their spectral fingerprint.

Fourier Transform Infra Red Spectroscopy: Application of a robust tool for cell wall characterisation

Infrared (IR) spectroscopy can be readily applied as a rapid non-destructive tool for the investigation of cell wall composition in eudicots and monocot grasses. IR spectroscopy measures the transitions between molecular vibrational energy states as a result of absorption of mid-IR radiation. The vibrational energy levels are unique to each molecule and the frequencies of these molecular vibrations depend upon atomic mass, the geometric arrangement of the atoms and the strength of their chemical bonds [85]. Mid-infrared (MIR) absorption spectra (4,000 to 400 cm^{-1}; 25000-2500nm) provide characteristic fundamental vibrational modes directly, generating sharp peaks that are more readily interpreted than near-infrared (NIR; 13,000 to 4,000 cm^{-1}; 700-2500nm) or far-infrared (FIR; 10 to 400 cm^{-1}) spectra which reflect the broad overtones or combined bands of fundamental vibrations [85].

Although Fourier Transform MIR (FT-MIR or FTIR) spectra have been used in the analysis of plant cell walls for more than 30 years [86], the power of this tool has only been realised more recently, when coupled with multivariate data analysis techniques. FTIR spectroscopy is a robust and accurate method for high-throughput screening of cell wall mutations in experimental plant tissues, such as the model species *Arabidopsis thaliana* [87] and *Zea mays* coleoptiles [88]. In Sorghum, variation in the cell wall composition of a range of phenotypically different ecotypes grown in the field was identified following Principal Component Analysis (PCA) of FTIR spectra generated from ground stem samples (Figure 11). The PCA separation of the different Sorghum ecotypes (Figure 11B) is based predominantly upon the wavenumbers corresponding to cellulose and lignin as shown in Figure 11C. Total lignin content was measured by the Acetyl Bromide method (Figure 11D), validating the differences in cell wall composition, particularly lignin, predicted by FTIR and PCA.

34

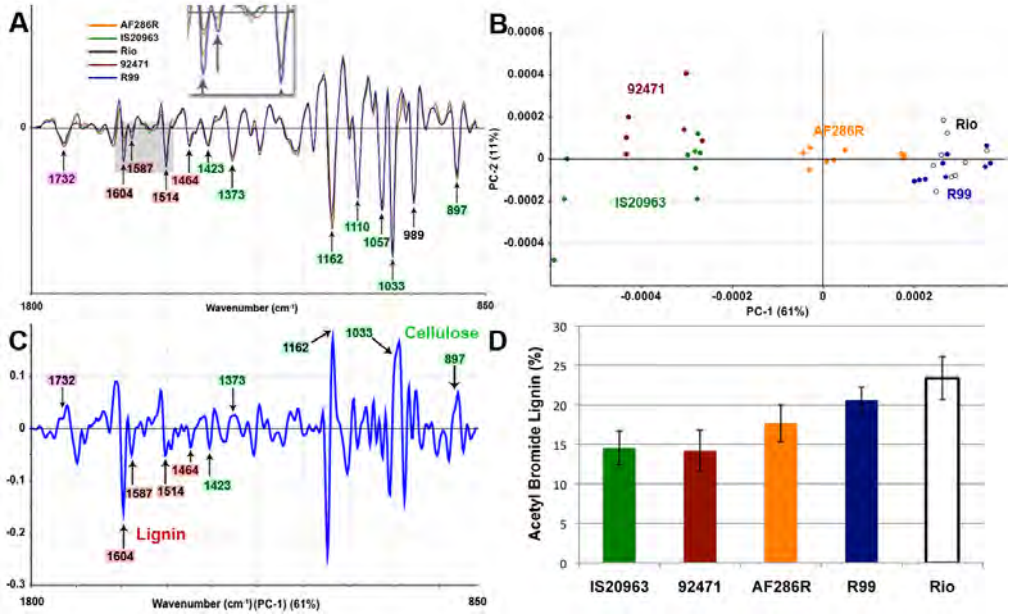


Figure 11. Principal Component Analysis (PCA) of FTIR spectra collected from *Sorghum bicolor* accessions grown under field conditions. A) Second derivative spectra with extended multiplicative scatter correction (EMSC) applied; Inset highlights the differences between accessions at wavenumbers 1587 and 1604 cm^{-1} (total lignin). **B**) Scores plot of PC1 against PC2 (**C**) PC1 loadings plot showing the main peaks contributing to PC1 and (**D**) Total lignin as determined by the Acetyl Bromide method (Martin, Byrt, Furbank and Grof; unpublished).

Development of a Partial Least Squares (PLS) model to predict cell wall composition in *Setaria*

FTIR has also been used as a tool to assess *Setaria* as a model for unravelling the complexity of cell wall composition, assembly and subsequently deconstruction in grasses. Drawing upon a set of 214 *S. italica* accessions a prediction of total lignin was made with an accuracy of close to 90% using FTIR spectra (Figure 12). This research is currently being extended to incorporate the prediction of the full gamut of carbohydrate components making up *Setaria* cell walls following hydrolysis.

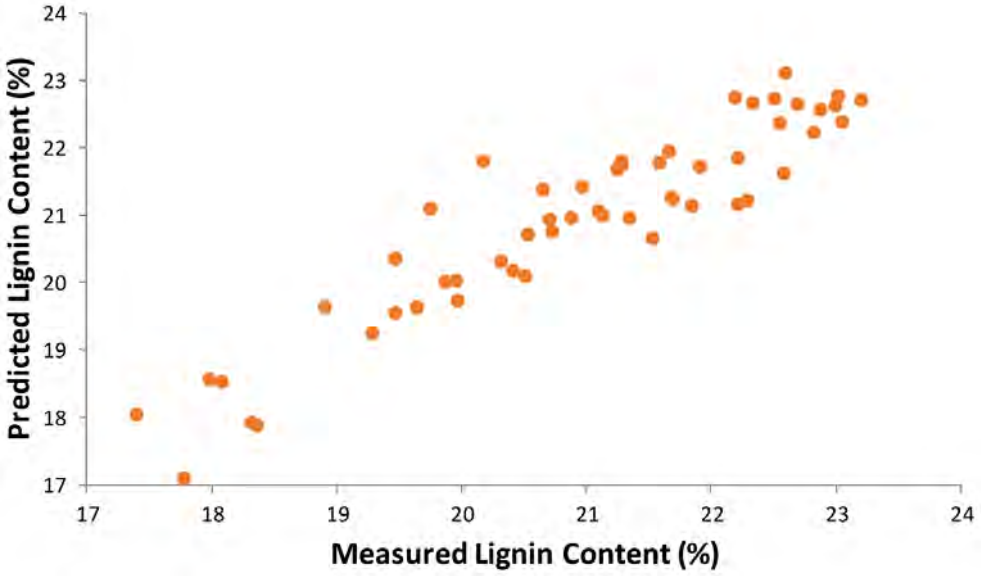


Figure 12. Prediction of the percentage of total lignin in accessions of *Setaria italica* using FTIR spectra. FTIR based predictions of lignin content in *Setaria* stem tissue that had been ground to a fine powder were made using a 2 factor partial least squared (PLS) predictive model. These values were plotted against wet chemistry data for lignin performed on the same samples.

Conclusion

The gene regulatory network controlling cell wall synthesis and modification in the developing *Setaria* stem internode is not yet known. One strategy to identify this network is gene expression profiling. In addition to the genes involved in the biosynthesis of components of the cell wall there are also upstream transcription factors influencing cell wall biosynthesis. Identification and verification of these 'master regulators' in grasses is lagging behind relative to information about model eudicot species. We have reported a high-throughput experimental platform that will aid in the discovery of key genes involved in photoassimilate partitioning in the largest contributor to monocot plant biomass, the culm. This experimental platform includes *Setaria* gene expression analysis

and characterisation of differences in biomass composition of *Setaria* stems using FTIR. The genetic co-linearity of grasses [89] ensures that orthologs of key candidate genes identified in *Setaria* can be rapidly identified in the closely related annual or perennial C_4 monocot species including economically important crop plants such as maize and Sorghum.

References

1. Collins HM, Burton RA, Topping DL, Liao M-L, Bacic A, Fincher GB. Variability in fine structures of noncellulosic cell wall polysaccharides from cereal grains: Potential importance in human health and nutrition. Cereal Chem 2010;87:272–282.
2. Brutnell TP, Wang L, Swartwood K, Goldschmidt A, Jackson D, Zhu X-G, Kellogg E, Van Eck J. Setaria viridis: A Model for C4 Photosynthesis. Plant Cell 2010;22:2537–2544.
3. Bar-Peled M, O'Neill MA. Plant nucleotide sugar formation, interconversion, and salvage by sugar recycling. Annu Rev Plant Biol 2011;62:127–155.
4. Parameswaran N, Liese W. On the fine structure of bamboo fibres. Wood Sci Technol 1976;10:231–246.
5. de Oliveira Buanafina MM, Cosgrove DJ (2013) Cell walls: structure and biogenesis. In: Sugarcane Physiol. Biochem. Funct. Biol. John Wiley & Sons Ltd, p. 307–329.
6. Penning BW, Hunter CT, Tayengwa R, et al. Genetic resources for maize cell wall biology. Plant Physiol 2009;151:1703–1728.
7. Carpita NC. Progress in the biological synthesis of the plant cell wall: new ideas for improving biomass for bioenergy. Curr Opin Biotechnol 2012;23:330–337.
8. Vogel J. Unique aspects of the grass cell wall. Curr Opin Plant Biol 2008;11:301–307.
9. Richmond T. Higher plant cellulose synthases. Genome Biol 1: reviews 3001.1 –3001.6.
10. Taylor NG. Cellulose biosynthesis and deposition in higher plants. New Phytol 2008;178:239–252.
11. Campbell JA, Davies GJ, Bulone V, Henrissat B. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochem J 1997;326:929–939.
12. Richmond TA, Somerville CR. The cellulose synthase superfamily. Plant Physiol 2000;124:495–498.
13. Kimura S, Laosinchai W, Itoh T, Cui X, Linder CR, Brown RM. Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant vigna angularis. Plant Cell 1999;11:2075–2086.
14. Emons AMC, Mulder BM. The making of the architecture of the plant cell wall: How cells exploit geometry. Proc Natl Acad Sci U S A 1998;95:7215–7219
15. Carpita NC. Update on mechanisms of plant cell wall biosynthesis: How plants

make cellulose and other (1→4)-β-D-Glycans. Plant Physiol 2011;155:171–184.

16. Cosgrove DJ. Growth of the plant cell wall. Nat Rev Mol Cell Biol 2005;6:850–861.

17. Niimura H, Yokoyama T, Kimura S, Matsumoto Y, Kuga S. AFM observation of ultrathin microfibrils in fruit tissues. Cellulose 2010;17:13–18.

18. Fernandes AN, Thomas LH, Altaner CM, Callow P, Forsyth VT, Apperley DC, Kennedy CJ, Jarvis MC. Nanostructure of cellulose microfibrils in spruce wood. Proc Natl Acad Sci 2011;108:1195–1203.

19. McFarlane HE, Döring A, Persson S. The cell biology of cellulose synthesis. Annu Rev Plant Biol 2014;65:69–94.

20. Hill JL, Hammudi MB, Tien M.The Arabidopsis cellulose synthase complex: A proposed hexamer of CESA trimers in an equimolar stoichiometry. Plant Cell 2014;26:4834–4842.

21. Oehme DP, Downton MT, Doblin MS, Wagner J, Gidley MJ, Bacic A. Unique aspects of the structure and dynamics of elementary Iβ cellulose microfibrils revealed by computational simulations. Plant Physiol 2015;168:3–17.

22. Mutwil M, Debolt S, Persson S. Cellulose synthesis: a complex complex. Curr Opin Plant Biol 2008;11:252–257.

23. Doblin MS, Pettolino F, Bacic A. Evans Review : Plant cell walls: the skeleton of the plant world. Funct Plant Biol 2010;37:357-381.

24. Tanaka K, Murata K, Yamazaki M, Onosato K, Miyao A, Hirochika H. Three distinct rice cellulose synthase catalytic subunit genes required for cellulose synthesis in the secondary wall. Plant Physiol 2003;133:73–83.

25. Paterson AH, Bowers JE, Bruggmann R, et al. The Sorghum bicolor genome and the diversification of grasses. Nature 2009;457:551–556.

26. Appenzeller L, Doblin M, Barreiro R, Wang H, Niu X, Kollipara K, Carrigan L, Tomes D, Chapman M, Dhugga K. Cellulose synthesis in maize: isolation and expression analysis of the cellulose synthase (CesA) gene family. Cellulose 2004;11:287–299.

27. Zhang QS, Cheetamun R, Dhugga KS, et al (2014) Spatial gradients in cell wall composition and transcriptional profiles along elongating maize internodes. BMC Plant Biol. 2014;14:27-46. doi: Artn 27Doi 10.1186/1471-2229-14-27

28. Arioli T, Peng L, Betzner AS, et al. Molecular analysis of cellulose biosynthesis in Arabidopsis. Science 1998;279:717–720.

29. Aohara T, Kotake T, Kaneko Y, Takatsuji H, Tsumuraya Y, Kawasaki S. Rice BRITTLE CULM 5 (BRITTLE NODE) is involved in secondary cell wall formation in the sclerenchyma tissue of nodes. Plant Cell Physiol 2009;50:1886–1897.

30. Sindhu A, Langewisch T, Olek A, Multani DS, McCann MC, Vermerris W,

Carpita NC, Johal G. Maize Brittle stalk2 encodes a COBRA-like protein expressed in early organ development but required for tissue flexibility at maturity. Plant Physiol 2007;145:1444–1459.

31. Kokubo A, Sakurai N, Kuraishi S, Takeda K. Culm brittleness of barley (Hordeum vulgare L.) mutants is caused by smaller number of cellulose molecules in cell wall. Plant Physiol 1991;97:509–514.

32. Park YB, Cosgrove DJ. Xyloglucan and its Interactions with Other Components of the Growing Cell Wall. Plant Cell Physiol 2015;56:180–194.

33. Faik A. Xylan Biosynthesis: News from the Grass. Plant Physiol 2010;153:396–402.

34. York WS, O'Neill MA. Biochemical control of xylan biosynthesis - which end is up? Curr Opin Plant Biol 2008;11:258–265.

35. Carpita NC. Hemicellulosic polymers of cell walls of Zea coleoptiles. Plant Physiol 1983;72:515–521.

36. McCann MC, Carpita NC. Designing the deconstruction of plant cell walls. Curr Opin Plant Biol 2008;11:314–320.

37. Mortimer JC, Miles GP, Brown DM, et al. Absence of branches from xylan in Arabidopsis gux mutants reveals potential for simplification of lignocellulosic biomass. Proc Natl Acad Sci U S A 2010;107:17409–17414.

38. Doering A, Lathe R, Persson S. An update on xylan synthesis. Mol Plant 2012;5:769–771.

39. Mitchell RA, Dupree P, Shewry PR. A novel bioinformatics approach identifies candidate genes for the synthesis and feruloylation of arabinoxylan. Plant Physiol 2007;144:43–53.

40. Zeng W, Jiang N, Nadella R, Killen TL, Nadella V, Faik A. A glucurono(arabino)xylan synthase complex from wheat contains members of the GT43, GT47, and GT75 families and functions cooperatively. Plant Physiol 2010;154:78–97

41. Kim J-B, Olek AT, Carpita NC. Cell wall and membrane-associated Exo-β-d-glucanases from developing maize seedlings. Plant Physiol 2000;123:471–486.

42. Burton RA, Wilson SM, Hrmova M, Harvey AJ, Shirley NJ, Medhurst A, Stone BA, Newbigin EJ, Bacic A, Fincher GB. Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1,3;1,4)-beta-D-glucans. Science 2006; 311:1940–1942.

43. Doblin MS, Pettolino FA, Wilson SM, Campbell R, Burton RA, Fincher GB, Newbigin E, Bacic A. A barley cellulose synthase-like CSLH gene mediates (1,3;1,4)-β-D-glucan synthesis in transgenic Arabidopsis. Proc Natl Acad Sci 2009;106:5996–6001.

40

44. Wilson SM, Ho YY, Lampugnani ER, Van de Meene AML, Bain MP, Bacic A, Doblin MS. Determining the subcellular location of synthesis and assembly of the cell wall polysaccharide (1,3; 1,4)-β-D-Glucan in Grasses. Plant Cell 2015;27:754–771.

45. Pauly M, Gille S, Liu L, Mansoori N, de Souza A, Schultink A, Xiong G. Hemicellulose biosynthesis. Planta 2013;238:627–642.

46. Scheller H V, Ulvskov P. Hemicelluloses. Annu Rev Plant Biol 2010;61:263–289.

47. Cheynier V, Comte G, Davies KM, Lattanzio V, Martens S. Plant phenolics: recent advances on their biosynthesis, genetics, and ecophysiology. Plant Physiol Biochem 2013;72:1–20.

48. Lattanzio V, Lattanzio VMT, Cardinali A. Role of phenolics in the resistance mechanisms of plants against fungal pathogens and insects. In: Phytochem. Adv. Res. 2006. pp 23–45.

49. Dai J, Mumper RJ. Plant phenolics: extraction, analysis and their antioxidant and anticancer properties. Molecules 2010;15:7313–7352.

50. Li X, Weng JK, Chapple C. Improvement of biomass through lignin modification. Plant J 2008;54:569–581.

51. Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W. Lignin biosynthesis and structure. Plant Physiol 2010;153:895–905.

52. Abreu HS, Latorraca JVF, Pereira RPW, Monteiro MBO, Abreu FA, Amparado KF. A supramolecular proposal of lignin structure and its relation with the wood properties. An Acad Bras Cienc 2009;81:137–142.

53. Koutaniemi S. Lignin biosynthesis in Norway spruce: from a model system to the tree. 2007. PhD thesis University of Helsinki.

54. Liu CJ. Deciphering the enigma of lignification: precursor transport, oxidation, and the topochemistry of lignin assembly. Mol Plant 2012;5:304–317.

55. Hao Z, Mohnen D. A review of xylan and lignin biosynthesis: foundation for studying Arabidopsis irregular xylem mutants with pleiotropic phenotypes. Crit Rev Biochem Mol Biol 2014;49:212–241.

56. Boerjan W, Ralph J, Baucher M. Lignin biosynthesis. Annu Rev Plant Biol 2003;54:519–546.

57. Zhong R, Ye Z-H. Secondary cell walls: biosynthesis, patterned deposition and transcriptional regulation. Plant Cell Physiol 2015;56:195–214.

58. Boija E, Johansson G. Interactions between model membranes and lignin-related compounds studied by immobilized liposome chromatography. Biochim Biophys Acta - Biomembr 2006;1758:620–626

59. Kaneda M, Rensing KH, Wong JCT, Banno B, Mansfield SD, Samuels AL. Tracking Monolignols during Wood Development in Lodgepole Pine. Plant

Physiol 2008;147:1750–1760.

60. Ehlting J, Mattheus N, Aeschliman DS, et al. Global transcript profiling of primary stems from Arabidopsis thaliana identifies candidate genes for missing links in lignin biosynthesis and transcriptional regulators of fiber differentiation. Plant J 2005;42:618–640.

61. Kang J, Park J, Choi H, Burla B, Kretzschmar T, Lee Y, Martinoia E. Plant ABC Transporters. Arabidopsis Book 2011;9:e0153.

62. Saha J, Sengupta A, Gupta K, Gupta B. Molecular phylogenetic study and expression analysis of ATP-binding cassette transporter gene family in Oryza sativa in response to salt stress. Comput Biol Chem 2015;54:18–32.

63. Sibout R, Höfte H. Plant Cell Biology: The ABC of Monolignol Transport. Curr Biol 2012;22:533–535.

64. Alejandro S, Lee Y, Tohge T, et al. AtABCG29 is a monolignol transporter involved in lignin biosynthesis. Curr Biol 2012;22:1207–1212.

65. Hatfield R, Vermerris W. Lignin formation in plants. The dilemma of linkage specificity. Plant Physiol 2001;126:1351–1357.

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

67. He L, Terashima N. Formation and structure of lignin in monocotyledons IV. Deposition process and structural diversity of the lignin in the cell wall of sugarcane and rice plant studied by ultraviolet microscopic spectroscopy. Holzforsch - Int J Biol Chem Phys Technol Wood 1991;45:191.

68. de O. Buanafina MM. Feruloylation in grasses: current and future perspectives. Mol Plant 2009;2:861–872.

69. Bunzel M, Ralph J, Brüning P, Steinhart H. Structural identification of dehydrotriferulic and dehydrotetraferulic acids isolated from insoluble maize bran fiber. J Agric Food Chem 2006;54:6409–6418.

70. Jacquet G, Pollet B, Lapierre C, Mhamdi F, Rolando C. New ether-linked ferulic acid-coniferyl alcohol dimers identified in grass straws. J Agric Food Chem 1995;43:2746–2751.

71. Ralph J, Grabber JH, Hatfield RD. Lignin-ferulate cross-links in grasses: active incorporation of ferulate polysaccharide esters into ryegrass lignins. Carbohydr Res 1995;275:167–178.

72. Wilkerson CG, Mansfield SD, Lu F, et al. Monolignol ferulate transferase introduces chemically labile linkages into the lignin backbone. Science 2014;344:90–93.

73. Harholt J, Suttangkakul A, Vibe Scheller H. Biosynthesis of Pectin. Plant Physiol 2010;153:384–395.

74. Park YB, Cosgrove DJ. A revised architecture of primary cell walls based on biomechanical changes induced by substrate-specific endoglucanases. Plant Physiol 2012;158:1933–1943.

75. Bootten TJ, Harris PJ, Melton LD, Newman RH. Solid-state 13C-NMR spectroscopy shows that the xyloglucans in the primary cell walls of mung bean (Vigna radiata L.) occur in different domains: a new model for xyloglucan-cellulose interactions in the cell wall. J Exp Bot 2004;55:571–583.

76. Martin AP, Palmer WM, Brown C, Abel C, Lunn JE, Furbank RT, Grof CPL. A developing Setaria viridis internode: an experimental system for the study of biomass generation in a C4 model species. Biotechnol Biofuels 2016;9:45.

77. Hirano K, Aya K, Morinaka Y, Nagamatsu S, Sato Y, Antonio BA, Namiki N, Nagamura Y, Matsuoka M. Survey of genes involved in rice secondary cell wall formation through a co-expression network. Plant Cell Physiol 2013;54:1803–1821.

78. Bosch M, Mayer CD, Cookson A, Donnison IS. Identification of genes involved in cell wall biogenesis in grasses by differential gene expression profiling of elongating and non-elongating maize internodes. J Exp Bot 2011;62:3545–3561.

79. Casu RE, Jarmey JM, Bonnett GD, Manners JM. Identification of transcripts associated with cell wall metabolism and development in the stem of sugarcane by Affymetrix GeneChip Sugarcane Genome Array expression profiling. Funct Integr Genomics 2007;7:153–167.

80. Hansen SF, Harholt J, Oikawa A, Scheller H V. Plant glycosyltransferases beyond CAZy: A perspective on DUF families. Front Plant Sci 2012;3:59.

81. Rennie EA, Scheller HV. Xylan biosynthesis. Curr Opin Biotechnol 2014;26:100–107.

82. Xu J, Li Y, Ma X, Ding J, Wang K, Wang S, Tian Y, Zhang H, Zhu XG. Whole transcriptome analysis using next-generation sequencing of model species Setaria viridis to support C4 photosynthesis research. Plant Mol Biol 2013;83:77–87.

83. Torres AF, Visser RGF, Trindade LM. Bioethanol from maize cell walls: genes, molecular tools, and breeding prospects. GCB Bioenergy 2015;7:591–607.

84. Martin AP, Palmer WM, Byrt CS, Furbank RT, Grof CP. A holistic high-throughput screening framework for biofuel feedstock assessment that characterises variations in soluble sugars and cell wall composition in Sorghum bicolor. Biotechnol Biofuels 2013;6:186.

85. Larkin P (2011) Infrared and Raman Spectroscopy; Principles and Spectral Interpretation, 1st Edition | ISBN 9780123869845.

86. Morikawa H, Senda M. Infrared analysis of oat coleoptile cell walls and oriented structure of matrix polysaccharides in the walls. Plant Cell Physiol

1978;19:327–336.

87. Mouille G, Robin S, Lecomte M, Pagant S, Hofte H. Classification and identification of Arabidopsis cell wall mutants using Fourier-Transform InfraRed (FT-IR) microspectroscopy. Plant J 2003;35:393–404.

88. Carpita NC, Defernez M, Findlay K, Wells B, Shoue DA, Catchpole G, Wilson RH, McCann MC. Cell wall architecture of the elongating maize coleoptile. Plant Physiol 2001;127:551–565.

89. Devos KM. Updating the “crop circle.” Curr Opin Plant Biol 2005;8:155–162.

90. Sattler SE, Funnell-Harris DL. Modifying lignin to improve bioenergy feedstocks: strengthening the barrier against pathogens?. Front Plant Sci 2013;4. doi: 10.3389/fpls.2013.00070.

2.3 Literature review

2.3.1. Building biomolecules

The composition of life at an atomic level consists of six key elements (carbon 34 %, oxygen 21.9 %, hydrogen 43.5 %, nitrogen 0.3 %, phosphorous 0.05 % and sulphur 0.02 %) and 19 trace elements that are considered essential. These atomic ratios are sourced from the bioenergy grass *Miscanthus* [6], however, elemental ratios are consistent in all living organisms with >99% of all life being composed of C, O, H and N [7]. Whilst these simple elements are the atomic constituents of life, they are the building blocks of a great diversity of carbohydrates, lipids, proteins, nucleic acids and other biomolecules which act in concert to produce the great diversity of life that surrounds us. Importantly it is the formation of biomolecules by autotrophic organisms, of which plants are the most prolific, that provides the entry point of energy into living systems. An understanding of this process (photosynthesis), the biomolecules it produces, how to best utilise them as a food source for humans and animal livestock and as a source of energy to fuel our industrialised world will be integral in meeting the demands of an over populated world.

2.3.1.1 An introduction to C_4 plants

The plant kingdom is natures energy sector, converting sunlight, H_2O absorbed from the soil, and CO_2 from the atmosphere into stored chemical energy, initially in the form of sugars. These sugars, either directly, or indirectly, fuel all other kingdoms of life, with the exception of small groups of bacterial autotrophs.

The plant kingdom can be divided into mosses, which lack vascular structure, ferns, which reproduce via spores, gymnosperms, which reproduce via 'naked' seeds, and angiosperms, the flowering plants. Angiosperms constitute the majority of plant life on earth and can be divided into class *Dicotyledoneae* (dicots) and class *Monocotyledoneae* (monocots). Most major agricultural crops are monocot grasses including wheat, rice, corn, barley, oats, sorghum and millet, thus this class of plant

is the energy producing backbone upon which human evolution, the birth of civilisations and our over-populated globe relies.

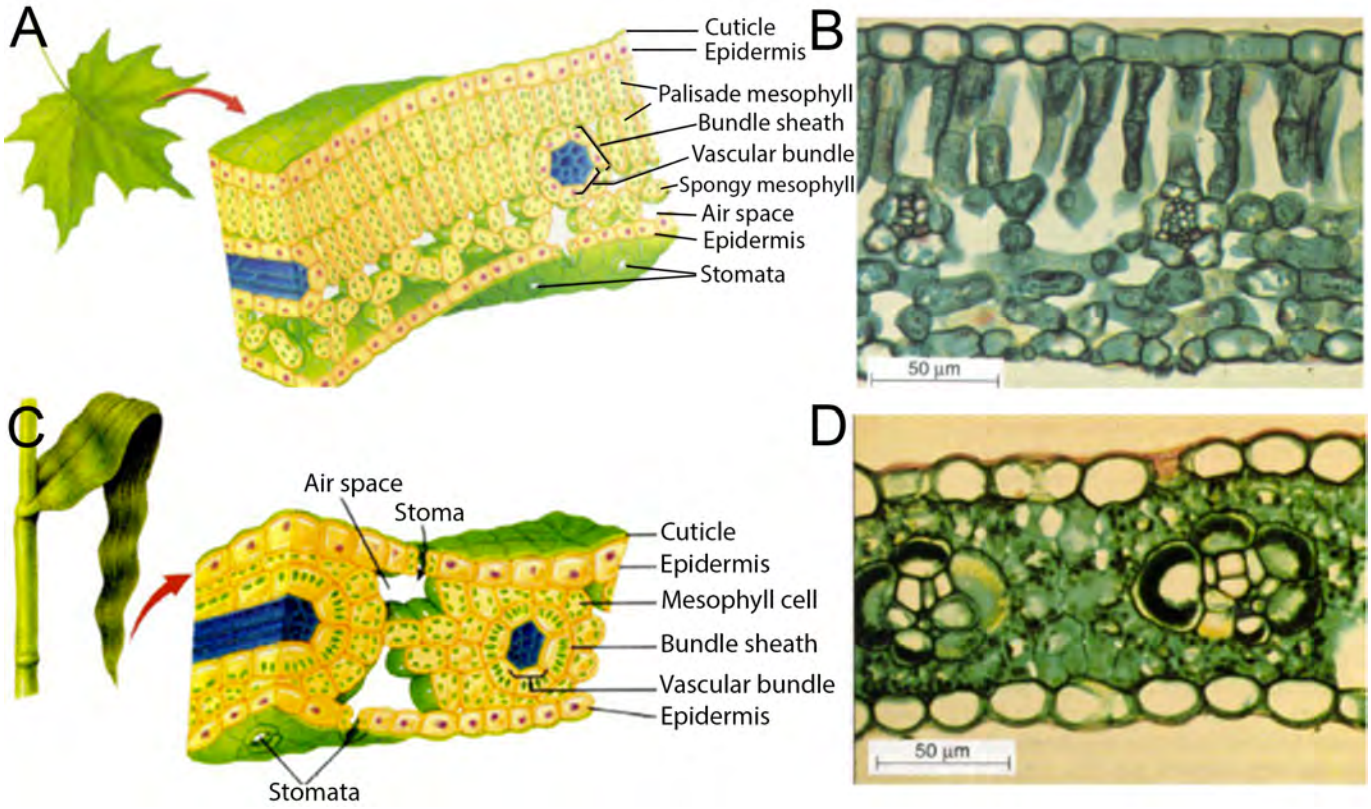


Figure 1. Representative leaf structure of a C_3 dicot (A) with its cross section (B), and a C_4 monocot (C) with its cross section (D) highlighting the differences in both vascular and chloroplast arrangement (adapted from [8]).

Monocot grasses first diverged from a common ancestor with dicots ~200 mya [4]. Importantly, grass species were the first to evolve the more efficient C_4 photosynthesis pathway approximately 24-35 mya [9] allowing these plants to convert sunlight into biomass more efficiently and allowing them to grow in harsher environments making many of these species ideal crops for domestication.

The C_3 photosynthesis pathway fixes CO_2 in chloroplasts of mesophyll cells via the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) [10]. This is an inefficient process since Rubisco competitively binds O_2 that is produced as a by-product of the light reactions of photosynthesis that also occurs in chloroplasts of mesophyll cells. Numerous plant species, of which the grasses were the first [9], have evolved molecular machinery and leaf anatomy that facilitates more efficient

CO_2 fixation by Rubisco [11] (see **Figure 1** for C_3-C_4 leaf anatomy). Briefly, CO_2 that diffuses into mesophyll cells through stomatal openings at the leaf surface is combined with H_2O by the enzyme carbonic anhydrase to form carbonic acid [11] (**Figure 2A**). Carbonic acid is then combined with phosphoenolpyruvate in the cytoplasm of mesophyll cells by the enzyme phospho*enol* pyruvate carboxylase (PEPC) to form the 4-carbon molecule oxaloacetate [11] (**Figure 2B**). Oxyloacetate is translocated to the mesophyll chloroplast where it is reduced to the C_4 acid malate, using the reducing power of NADPH that is generated by the light reactions of photosynthesis [12] (**Figure 2C**). Mesophyll chloroplasts contain granal stacks of thylakoids, which generate the NADPH by harvesting light through photosystems I and II, and these are linked together by lamellar thylakoids which contain ATP synthase that utilise proton gradients to produce ATP [13]. Malate is then translocated out of the mesophyll chloroplasts, through plasmodesmata into neighbouring bundle sheath cells, and into specialised bundle sheath chloroplasts where Rubisco is present [12] (**Figure 2D**). Malate is decarboxylated in the chloroplast stroma releasing CO_2 to effectively concentrate it around Rubisco [11] (**figure 2E**). Bundle sheath chloroplasts do not contain granal stacks of thylakoids and are specialised to harvest light for ATP production only [12, 13]. This ATP, and NADPH generated by the reducing power of malate, drives the Calvin cycle to produce an excess of 3 carbon sugars (3-phosphoglyceric acid; 3-PGA and glyceraldehyde-3-phosphate; G3P) as CO_2 is fixed by Rubisco [13]. The 3-PGA is translocated back to the mesophyll chloroplasts (**Figure 2F**) where it uses NADPH and ATP produced from the light reactions to generate G3P before being translocated back to the bundle sheath chloroplasts [11] (**Figure 2G**). These 3 carbon G3P sugars are used to synthesise glucose-1-phosphate, glucose-6-phosphate and/or fructose-6-phophate [14] (**Figure 2H**) which are the building blocks of complex carbohydrates such as starch, cellulose, sucrose and all other plant biomolecules. By separating the light reactions of photosynthesis between mesophyll and bundle sheath chloroplasts, O_2 is only produced in mesophyll chloroplasts thus eradicating competitive binding of Rubisco by O_2 and in combination with increased CO_2 concentrations this results in highly efficient CO_2 fixation and production of sugars [11, 12, 14]. This C_4 photosynthesis mechanism is

the most utilised mechanism for introducing energy into biomolecules in domesticated bioenergy crops.

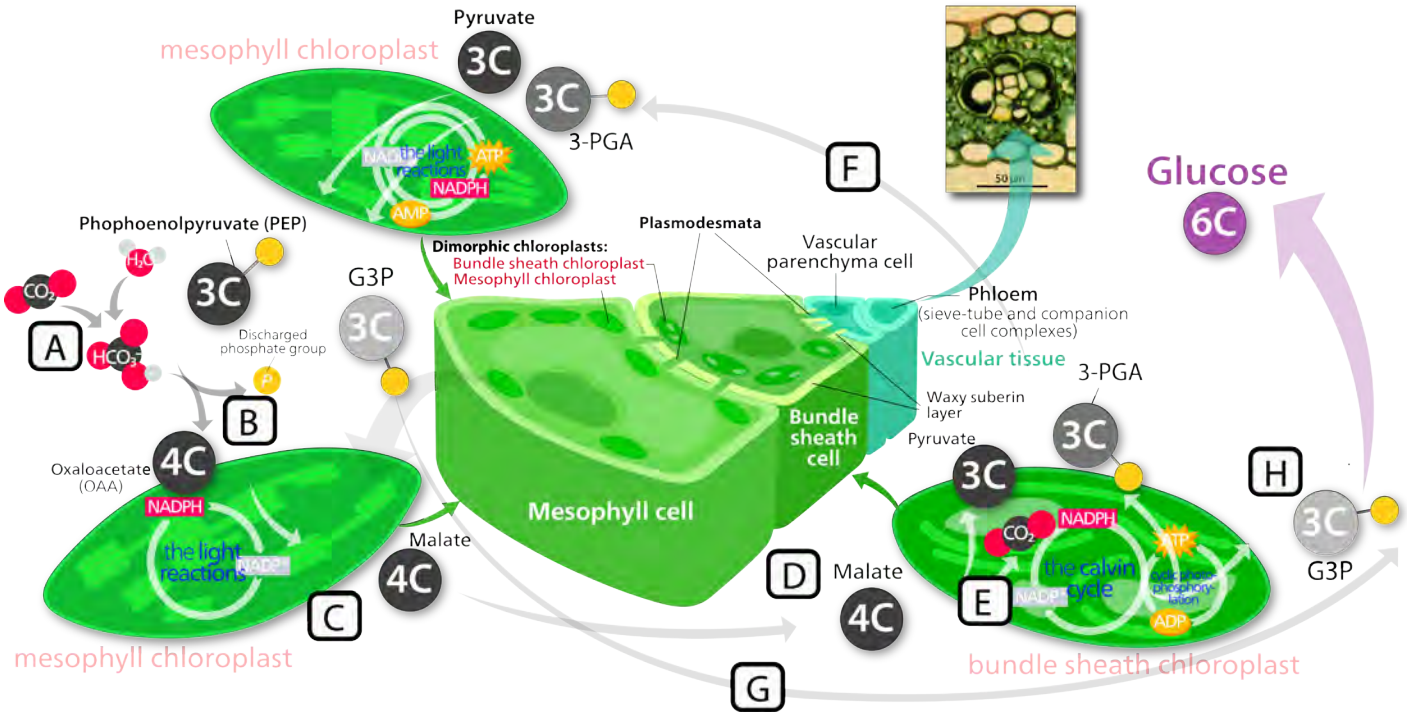


Figure 2. An overview of the complex C_4 NADP-ME photosynthesis pathway (utilised by Sorghum, Maize, Sugarcane, and Setaria). A-F are described in-text. 3-PGA; 3-Phosphoglyceric acid, G3P; Glyceraldehyde-3-phosphate. Cell cartoons were downloaded under a creative commons license (https://creativecommons.org/licenses/by-sa/3.0/) and annotation details were added.

Whilst C_4 photosynthesis has evolved separately over 45 times in 19 families of angiosperms including both dicots and monocots [9] (**Figure 3**), monocots comprise 79% (~5900 species) and monocotyledonous *Poaceae* (true grasses) family comprise 61% (~4600 species) of all C_4 species [15]. This is likely why they are among the most efficient biomass producing plants on earth, rapidly and efficiently accumulating cell wall material, starch and soluble sugars with minimal nutrient and water input. C_4 grass crop species such as *Sorghum bicolor*, *Zea mays*, *Saccharum officinarum* and *Panicum virgatum* are, therefore, ideal candidates for biofuel feedstocks. *S. bicolor* has been shown to be the highest yielding of these when relating a combination of cell wall and soluble sugar derived ethanol yields to fertiliser, water and economic inputs [1]. These crops have been domesticated and selectively bred over thousands of years to optimise yields of their food bearing organs. However, quality, high yielding biofuel feedstocks will possess

different characteristics namely, large biomass, high concentrations of soluble sugars and cellulose rich cell walls that are less recalcitrant to hydrolysis. Since the need for bioenergy crops has only emerged in the last decade there is a great need to explore the genetic potential of these crops.

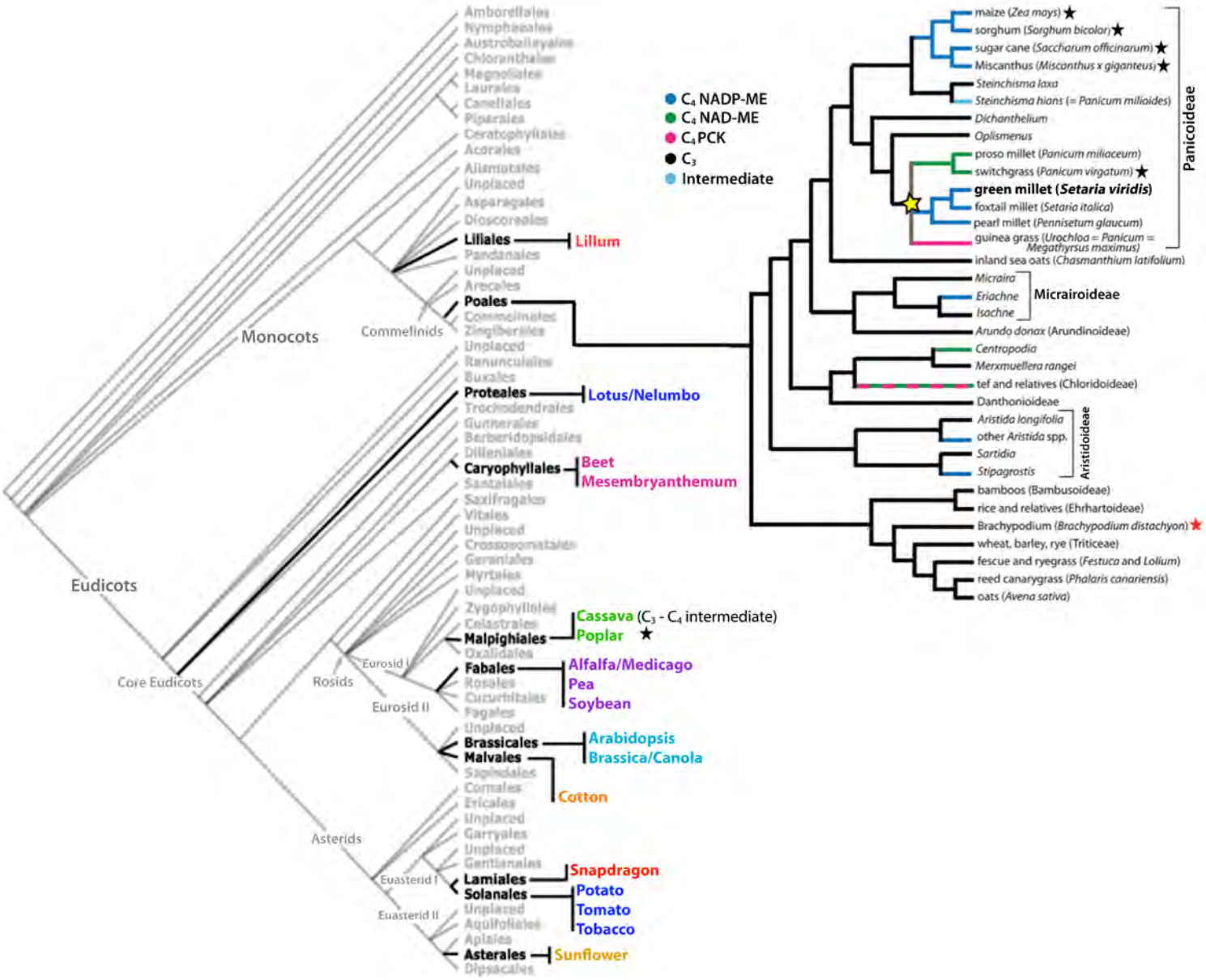


Figure 3. Phylogenetic tree of all plant taxonomic families. Model species within each family are listed and the *Poales* family is expanded to show the graminoids including C_4 grasses suggested as biofuel feedstock crops. Black stars indicate species suggested as biofuel feedstocks. A red star highlights the C_3 model grass species *Brachypodium distachyon* and bold type highlights the C_4 model species *Setaria viridis*. Coloured phylogenetic arms indicate species possessing C_4 photosynthesis and their subtype (C_4 NADP-ME, C_4 NAD-ME, C_4 PCK and intermediates) as indicated in the legend. Adapted from [16] and [17].

2.3.1.2 Sugar partitioning and storage

Consuming CO_2, H_2O and energy from sunlight, the net products of the light and dark reactions of photosynthesis are O_2 and the sugar phosphates glucose-1-phosphate, glucose-6-phosphate and/or fructose-6-phosphate (each formed by joining two G3P molecules) [10, 14]. Whilst these sugars are the building blocks of all organic plant biomolecules, the plant initially uses them either a foliar energy store in the form of soluble sugars and starch or to transport to growing sink tissues in the form of sucrose.

Starch molecules can be found in many plant tissues, and accumulates during daylight in leaves to be used for growth at night when sunlight is not available. In seeds it is an energy store which is used upon germination until the first photosynthetically active leaves are formed [10]. Glucose-1-phosphate molecules are converted to ADP-glucose, consuming energy in the form of ATP, which is then polymerised via α-1,4 glycosidic bonds, to form the linear polymer amylose [14]. α-1,6 branch points are then introduced to form the branched glucose polymer, amylopectin [10] (see **Figure 5** for starch chemical structures). Starch in plants generally consists of ~20-25 % amylose and 75-80% amylopectin [18]. Starch is insoluble in water [18] thus is osmotically inactive, unlike glucose, therefore it is found in chloroplasts as precipitated starch granules. This, among other advantages, provides an energy store that does not affect the osmotic components of the translocation of sugars from source leaves to growing sink tissues [10]. Alpha bonds linking glucose molecules are easily broken by amylase enzymes which allows the plant to easily access glucose when it is required [10] (i.e. for growth in the absence of sunlight). Amylases are abundant in human saliva [14] making the starch in grains such as maize, wheat, and rice an ideal food source for humans. Recently, however, starch from maize kernels has been used extensively as a source of glucose to fuel fermentation for bioethanol production in the United States [19]. This, however, generates competition for maize as a food source and drives food prices up igniting the food vs. fuel debate [20].

The second major use of the sugars produced by the light and dark reactions of photosynthesis is the production of sucrose for transport to plant sink tissues where carbon is required for growth, storage in fruits and grains etc., or response to external stimuli such as pathogen attack [10]. To synthesise sucrose, G3P molecules are translocated from the chloroplast stroma to the cytoplasm via the triose phosphate translocator. From G3P, fructose-6-phosphate and UDP-glucose are synthesised, combined and dephosphorylated to form the disaccharide, sucrose [10]. Since sucrose is a 12-carbon sugar, composed of a glucose and fructose molecule (see **Figure 5** for chemical structure), it is a more osmotically efficient transport molecule than glucose or fructose.

Sugar transport and carbon partitioning in higher plants involves complex interactions between active and passive transport across membranes, osmotic gradients and hydrostatic pressures of xylem and phloem transport networks. A detailed description of these processes is beyond the scope of this review however, it is important to understand how plant tissues accumulate biomolecules such as sucrose stores in *S. bicolor* stems and how sugar molecules are delivered to growing plant tissues for incorporation into cell wall polymers such as cellulose. For recent reviews see [21-25].

Once sucrose reaches sink tissues it can be:

- Directly stored, for example in sugarcane and sorghum stems
- Hydrolysed by invertase enzymes to form glucose and fructose, which are often the major storage sugars of commercial importance in fruits (since fructose gives a sweeter taste than sucrose)
- Hydrolysed by the enzyme sucrose synthase (SuSy) to produce UDP-glucose which is required to synthesise many of the biomolecules essential for plant growth including many cell wall polymers such as cellulose and hemicelluloses which provide structural support to a growing plant and also act as energy stores.

2.3.1.3 Plant cell walls

The plant cell wall is a complex, highly specialized structure that encases every plant cell and is essential to all higher plants. It provides mechanical strength to plant structures, determines cell shape and is involved in fundamental cellular processes such as cell expansion and cell division. Furthermore, it helps regulate nutrient partitioning and provides a barrier to pathogen attack, gaseous diffusion and macromolecule transport [10].

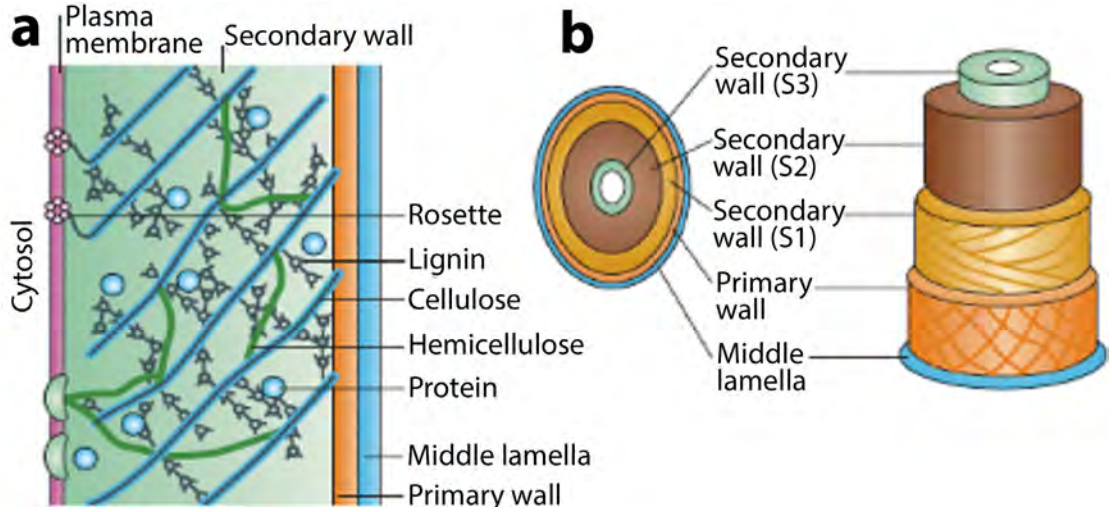


Figure 4. Generic cell wall structure (adapted from [26]).

A typical plant cell maintains a primary cell wall throughout cell expansion followed by the deposition of a secondary cell wall, often in three distinct layers (S1, S2 & S3), which is usually thickened by lignification and/or suberisation [26] (**Figure 4**). Primary and secondary cell walls therefore differ greatly in composition and their structure and deposition must be considered separately. Grass cell walls (type II) are also distinctly different in composition from dicot cell walls (type I) and thus must also be considered separately.

The biomolecules that make up the plant cell wall are a rich source of fermentable sugars, particularly glucose, however, these molecules are generally β linked polymers which are much more recalcitrant to degradation than the α linked polymers of starch and the disaccharide bonds of sucrose. Plant cell walls are therefore not currently economically viable as a source of fermentable sugars for

bioethanol production. An in-depth understanding of the synthesis, structure and composition of the cell wall as a whole and the molecular switches that signal the transition from primary to secondary cell wall synthesis will be integral to engineering crops with higher ratios of fermentable, six carbon sugars and a structure that lends itself to more efficient deconstruction.

2.3.1.3.1 Cell wall biomolecules of grasses

Cell wall biomolecules are produced from a large pool of nucleotide sugars synthesised from sucrose that is transported to growing cells. Sucrose is hydrolysed by SuSy to generate UDP-glucose which can be interconverted to a range of five and six carbon nucleotide sugars through a series of complex enzymatic pathways [27]. Nucleotide sugar interconversions and transport to sites where they can be used to synthesise cell wall polymers are also vitally important processes in cell wall synthesis, however they are beyond the scope of this report and have recently been reviewed [27].

The cell wall biomolecules produced from these nucleotide sugars are generally long chain polymers that are arranged in a complex matrix with the cellulose, hemicellulose and phenolic fractions highly integrated (see **Figure 5** for an overview of the major polysaccharides that make up the cell wall and their nucleotide sugar substrate). In grasses, ~85% of the primary cell wall is composed of (1,4)-β-glucans (cellulose), and the hemicelluloses, glucurono(arabino)xylan (GAX) and xyloglucan (XyG), and (1,3;1,4)-β-glucans (mixed linkage glucans; MLG) [28]. The remainder is composed of small amounts of pectins, structural proteins, unpolymerised phenolics and silica [28]. Secondary cell wall composition is lignified to provide structural rigidity and therefore its main components are cellulose, GAX, silica and the phenolic polymer, lignin, which together comprise ~95% of secondary cell wall dry weight [28]. It also contains small amounts of pectins and unpolymerised phenolics such as ferulic acid and ρ-coumaric acid [28] (**table 1**).

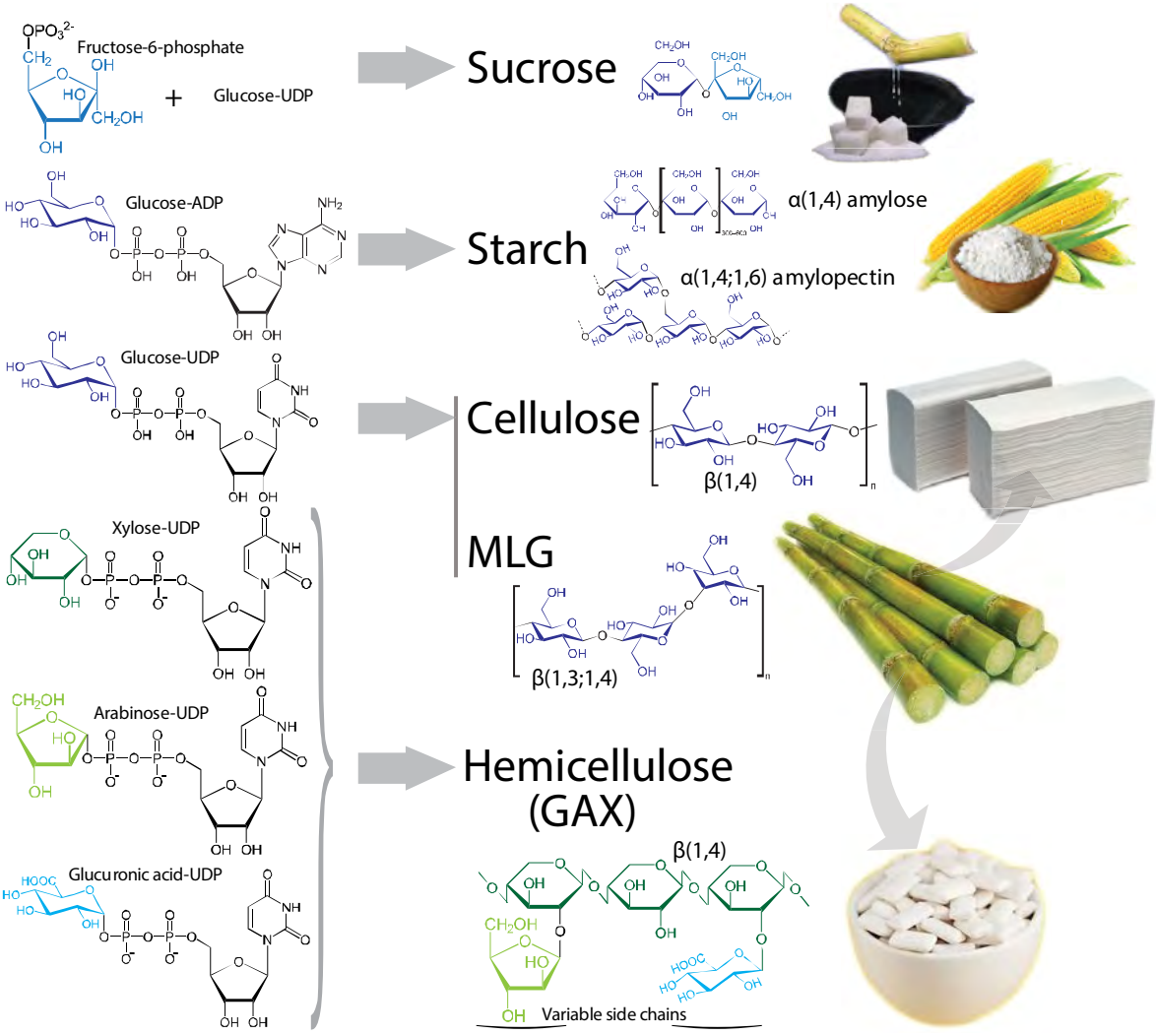


Figure 5. Major nucleotide sugar substrates required to build the major storage biomolecules and cell wall polysaccharides of grasses. MLG; β(1,3;1,4)-glucans or mixed linkage glucans, GAX; glucurono(arabino)xylan. Six carbon monomers are coloured in shades of blue and five carbon monomers are coloured in shades of green whilst the nucleotide functional groups are black. An image of the natural product of each polymer is displayed.

2.3.1.3.1.1 Cellulose

Cellulose microfibrils are universally present in almost all cell walls in high proportions making cellulose the most abundant polymer on earth. Cellulose is an un-branched, β-1,4 linked polymer chain of D-glucose, successively inverted 180° and ranging in size from 2,000 – 25,000 glucose residues [29, 30]. The polymerisation of cellulose is catalysed by cellulose synthase (CesA) enzymes from the large cellulose synthase-like (Csl) gene family [30]. Immunogold labelling of CesA enzymes has revealed that CesAs, are embedded in the plasma membrane

and form a particle rosette [31] consisting of six rosette subunits which are each composed of six CesA enzymes [32, 33] (**Figure 6**). The eight highly conserved transmembrane domains of the CesA gene family defines the barrel shape of the rosette and the cytosolic orientation of the enzymes active site [29] so that glucan chains are produced and secreted from the centre of the barrel into the extracellular space [29, 30, 33]. This cellulose synthase enzyme complex (CSC) produces multiple glucan chains that weave together like a piece of string, via hydrogen bonds, to produce cellulose microfibrils [29] typically consisting of 18 or 24 interwoven glucan chains [29, 30, 34] (**Figure 6**). The CSC uses UDP-glucose as a substrate for cellulose synthesis, in comparison to starch synthesis which uses ADP-glucose, and is guided through the plasma membrane by cytosolic microtubules which ensure that the orientation of cellulose microfibrils is organised [35]. This facilitates anisotropic cell wall expansion perpendicular to cellulose microfibril orientation and maximises strength in secondary cell walls [10].

Table 1. Primary and secondary cell wall composition (% dry weight) of grasses and genes that have been implicated in the biosynthesis of each component in grasses

	Primary cell wall		Secondary cell wall	
	%	Genes	%	Genes
Cellulose	20-30	CesA1-9, KOR A3 & B5, BC1L4	35-45	CesA10-12, KOR, COBRA
Hemiellulose				
MLG	10-30	CslF & CslH families	minor	
Xylans	20-40	GT43, 47, 61, 75	40-50	GT43, 47, 61, 75
XyG	1-5		minor	
Pectin	5		0.1	
Structural proteins	1		minor	
Phenolics				
Ferulic acid & ρ-coumaric acid	1-5	PAL pathway **(Figure 7)**	0.5-1.5	PAL pathway **(Figure 7)**
Lignin	minor		20	PAL pathway **(Figure 7)**
Silica	?		5-15	

Ten CesA genes have been identified in *A. thaliana* [36], 10 in rice [37], 12 in sorghum, 14 in maize and 12 in *S. viridis* all with high homology [38] suggesting

that the CSC is highly conserved among all land plants. It is has been shown by mutant analysis that there are at least three non-redundant CesAs that form a functional CSC [37]. In rice, mutant analysis has suggested OsCesA 4, 7 and 9 are required for secondary cell wall synthesis [37]. Expression analysis in maize has suggested that ZmCesA1-9 are involved in primary cell wall deposition whilst ZmCesA10-12 are involved in secondary cell wall cellulose deposition [39]. In *A. thaliana*, homologous CesAs have been shown to be co-expressed in a range of tissues and have protein-protein interactions [36]. This suggests that three CesAs may form a protein complex that forms half of the rosette, which is known to contain six subunits, and that combinations of CesAs (in lots of three) are tightly transcriptionally co-regulated in specific tissues corresponding to primary and secondary cell wall synthesis (and possibly different plant tissues).

Mutational analysis of *A. thaliana* lines has implicated the KOBITO (unknown function), KOR (a membrane localised β-1,4 glucanase), COBRA (COB - has been shown to act as a chaperone to facilitate cellulose crystallisation in the β-1,4 glucan polymers emerging from the CesA rosette [40, 41]) and POM/CSI (chitinase-like protein that encode β-1,3 glucanases) genes in cellulose synthesis, however, their exact mechanisms remain elusive [35, 36, 42, 43]. Mutations affecting primary cell wall cellulose synthesis tend to result in swelled cells, most likely due to isotropic cell expansion and inability to regulate cell turgor in an expanding cell [44]. In grass species, barley, maize and rice, mutations in secondary cell wall cellulose synthesis genes consistently result in a brittle stem phenotype [37, 45-47] and in *A. thaliana* collapsed xylem elements are observed [36].

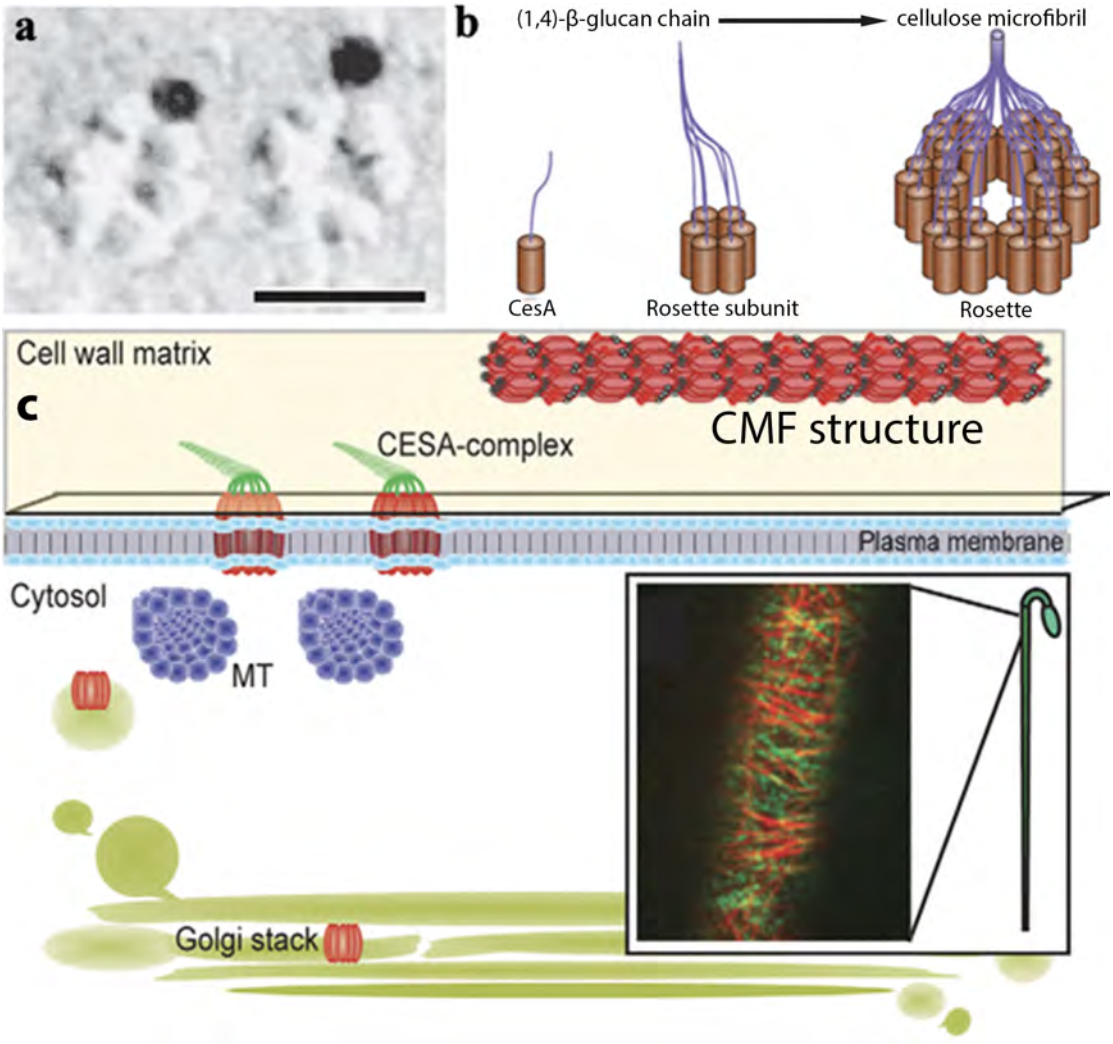


Figure 6. Overview of cellulose synthesis (a) immunogold labelling of CesA rosettes embedded in the plasma membrane (b) diagram of particle rosette structure and mechanism of CMF synthesis and (c) cellular overview of cellulose synthesis (adapted from [33] and [35]).

2.3.1.3.1.2 Hemicelluloses

Hemicelluloses are named as such because they are polymers of β linked sugars, often with a glucan backbone, making them chemically similar to cellulose and also a rich source of fermentable sugars. Hemicelluloses however, are often composed of a mixture of six and five carbon sugars and their backbone is often highly substituted with monosaccharide or disaccharide side chains [36]. Whilst most work on hemicellulose structure and function within the cell wall has been done on dicot species (i.e. *A. thaliana*), in grasses it is generally accepted that hemicelluloses coat cellulose microfibrils, linking them together via hydrogen

bonds and forming the cellulose-hemicellulose network that confers mechanical strength to a cell wall.

2.3.1.3.1.2.1 Xylans

Glucurono(arabino)xylan (GAX) is a polymer of β-1,4 linked D-xylose (five carbon sugar) with variable non-repeating arabinose and glucuronic acid side chains [48]. It is the predominant hemicellulose in grass cell walls [28] (**Figure 5**).

Synthesis of xylans occurs in the golgi using UDP-xylose as the substrate for the xylan backbone, however, the enzymatic processes involved are only recently being revealed [49]. The xylose backbone of GAX is highly substituted with arabinose and glucuronic acid side chains during synthesis in the golgi, however, these side chains are removed upon excretion to the wall and ferulic acid is often linked to the remaining arbinose side chains [28]. This method of synthesis has implications for cell wall structure. Since GAX is the predominant hemicellulose, rather than a cellulose-XyG network linking CMFs together, as in dicots, there is a cellulose-GAX network. The GAX is thought to bind to CMFs via hydrogen bonds forming a chain fence like mesh which is also held together by hydrogen bonds (**Figure 5**) [50]. The arabinose and xylose side chains are usually attached at the O-2 and O-3 position [28] which blocks efficient hydrogen bonding to CMFs and other GAX [51, 52] and this is thought to be the basis for cell wall expansion in type II primary cell walls.

Xylan biosynthesis requires initiation, elongation and termination of the xylan backbone and, in the case of GAX, addition of arabinose and glucuronic acid side chains [36]. In *A. thaliana* GUX1 and GUX2 genes have been identified that localise to the golgi and when knocked down result in reduced addition of side chains to xylan backbones but without a reduction in xylan backbone synthesis [53]. This indicates that xylan backbone synthesis occurs via an independent mechanism to side chain substitution. Recently a number of glycosyltransferases (GT) families have been implicated in both xylan backbone synthesis and side chain substitution of xylans in grasses [54]. The glycosyltransferase 43 and 47 (GT43 and 47) gene families have been implicated in xylan backbone synthesis [55]. The GT61 gene

family, although its precise role is still unclear, has been implicated in side chain addition of both arabinose and glucuronic acid, and the GT75 gene family has been implicated in side chain addition of glucuronic acid [54, 55]. Mostly these implications are based on heterologous expression in *A. thaliana*, rather than in a grass species.

Most yeast cannot efficiently ferment five carbon sugars such as xylose, therefore grass cell wall xylans such as GAX, which comprise ~40% of grass cell walls [28], are not particularly useful in the production of bioethanol. Bioengineering of yeast strains that efficiently ferment five carbon sugars is currently in progress however [56], and will greatly increase the value of this cell wall fraction in grasses.

2.3.1.3.1.2.2 Mixed linkage glucans (MLG)

Mixed linkage glucans (MLG) are a polymer of glucose with mixed β-1,3 and β-1,4 glycosidic linkages making their structure very similar to cellulose with similar hydrogen bonding properties. This hydrogen bonding potential is because, unlike other hemicelluloses, they do not have any side chains [28]. The mixed linkages between glucose monomers result in a polymer very similar to cellulose but with regular 'kinks' or bends along the chain (**Figure 5**). They are specifically present in high proportions (10-30% cell wall dry weight) in expanding primary cell walls of grasses with maximum levels occurring at the peak of cell expansion [57, 58] while mature organs with secondary cell walls have very low amounts with measurements in mature *Sorghum bicolor* leaves and stem of 0.12 and 0.34 % (w/w) respectively [59]. This suggests a major role in cell expansion of grass primary cell walls for MLGs. It has been shown that MLGs tightly bind to the surface of CMFs in expanding cell walls [60]. I propose a novel mechanism by which CMFs expand by a 'ratchet' system where substituted GAX slides along 'kinks' in the MLG polymers providing resistance to recoiling. This 'ratchet' system would provide a mechanism for expansion that maintains cell wall strength during cell expansion. This requirement of expanding cell walls to maintain strength and allow expansion simultaneously is clearly observed in expanding leaf epidermal cells, which are also the major load bearing cells. These expanding leaf epidermal cell wall have been shown to contain higher levels of MLG in comparison to mesophyll cells [60].

Since MLGs are mostly specific to grasses, relatively little is known about the molecular mechanisms of their synthesis. Recently immunocytochemistry has revealed that MLG polymers are synthesised at the plasma membrane [61, 62]. Identification of a quantitative trait loci for MLG content in barley grain has also led to the discovery of six rice cellulose synthase-like F (CslF) genes in the syntenic region of the sequenced rice genome [63]. These CslF genes were heterologously expressed in *A. thaliana*, which naturally produces no MLG, and small amounts of MLG were detected [63]. In addition, the CslF family is one of three grass specific Csl gene families supporting the argument that CslF genes are responsible for biosynthesis of MLG, which is specific to grasses. Similarly, the CslH gene family that is also specific to grasses has been shown to produce small amounts of MLG when heterologously expressed in *A. thaliana* [64]. Expression studies have resulted in the suggestion that CslF genes are required in expanding primary cell walls, whilst CslH genes are required in seed endosperms or secondary cell walls where MLG is used as a carbon energy store [36]. Both gene families are thought to only introduce β-1,4 glycosidic bonds into MLGs and since heterologous expression in *A. thaliana* only produces minute quantities of MLG it is thought that another unidentified gene family is responsible for producing the β-1,3 linkages in the MLG polymer [36]. A recent study, however, has suggested that *Brachypodium distachion* CSLF6 is capable of producing both the β-1,4 and β-1,3 linkages [65].

MLGs are a rich source of the fermentable six carbon sugar glucose, therefore are of interest in bioethanol production, however, in mature *Sorghum bicolor* stem tissue they are only found in low amounts [59]. It has, however, been suggested that MLG is present in high proportions in mature rice stems although the only quantative data available reports levels of no greater than 5 % (w/w) of the acid insoluble residue (i.e. cell walls) [66, 67].

2.3.1.3.1.2.3 Xyloglucan (XyG)

Xyloglucan (XyG) is a polymer of β-1,4 linked D-glucose with variable xylose side chains substituted in repeating units of four glucose backbone monomers. It is

only present in small proportions in grass cell walls (**Table 1**) and is therefore only mentioned for completeness, however it is a major component of dicot cell walls.

2.3.1.3.1.2 Phenolics

Phenolics are a broad class of chemical compounds produced by plants and microorganisms. Their fundamental unit is an OH group bound to an aromatic hydrocarbon ring which is very different from the mono/polysaccharides which are based on long chain carbons joined by glycosidic bonds. They are produced from the phenylpropanoid pathway yielding a large variety of aromatic compounds that are generally responsible for the myriad of colours, odours and tastes produced by the plant kingdom. In the cell walls of grasses three significant phenolics have been observed. These are the hydroxycinnamates, ferulic acid and ρ-coumaric acid. The polyphenol lignin is a complex polymer that varies in monomer composition between species and cell types, and is also a major component of secondary cell walls.

2.3.1.3.1.2.1 Lignin

Lignin is the second most abundant biomolecule on earth after cellulose and presents the greatest hurdle to efficient release of fermentable sugars from cell wall polysaccharides. It is a complex aromatic polymer generated by irregular linkage of three main phenylalanine derived monomers (H, G & S lignin) [68-70]. The resultant polymer is highly branched due to a variety of possible linkages (β-O-4, β-β, β-5) between monomers. Varying linkage ratios and varying H, G and S lignin ratios are observed between different species, cell types and even within a cell wall [71]. H lignin is uniquely present in grass cell walls with a typical grass cell wall containing ~35-49% G units, 40-61% S units, and 4-15% H units [28]. Monomers polymerise through pores in the cellulose-hemicellulose network and covalently bind to hemicelluloses to rigidify cell walls and render them impermeable to water and solutes [72]. The ratio of S:G lignin has been implicated in defining cell wall characteristics since G lignin is more highly branched than S lignin and therefore has greater opportunity to cross-link to cell wall polysaccharides. A lower S:G ratio therefore, typically results in a more rigid and less digestible cell wall [73]. Lignin is an aromatic polymer thus can be used in the

production of a variety of aromatic compounds, however, unlike polysaccharides it does not contain fermentable sugars therefore it cannot be used as a fuel source for bioethanol production.

Lignin biosynthesis begins in the cytosol where the amino acid phenylalanine is deaminated by the enzyme phenylalanine ammonia lyase (PAL) to form cinnamic acid [70]. PAL is an important enzyme because its substrate, phenylalanine, is not only a precursor to lignin but is also a precursor to an array of phenylpropanoids such as flavanoids, stillbenes, coumarins and benzoic acid derivatives [68, 74, 75]. In angiosperms, PAL gene families are usually quite large with up to 40 members, suggesting gene regulation of phenylpropanoid biosynthesis via differential expression of these 40 isoforms [68]. Sorghum is thought, by homology, to contain 9 PAL genes [76] as does *S. viridis* (unpublished data).

PAL has been shown to associate with cinnamate 4-hydroxylase (C4H) in microsomal membranes [77-79]. Thus it is thought that the substrate of PAL, phenylalanine, is 'channelled' through a membrane complex [79-83] consisting of a series of enzymes which modify it to form the monolignols, ρ-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol [70, 84, 85]. These act as the building blocks for ρ-hydroxyphenyl (H), guaiacyl (G) & syringyl (S) lignin, respectively [68, 70]. ρ-coumaroyl CoA is an important substrate because the subsequent pathway flux can be directed, depending on the enzyme catalyst, either towards monolignols, catalysed by hydroxycinnamoyl CoA shikimate:quinate hydroxycinnamoyl transferase (HCT), or towards flavanoids, catalysed by chalcone synthase (CHS) [86] (see **Figure 7A** for an overview of the lignin biosynthesis pathway and **Figure 7B** for an example of a membrane complex specific for monolignol synthesis).

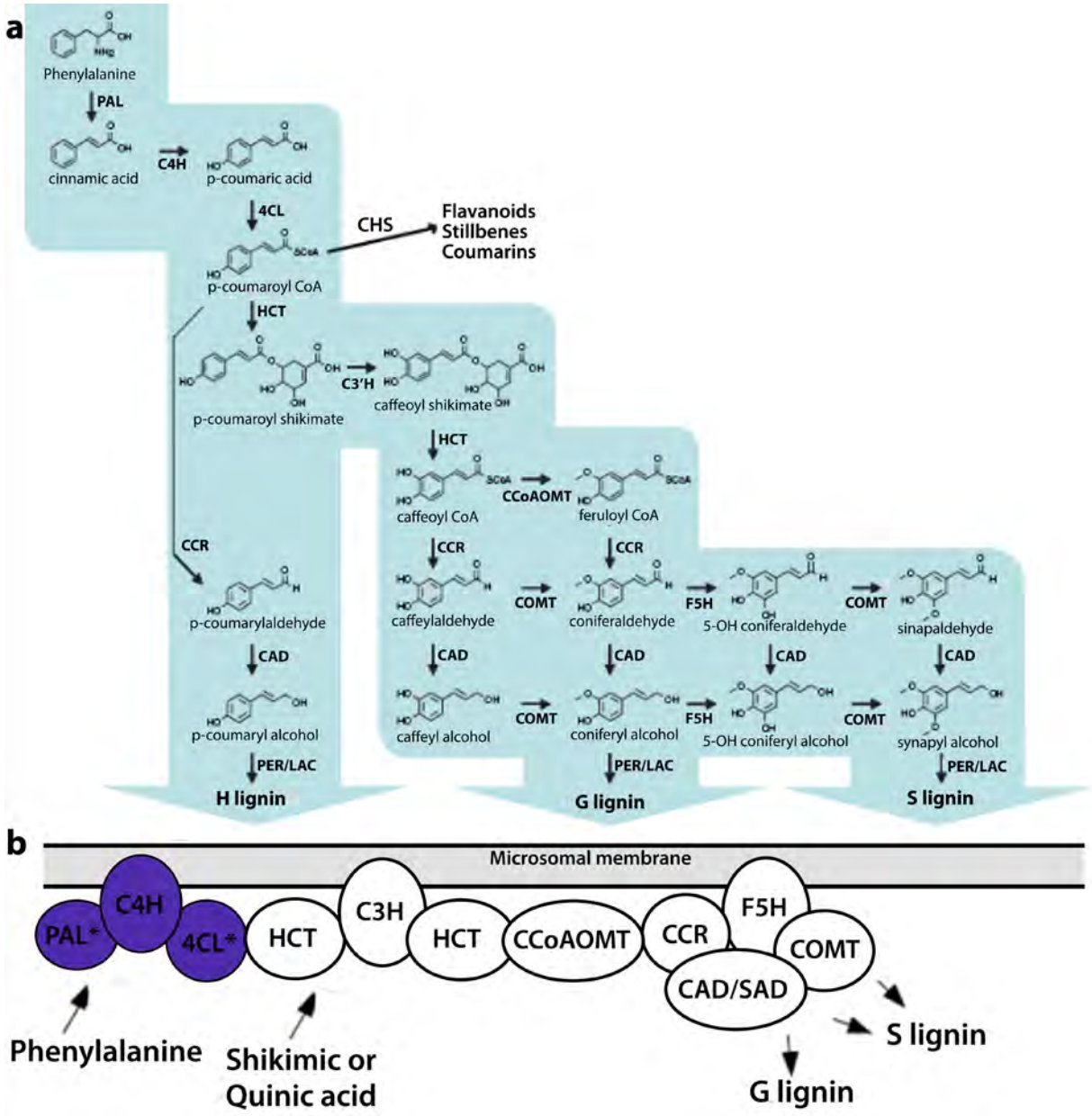


Figure 7. The highly conserved lignin biosynthesis pathway in flowering plants (a) (adapted from [70] and a model for membrane channelling of intermediates through the phenylpropanoid biosynthesis pathway (b) (adapted from [79]. Enzymes involved in multiple phenylpropanoid synthesis pathways are in purple with asterix (*) representing enzymes for which multiple isoforms have been identified; enzymes specific to a particular phenylpropanoid biosynthesis pathway (i.e. monolignol biosynthesis in this case) are in white. PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-(hydroxy)cinnamoyl CoA ligase; HCT, hydroxycinnamoyl CoA shikimate:quinate hydroxycinnamoyl transferase; C3'H, ρ-coumaroylshikimate 3'-hydroxylase; CCoAOMT, Caffeoyl CoA O-methyl transferase; CCR, (hydroxy)cinnamoyl CoA reductase; F5H, ferulic acid/coniferaldehyde/coniferyl alcohol 5-hydorxylase; COMT, caffeic acid/5-hydroxyferulic acid O-methyl transferase; CAD, (hydroxycinnamyl alcohol dehydrogenase; PER, peroxidase; LAC, laccase; CHS, chalcone synthase.

Whilst monolignol biosynthesis is fairly well characterized (at least in terms of the chemistry and enzymology), the mechanism of lignin polymerization has only recently begun to be elucidated. This is partly due to the fact that polymerization occurs within the complex cell wall matrix and attempts to mimic these conditions *in vitro* have been unfruitful [87].

It has long been accepted that monolignols are transported and excreted into the extracellular space by golgi bodies and this was supported by immunolocalisation studies of monolignol biosynthesis enzymes [88, 89]. However, recently TEM autoradiographs of radiolabelled monolignols in Lodgepole pine tracheids (cells undergoing lignification) have allowed monolignol transport to be visualized, revealing that they are not, in fact, translocated via the golgi [90]. Instead, a number of ATP-binding cassette (ABC) transporters have been suggested as candidates for energy-dependant monolignol export across the plasma membrane [90]. This is based on transcript profiling, which shows tight co-expression of some ABC transporters with monolignol biosynthesis genes [91] and subsequently, several gene knockout studies have also supported the role of ABC transporters in lignin excretion into the extracellular space [92]. Once the monolignols have been excreted into the extracellular space (cell wall), irregular polymerization of H, G & S lignin occurs.

Once lignin monomers have been delivered to the cell wall they must be polymerised, however the exact mechanism remains elusive. Currently, a model of random polymerisation where monolignols are 'primed' for polymerization before being randomly incorporated into the growing lignin polymer has been proposed [93]. Cell wall peroxidases and laccases are capable of oxidising monolignols in vitro (utilizing H_2O_2 produced by cell wall NADPH oxidases, and O_2 respectively) allowing polymerisation to occur and are therefore the primary candidates for the proposed 'priming' agents [94, 95]. A number of peroxidases and laccases have also been shown to be expressed in tissue undergoing lignification [96, 97], and several apoplastic peroxidases bind strongly to Ca^{2+} rich pectin gels which are

present in the middle lamella [96-98]. The later observation has prompted the claims that these peroxidases act as nucleation sites for lignin polymerization [94]. It has been very difficult, however, to identify specific peroxidases or laccases which act in lignin polymerization *in vivo* since most act on numerous substrates without having high specificity for monolignols [99]. Recently, LAC4 and LAC17 (laccases) have been shown to be required for lignin polymerisation [100-103], casparian strip domain proteins (CSDPs) in complex with NADPH peroxidases have been shown to direct site specific lignin polymerisation in *Arabidposis* roots [104], and a loss of function AtABCG29 mutant in *Arabidopsis* has been shown to reduce ligninification [105], thus strengthening the proposed mechanisms of lignin monomer export to cell walls and their polymerisiation.

Whilst the peroxidase/CSDPs complex has been shown to act as a nucleation site in developing *Arabidopsis* roots, it has been suggested that ferulic acid, which cross-links lignin to the cell wall matrix [106], acts as a nucleation site for lignin polymerisation within the secondary cell wall [94]. In fact, it has been observed that ferulic acid binds to arabinose side chains of GAX [106]. Thus, regions where peroxidase/laccase, monolignols and ferulic acid bound to GAX are present will be able to initiate lignin polymerisation from which large lignin polymers can grow. This is by random incorporation of monolignols which have been translocated to the cell wall matrix [106] (**Figure 8**).

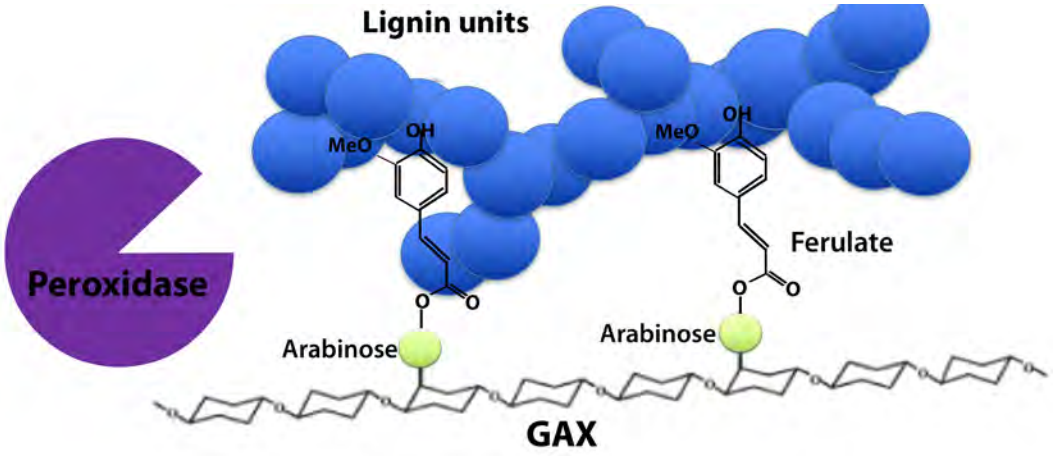


Figure 8. Schematic of lignin polymerisation facilitated by peroxidises, at nucleation sites where ferulic acid cross-links the growing lignin polymer to the cell wall matrix via arabinose side chains of GAX (adapted from [106].

2.3.1.3.1.2.2 Hydroxycinnamates

The hydroxycinnamates are intermediates in the monolignol synthesis pathway and therefore are phenolics which contain no fermentable sugars for bioethanol production. Significant amounts of ferulic acid (up to 4%) and ρ-coumaric acid (up to 3%) are uniquely incorporated into both primary and secondary cell walls of grasses [28]. The function of these hydroxycinnamates is relatively unknown, however, ferulates can dimerise and form cross links to GAX and to lignin. It has, therefore been suggested that ferulic acid replaces the functional role that structural proteins play in type I cell walls, cross linking hemicellulose, cellulose and lignin fractions [28]. It has also been suggested that ferulic acid acts as initiation sites for lignin polymerisation within the secondary cell wall [94]. Hydroxycinnamates are a major contributor to the recalcitrance of grass cell walls, particularly in ruminant digestion [107]. They are also inhibitory to yeast fermentation [28], thus they are a target for genetic manipulation of cell walls for increased digestibility.

2.3.1.3.1.3 Pectins

Pectins are the most structurally complex polysaccharides. They are galacturonic acid-rich and contain many and varied branch points producing complex polymers of many acidic and neutral five and six carbon sugars monomers [10]. Pectins form the middle lamellae of all plant cells acting as a glue that hold cells together [10]. They also facilitate plasmodesmatal connections between neighbouring cells [10]. Whilst pectins contain many sugar moieties, some of which are easily fermentable, they only comprise a small proportion of the cell walls of grasses [28]. This small proportion is greater in thin primary cell walls than thick secondary cell walls and therefore is probably only present in the middle lamellae. Therefore, whilst pectins in the middle lamellae play an important physiological role in the middle lamellae, the proportion of pectins in the grass cell wall is not enough to warrant interest as product for bioethanol production and will not be discussed further here. For a recent review see [108].

2.3.1.3.1.4 Structural proteins

Proteins are highly ordered polymers built from the 20 standard amino acids that are encoded in the genetic code of deoxyribonucleic acid (DNA). Whilst many of these amino acids contain sugar moieties and are able to be converted to the fermentable sugar glucose by gluconeogenesis, this is a complex metabolic pathway and would not be an efficient means of producing fermentable sugars for bioethanol production. Amino acids do, however, form an essential component of animal nutrition and therefore should be considered when assessing plant cell walls as ruminant feedstocks (however amino acids are supplemented in ruminants, to an extent, by gut bacteria that synthesise them from ammonia and urea).

The plant cell wall is an active organ that contains numerous proteins and enzymes that are specialised in individual cell types to facilitate cell wall deposition, cell expansion, respond to environmental stimuli, transport solutes and interact with neighbouring cells. These proteins can be soluble, loosely bound, or covalently bound to the cell wall, however, in grasses they are generally not structural [28]. In fact significant amounts of structural proteins have only been observed in primary cell walls of grasses and only make up ~1% of the dry weight. Recently, proteomic studies of the cell wall elongation zone of a primary maize root examined the soluble and loosely bound protein fractions. While many proteins had previously been identified in dicot cell walls, 18% were found to be unique to the type II maize cell wall [109]. Specifically an endo-1,3;1,4-β-D-glucanase was found which is implicated in MLG mediated expansion of primary cell walls, and an α-L-arabinofuranosidase which has been implicated in modification of GAX were identified [109] and are of specific interest since MLG and high proportions of GAX are unique to the grass cell wall. Further characterisation of these proteins and of the many other proteins unique to the grass primary cell wall will be of great interest in dissecting the molecular mechanisms involved in grass cell wall expansion and synthesis. Proteomic studies of grass cell walls in various species and tissues will also lead to a greater understanding of the role of proteins in the grass cell wall.

2.3.1.3.1.5 Silica

Silica, or silicon dioxide (SiO_2) is another component of the grass cell wall that is uniquely present in high proportions (2-15%) [28, 110] in comparison to most type I cell walls. Silica is mostly found in shoots of grasses, particularly leaves but also in stems [111]. It is also present in the endodermis and casparian strip of roots [111]. In shoots, silica is found in silica bodies (phytoliths) that are formed after secondary cell wall thickening, silicification of the cell wall and cytoplasmic degeneration of leaf epidermal cells [112]. Cells are slowly 'filled' with silica to form dumbbell, saddle and bulliform shaped clear silica bodies [113] at leaf maturity [112]. Silica is also found in intracellular spaces between cells and importantly, in the cell wall where silicification occurs. Very little is known about silicification of grass cell walls.

It has been shown that silicon may combine with the phenol (ferulic acid) or lignin fraction of the cell wall rather than the carbohydrate fraction [114] through a SiO-aromatic ring bond [115], however, it is unknown what form this silicon is in. It has also been shown that silicon promotes cell wall extensibility in elongating (primary cell walls) in rice [116], oats [116, 117], wheat [116] and sorghum [118] but conversely aids in rigidifying cells undergoing secondary cell wall deposition thus decreasing their extensibility [118]. The mechanisms by which this occurs is unknown however, it has been suggested that increased extensibility of primary cell walls could be due to the reduced density of cell wall polysaccharides when silica is present [117].

It is well established that silica is beneficial for plants [119]. It has been shown to: enhance resistance to herbivory, pathogen, metal toxicity, salt, and temperature stress [120, 121]; provide structural support to the plant [121]; allow regulation of leaf shape in drought conditions [121]; and increase canopy photosynthesis rates and grain yields of rice, wheat and maize [122]. The increased resistance to stresses, particularly resistance to herbivory by producing a stronger, more rigid cell wall, however, results in silica negatively effecting cell wall digestibility. In one case it was reported to have a greater effect than lignification [123], therefore regulation of silicification is a poorly studied area which presents opportunities for re-

engineering cell walls for greater digestibility, however, the beneficial effects of silica make this task a difficult balancing act.

On a molecular level, silica transport and enzymatic control of its localisation and polymerisation are poorly understood. Only recently three silica transport genes were functionally identified in rice, and their homologues were functionally confirmed in maize [124], barley [125], wheat [126] and pumpkin [127]. The three genes are the silica influxer *Lsi1* [128], the silica effluxer *Lsi2* [129] and *Lsi6* which is likely involved in xylem unloading of silica [124, 130]. It has been suggested that the mechanism of silica transport is slightly different in maize and barley in comparison to rice [131].

2.3.1.3.2 A model for primary cell wall expansion

Cell wall biosynthesis and the structural arrangement of its polymers is highly regulated by the plant in concert with developmental and environmental cues such as cell division, cell expansion, cell maturity, response to pathogens and so on. This regulation is complex, involving hormones, transcription factors and microRNAs [132] which act as master regulators of large suites of genes.

Once a cell divides in the meristem it must immediately begin synthesis of the cell wall. Following determination of cell fate the cell must then expand, depositing cell wall polymers in a complex manner that involves deposition of newly synthesised polymers and 'slippage' or cell wall 'creep' of existing polymers to accommodate and control an expanding cell. Synergistic action of hormones has been shown to regulate this process activating genetic switches that control gene expression [132]. Recently a brassinosteroid response transcription factor, BES1, was shown to bind to primary cell wall CesA gene promoters and regulate their expression in *A. thaliana* [133] implicating it as a transcriptional regulator of cell wall expansion. In rice a phytochrome-interacting factor-like protein PIL1/13 was shown to regulate cell expansion, and therefore plant height in the elongating regions of internodes [134].

Molecular regulation of primary cell wall expansion is clearly not well defined and is an open area of research. Also of interest is the mechanism by which the primary cell wall expands. Models have been proposed for dicot cell walls involving cell expansion by α-expansin mediated 'slippage' of CMFs at an auxin induced acidic pH of ~5 [10, 135]. Grass cell walls, however, have a very different primary cell wall composition and therefore the mechanism of cell wall expansion is poorly defined.

GAX is the predominant hemicellulose in grasses and it forms a cellulose-GAX network via hydrogen bonds to CMFs. The GAX is thought to form a chain fence like mesh which is also held together by hydrogen bonds [50]. The arabinose and glucuronic acid side chains of GAX [28] block efficient hydrogen bonding to CMFs and other GAX [51, 52] and this is currently thought to be the basis for cell wall expansion in grass cell walls.

Highly substituted GAX is secreted into the cell wall space thus it is not able to bind to CMFs or other GAX efficiently until its side chains are removed by cell wall enzymes [60]. Addition and removal of these side chains, therefore, presents a model for cell wall loosening and cell expansion where CMFs are able to 'slip' by each other and thus expand the cell wall (driven by turgor pressure) when side chains are present on GAX. As the cell wall ceases expansion it is strengthened and held into place by the removal of GAX side chains. This is supported by analysis of maize coleoptile cell walls by gas chromatography-mass spectroscopy showing that the amount of side chains present on GAX in expanding cell walls is much higher than that in mature, expanded cells [52].

The expansion of the grass cell wall is however, more complex than this simplified version. MLGs closely follow cell expansion with maximum levels occurring at the peak of cell expansion [57] suggesting a major role in grass primary cell wall structure and expansion. Hydroxycinnamates (phenolics such as ferulic acid and ρ-coumaric acid) are also present in much higher amounts in grass cell walls in comparison to dicots and are thought to aid in cross-linking GAXs, facilitating binding to CMFs [28, 51].

We can, therefore, present a more current model of the grass primary cell wall as consisting of CMFs densely coated in MLGs [60] and low-substituted GAXs, cross-linked by phenolics such as ferulic acid, embedded in a matrix of highly substituted GAX, and silica [28, 50, 51, 60] (**Figure 9**). In this model, cell expansion is facilitated by 'slippage' of CMF bound polymers, which is most likely modulated by regulating side chain removal from GAX and almost certainly involves MLG presence or modification [60]. β-expansin genes homologous to the well-characterized *A. thaliana* α-expansin gene family have also been identified in maize [136-138], however their precise role in cell wall loosening of grass cell walls is yet to be elucidated. Since no proteolytic or hydrolase activity of β-expansins has been detected [139] and β-expansins in maize have been shown to solubilise GAX [138], the current theory of cell wall loosening is based on the ability of β-expansins to reduce non-covalent bond strength of GAX, thus making GAX 'slippery' and allowing turgor pressure to force cell wall expansion.

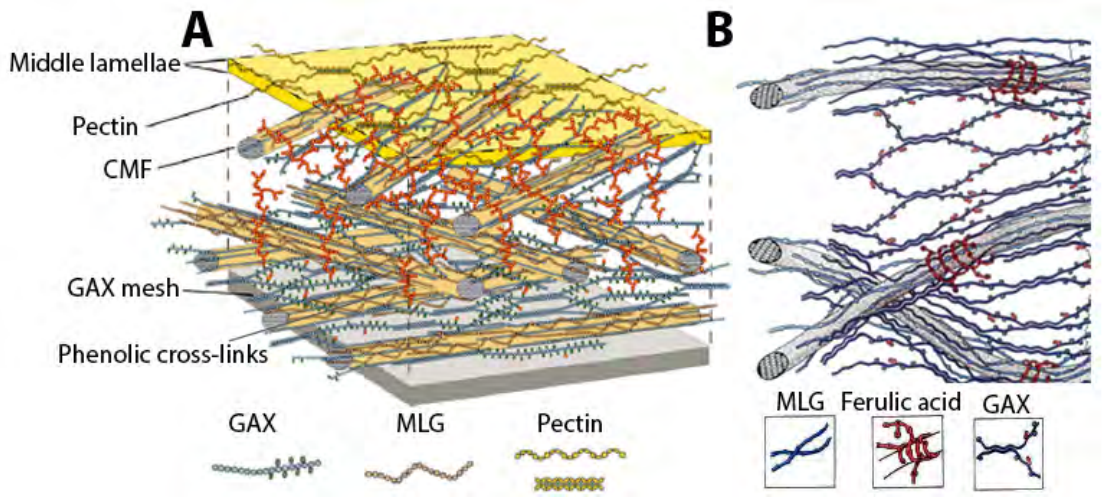


Figure 9. Current model of the grass primary cell wall (**A**) and a more detailed sketch of the structure of the GAX mesh that bonds (via hydrogen) to CMFs (**B**). Adapted from [50, 51].

Based on additional information showing that MLG and silica increase cell wall extensibility in grass primary cell walls, a novel mechanism for grass cell wall expansion is presented here. In an expanding primary cell wall, GAX polymers are highly substituted [140] and therefore do not bind CMFs tightly. MLG is synthesised and tightly coats CMFs [60]. The 'kinks' in the MLG polymers act like a

ratchet system allowing CMFs to stretch and slide along each other without recoiling. Silica has been shown to aid cell wall extensibility [116-118] and silicone oil (a polymer of silica combined with organic methyl groups, which are possibly provided by increased amounts of ferulic acid in primary cell walls) is widely used as a lubricant. Silica could therefore act as a lubricant in expanding cell walls facilitating efficient 'slippage' of CMFs along the MLG derived ratchet system that coats it, thus increasing cell wall extensibility. Since β-expansins have been shown to solubilise GAX [138], they could act to reduce hydrogen bonding specifically between the GAX-GAX mesh structure, thus this hypothesis explains cell wall creep of both the CMFs and GAX fraction of the network. As the cell ceases expansion, MLG is hydrolysed and GAX side chains are removed enzymatically, causing GAX to bind strongly to CMFs. Ferulic acid concentration is decreased and lignin is polymerised through the spaces in the cell wall matrix to rigidify and water proof the cell walls. This reduction in ferulic acid accompanied by an increase in silica, and the availability of methyl groups of lignin for interaction with silica could cause a chemical change in the silicon polymer that alters its properties to become rigid like sand or quartz. This would be in agreement with the observation that silica decreases cell wall extensibility [118] and cell wall digestibility [123] of secondary cell walls and provides protection against abiotic and biotic stresses [119].

2.3.1.3.3 Transition to secondary cell wall deposition

The secondary cell wall contains the greatest proportion of biomass [141], thus it is of most interest as a renewable source of carbon for bioenergy. The transition from primary to secondary cell wall synthesis is a highly regulated process. In *A. thaliana* many transcription factors involved in activation or repression of secondary cell wall synthesis have been identified [142]. These transcription factors are generally within the MYB or NAC domain gene families [142] and bind to specific AC elements of other MYB/NAC promoters and numerous cell wall biosynthesis gene promoters [143]. Since grass cell wall composition, anatomy and cellular distribution of secondary cell walls is vastly different to that of dicots, it is thought that novel transcriptional regulation will be acting in grasses [143].

To date only two transcription factors, which are homologous to those identified in *A. thaliana*, have been functionally tested in grass species suggesting that novel secondary cell wall regulatory networks remain to be discovered in grasses. The transcription factor, ZmMYB31 has been shown to reduce lignin deposition in secondary cell walls by regulation of the phenylpropanoid pathway in maize [144]. Similarly, overexpression of the switchgrass transcription factor PvMYB4 was shown to decrease lignin and phenolic contents of secondary cell walls by binding to AC-I, AC-II and ACII elements of monolignol biosynthesis genes [145].

Indirect evidence for the role of a number of other *A. thaliana* MYB and NAC homologues in secondary cell wall synthesis has also been provided. Complementation studies in *A. thaliana* show that rice and maize MYB46 and a number of secondary cell wall NAC domain (SWN) transcription factors, namely OsSWN1,3 and 7, and ZmSWN1, 3, 6 and 7, activate secondary cell wall biosynthesis [146] however, functional roles and physiological effects of altered expression has not been determined within a grass system. Similarly, by heterologous expression in *A. thaliana*, ZmMYB42 was shown to decrease cell wall lignin content [147], however this effect has not been reproduced in a grass species.

Once the regulatory factors have signalled a transition to secondary cell wall synthesis, the secondary cell wall is deposited inside the primary cell wall in the extracellular space. This is usually in three distinct layers, a thin primary layer (S1), a thick multilamellar secondary layer (S2), and occasionally a third layer (S3) (**Figure 4**). As opposed to the primary cell wall, which is designed for malleability, thus facilitating growth and defining cell shape, the secondary cell wall is designed for strength and its composition reflects this function [10, 26]. A typical grass secondary cell wall is composed of cellulose in the form of CMFs, embedded in a matrix consisting of GAX, lignin, silica and phenolics (i.e. ferulic acid & ρ-coumaric acid) [28, 110]. Since secondary cell walls are not required to expand (a very complex process), much of the complexity of the primary cell wall is absent. This can be seen with the absence of MLGs, reduced amounts of hydroxycinnamates and a reduction of side chain substitutions of GAX in secondary cell walls [28].

Secondary cell walls are generally deposited in a highly ordered manner with CMFs in each layer being deposited parallel to each other and offset by a set angle in each subsequent layer [10].

There is limited description in the literature of secondary cell wall structure. Despite this, an abbreviated model of secondary cell wall structure consists of CMFs bonded via hydrogen to a matrix of low-substituted GAX, which is cross-linked in a mesh structure, and is strengthened and rigidified by lignin, silica and phenolics [28, 87, 148]. Since lignin is probably the most prominent defining feature of secondary cell walls and due to its agricultural and industrial implications (i.e. reducing digestibility of lignocellulosic feedstocks for bioethanol production and rumen digestibility), most of the research in this area has focussed generally on the incorporation of lignin into the cell wall.

Lignification begins at the cell corners in the middle lamella where lignin polymers form spherical structures that infiltrate the primary, followed by the secondary, cell wall. Lignin forms lamella which coat CMFs and cross-link with hemicelluloses to strengthen the cell wall and create a hydrophobic environment [94]. It is thought that ferulic acid, which has been shown to cross-link lignin to the cell wall matrix [106], acts as a nucleation site for lignin polymerisation within the secondary cell wall [94]. In fact, it has been observed that ferulic acid binds to arabinose side chains of GAX [106]. Thus, regions where peroxidase/laccase, monolignols and ferulic acid bound to GAX are present, will be able to initiate lignin polymerisation from which large lignin polymers can grow through the deposited primary cell wall and the growing secondary cell wall. This is thought to be by random incorporation of monolignols which have been translocated to the cell wall matrix [106] although this is yet to be fully unravelled.

2.3.2. Harnessing energy from biomolecules

So far we have reviewed the process by which C_4 plants harness energy from the sun to build the three main plant biomolecule stores, that is: starch; soluble sugars; and cell walls. By breaking down these biomolecules, we are able to harness the energy stored within them and an understanding of how to do this efficiently will

be key to producing cost effective, carbon neutral fuels. Natural systems have been harnessing the energy from these molecules for millions of years, therefore it is helpful to have an understanding of how these processes work so that we can apply an understanding of these finely tuned evolutionary pathways to industrial processes.

2.3.2.1 Natural systems

Many natural systems are capable of breaking down, and therefore harvesting energy from the biomolecules produced by plants. In fact, in most cases these molecules are a major food source and are essential for the survival of almost all animal, fungal, bacterial and protozoan species.

The simplest food sources to deconstruct are disaccharides such as sucrose which can be metabolised by most animals, fungi, bacteria and protozoa by simple cleavage into its monosaccharide units of glucose and fructose. Heat and acids are able to cleave sucrose, however there is often an enzyme catalyst involved to increase the efficiency and gain biological control over the process. In mammals sucrose is hydrolysed by sucrases and the resulting monosaccharides are absorbed into the blood stream [149]. Bacteria and some animals use invertases to cleave sucrose into its monosaccharides. Plants also possess a number of invertases and sucrose synthase enzymes that are capable of cleaving sucrose [10]. The enzyme kinetics of sucrose hydrolysis and metabolism is well understood in many species and invertases are commonly used in efficient industrial processes to produce inverted sugar syrup which is a mixture of glucose and fructose that has a sweeter taste than sucrose [150]. Sucrose is also efficiently used directly as a fermentation substrate by yeast to produce ethanol and CO_2 [151].

Starch is also a relatively simple molecule to deconstruct. Most mammals contain α-amylase enzymes in their saliva that are capable of cleaving the α-1,4 glycosidic bonds of starch to yield glucose and maltose [149]. Maltose is then further digested by maltase and isomaltase enzymes that cleave the maltose and the α-1,6 glycosidic bonds that are not cleaved by α-amylase respectively [149]. Plants also contain α-amylases and α-glucosidases capable of degrading starch to yield

glucose [10]. The kinetics of starch degrading enzymes is well understood [152] and α-amylases are used in industrial processes, including bioethanol production from corn kernels to yield fermentable sugars.

Much fewer species are able to deconstruct the polysaccharides of the cell wall due to the β-glycosidic bonds from which cellulose, hemicellulose and pectin polymer backbones are constructed. Ruminants are capable of digesting cellulose and hemicellulose to some extent, however this is due to the presence of bacteria and fungi that live in their rumen (the first of four chambers of the ruminant stomach) [153]. The ruminant firstly masticates the plant cell wall material to disrupt the protective barriers of the tissue and release saliva that hydrates the material at alkaline pH [154]. Alkaline pretreatment has been shown to remove lignin from biomass and alter the hemicellulose and cellulose to allow better access of degrading enzymes [155]. The cell wall material is then swallowed exposing it to the acidic conditions of the foregut. The acidic conditions of the foregut solubilise hemicelluloses and also aid in the removal of lignin from the cell wall matrix reducing hydrophobicity and increasing the accessible surface area of cellulose [156]. This allows fungal and bacterial β-glucoside hydrolases to access the cell wall polysaccharides. The ruminant then regurgitates the cell wall material and re-masticates to further reduce particle size, thus increasing surface area of accessible cellulose which has been shown to greatly increase digestibility of lignocellulose [157]. The 'cud' is then swallowed again for further degradation by bacterial and fungal β-glucoside hydrolases and fermentation [158]. The degradation and fermentation products are then absorbed into the blood stream in the hindgut [158] (**Figure 10**). This process is a highly evolved mechanism for the deconstruction of cell wall polysaccharides and, in fact, the mechanism and process chemistry varies between ruminant species that have evolved from diets of different plant species with different cell wall types [158].

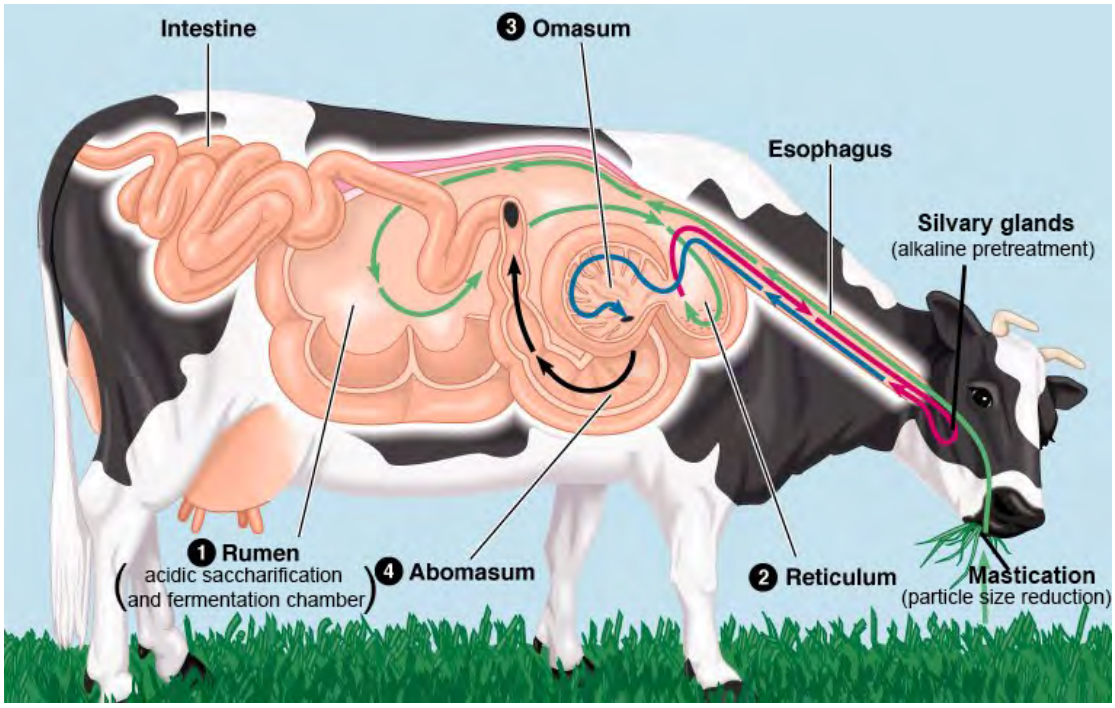


Figure 10. The ruminant digestive system that is capable of digesting plant cell wall polymers. The four chambers of the stomach are numbered in order from foregut to hindgut. This is the model from which most industrial bioethanol plants are based. Adapted from Addison Wesley Longman, 1999.

Whilst the ruminant pre-treats the lignocellulosic material and provides a fermentation/reaction vessel where bacteria and fungi thrive, it is clearly these microorganisms that possess the unique β-glucoside hydrolase enzymes for deconstruction of β linked cell wall polymer backbones. In fact numerous species of bacteria and fungi have been identified that possess numerous β-glucoside hydrolase gene families. Digestion of cellulose requires three main enzymes: exocellulases which cleave two to four unit chains of glucose from the exposed ends of cellulose polymers; endocellulases which randomly cleave internal bonds of cellulose polymers; and β-glucosidases which then convert these products into monosaccharides for metabolism or fermentation [159]. These enzymes have been identified in many fungal and bacterial species, thus a large number of glycoside hydrolase (GH) gene families have been identified, each with different enzyme kinetics and specificities for cellulose containing substrates [153, 159].

Complete digestion of GAX, the major hemicellulose in grass cell walls, requires six enzymes: endoxylanase, which randomly cleaves the xylan backbone of GAX; β-xylosidases, which release xylose monomers from the non-reducing ends of the xylan backbone of GAX; α-arabinofuranosidases, which catalyse the removal of arabinose side chains from GAX; α-glucoronidases, which catalyses the removal of glucuronic acid side chains from GAX; acetylxylan esterases, which remove acetyl groups from the xylan backbone of GAX; and ferulic acid esterases which catalyse the hydrolysis of arabinose-ferulate ester bonds which link GAX to lignin [159]. These enzymes have also been identified in many fungal and bacterial species, thus a large number of GH gene families exist with a range of enzyme kinetics and substrate specificities [159]. In addition, most of these enzymes, and cellulase enzymes, contain a carbohydrate-binding module (CBM) which increases substrate specificity and hydrolysis efficiency [153]. Many CBMs have been identified with different substrate specificities ranging from cellulose to hemicellulose to starch [160].

Fungi and bacteria are known to assemble these cellulose and hemicellulose degrading enzymes into complex structures called cellulosomes which contain multiple catalytic sites [153]. The arrangement of enzymes around structural cellulosome proteins has synergistic effects for the fungi [153] and allows the secretion of complex, but controlled enzyme cocktails. The range of GH gene families required for cellulose and hemicellulose deconstruction are also found in most plants, but rather than being assembled into cellulosomes, have been implicated in cell wall expansion and deposition [160].

Deconstruction of the phenolic component of cell walls, namely lignin, is a much more difficult process. Lignin is hydrophobic, non-porous and has a highly variable structure. Enzymatic degradation therefore seems impossible since the hydrophobicity of lignin reduces enzyme access to the polymer, the small pores of lignin do not allow the penetration of enzymes which are relatively large molecules, and the variable structure is incompatible with enzymatic cleavage which is usually highly specific to a certain bond type. For this reason only very few

organisms have been identified that can deconstruct and metabolise lignin [161]. Initially only the white-rot basidiomycetes (fungi) were thought to be able to metabolise lignin, however recently it has been suggested that bacterial species exist which are also capable of this feat [161]. By analysing the deconstruction of lignin by white-rot fungi it has been suggested that, rather than a direct enzyme catalysed reaction, deconstruction occurs via oxidative mechanisms [162]. Lignin peroxidase, manganese peroxidase and glyoxyl oxidase have been identified in white-rot fungi and have been shown to be capable of deconstructing synthetic lignin polymers (whose structure closely resembles *in vivo* lignin but does not represent its complexity) [162]. These enzymes produce oxidative species, such as Mn^{3+}, that are small and able to penetrate the small pores of the lignin polymer [162] . They are also less specific and are able to act on moderately activated aromatic rings that are present throughout the variable lignin polymer [161]. *In vitro* degradation of lignin using synthesised enzymes has not been demonstrated (at least not with any efficiency) and therefore enzymatic lignolysis cannot currently be applied to the pretreatment of biomass.

A comprehensive understanding of the biomass pretreatments applied by ruminants, the range of fungal and bacterial GH gene families and the specificities of their CBMs, and a better understanding of lignolysis in bacteria and fungi will lead to the engineering of efficient enzyme cocktails for industrial deconstruction of lignocellulose.

2.3.2.2 Human systems

Humans have utilised the energy stored in plant derived biomolecules as a tool ever since the mastery of fire. In modern times, many processes have been developed that utilise this stored chemical energy, however, by far the most common method is the combustion reaction. Hydrocarbons (i.e. methane and octane) consist of C and H atoms, alcohols (i.e. ethanol and butanol) consist of C and H atoms with a terminal OH group, and carbohydrates (i.e. sucrose, glucose, fructose, starch, cellulose and hemicellulose) consist of C, O and H atoms with higher proportions of O atoms than alcohols [14] (see **Figure 11**). If supplied with sufficient energy to initiate the reaction (auto ignition temperature), all carbon

based molecules will readily combine with O_2 to form the less complex, more stable, CO_2 and H_2O [163]. In fact, in reality most combustion is incomplete, since the O_2 supply is not always optimal, and reaction intermediates such as CO, H_2 and C are released [163].

The proportions of O and H atoms greatly affects the combustion properties of a fuel. Generally the proportion of H atoms in a molecule is proportional to the volatility and heat of combustion (energy released upon combustion MJ/kg) of that fuel. For example, carbohydrates have increased proportions of O atoms and decreased proportions of H atoms on the carbon backbone resulting in less volatile substances with a lower heat of combustion (viewed MSDS for sucrose, glucose, fructose, cellulose and starch). Hydrocarbons and alcohols such as ethanol, however have higher proportions of H atoms and are therefore more volatile and release more energy upon combustion (viewed MSDS for methane, methanol, ethane and ethanol). Their increased volatility also results in alcohol and hydrocarbon fuels being a liquid or gas at room temperature making them ideal for combustion engines used in transportation. For this reason, sugars, starch and cell wall polymers must be deconstructed and fermented into reduced molecules with higher proportions of H atoms such as ethanol for efficient use as transportation fuels.

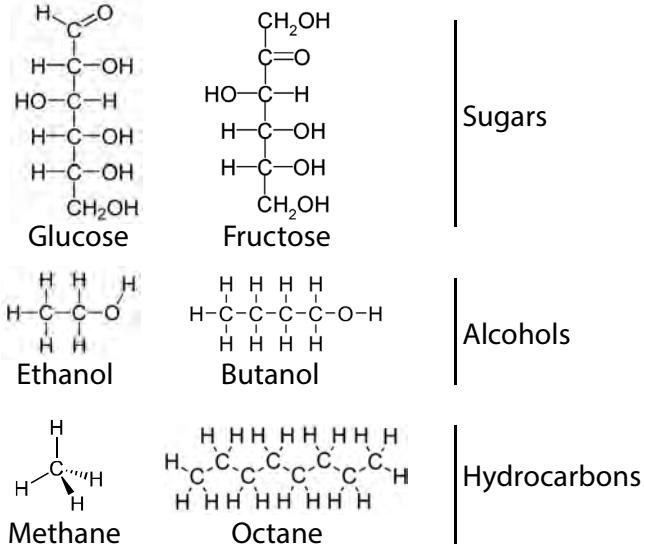


Figure 11. Chemical structures of common sugars and transportation fuels highlighting the requirement for fermentation due to the difference in H:O atom ratios.

2.3.2.3 Industrial bioethanol production

Industrial bioethanol production can be either 1st generation or 2nd generation. 1st generation is bioethanol produced from simple sugars such as sucrose, glucose, fructose and starch (which is easily hydrolysed by α-amylases). Corn kernels (mainly starch) are widely used in the US and sugarcane sucrose is used extensively in Brazil to efficiently produce 1st generation bioethanol that is economically competitive. The use of these foodstuffs as fuel however, reduces their availability as food and drives prices up, thus it is not thought to be sustainable on a large scale in most countries [20]. Sucrose from sugarcane can be directly fermented whilst corn starch is processed by either dry milling or wet milling [164]. These processes are well established, efficient industrial processes and will not be discussed further here.

2nd generation bioethanol is produced from lignocellulosic plant biomass and requires the additional steps of weakening the cell wall matrix and hydrolysing its β linked polysaccharide backbones. The industrial process for 2nd generation bioethanol production is currently an active field of research since lignocellulosic biomass is a large un-tapped resource but its conversion to fermentable sugars is not currently commercially viable. As of 2009 there were four functional non-commercial 2nd generation bioethanol pilot plants operational in the US and there is currently one in Australia [165]. More recently, 2nd generation bioethanol plants have become more widespread in the US.

Most industrial processes mimic the process employed by ruminants, that is, particle size reduction, alkaline and acid pretreatment, enzyme digestion and fermentation followed by distillation of high purity ethanol [159]. Many pretreatments have been trialled including liquid hot water, steam, steam explosion, hydrogen peroxide, ionic liquids, ammonia (alkaline), lime (alkaline) and dilute sulphuric acid [155, 156]. Alkaline treatment (specifically ammonia) has been shown to be effective at removing and altering the structure of lignin, and reducing the crystallinity of cellulose [155]. Yields of 90% of theoretical were reported after ammonia treatment and enzyme digestion [155]. Dilute sulphuric

acid has been shown to be effective at removing hemicellulose and altering the structure of lignin to increase the enzyme accessible surface area [155]. Yields of over 90% theoretical have been reported [166]. Like in ruminants, each pretreatment method must be optimised to specific substrates and processing plants. It is therefore difficult to suggest that any method is more efficient than another, however recently a method that closely mimics ruminant pretreatment of a two-step acid-alkaline treatment shows great promise with high yields of both glucose and xylose with few inhibitors of fermentation formed as by-products [167].

Enzyme digestion of pretreated lignocellulose is another area of active research. The most efficient methods employ simultaneous saccharification and fermentation (SSF). Recently, with the development of yeast strains with increasing pentose fermentation efficiency [56], simultaneous saccharification and co-fermentation of hexoses and pentoses (SSCF) is now possible [159]. Furthermore, efforts have been made to engineer a single organism that carries out both saccharification and fermentation (consolidated bioprocessing; CBP), however there are no current reports of commercially viable CBP organisms [168]. Attempts at bioengineering cellulosome complexes have also been made but none have been reported to increase cellulase activity of crystalline cellulose [169]. For recent reviews on cellulase developments see [153, 159, 168-170].

Continued progress towards more efficient, substrate specialised pretreatment and enzyme saccharification along with more efficient fermentation of pentoses by engineered yeast strains will lead to productive economically viable commercial scale lignocellulosic bioethanol plants.

2.3.3. Bioenergy feedstocks

So far we have reviewed the processes by which C_4 plants build biomolecules, natural systems that have evolved to harness energy from these molecules and the use of these molecules by humans as an energy source. To efficiently utilise these biomolecule production and degradation systems we must also identify optimal feedstock species and employ targeted strategies for improvement of these crops.

In the case of bioethanol, energy feedstocks must contain biomolecules that are easily deconstructed to yield fermentable molecules (i.e. six carbon sugars, and to a lesser extent five carbon sugars). This means that feedstocks containing high levels of soluble six carbon sugars such as glucose and fructose are ideal, as are those that accumulate high levels of disaccharides such as sucrose that is readily metabolised into its monosaccharide components. Polymers containing glucan backbones such as starch, cellulose and some hemicelluloses (i.e. XyG), and to a lesser extent polymers containing xylan backbones such as arabinoxylan are also of great value and coincidently are the main storage molecules of plants [10]. In considering potential plant based biofuel feedstocks it is therefore essential to find species that store energy in molecules based on six carbon sugars and that accumulate these molecules rapidly whilst using water and nutrients efficiently. It will also be pertinent to consider the potential for crop improvement for each species and the environments in which they prosper. Rather than choosing one energy feedstock to meet world fuel supplies, a more effective approach will involve the improvement of multiple technologies and crop species each selected to thrive in a particular environment or community.

Microalgae are effectively a single cell plant that reproduces rapidly and when grown in a nutrient rich medium can accumulate biomass more rapidly than terrestrial plants [171, 172]. Microalgae are capable of photosynthesising (thus fixing CO_2) and have been used as a bioenergy feedstock, however most research has focused on using its lipid fraction to produce biodiesel [171-175]. Lipids are not readily fermentable therefore cannot be used to produce ethanol. Since naturally occurring algae store up to 70% of their energy as lipids [173] they are not currently ideal for bioethanol production, however they are easily transformed and therefore their metabolic pathways can be easily manipulated. An example of this is the Algenol™ technology which has engineered the DIRECT TO ETHANOL® pathway into cyanobacteria (blue-green algae) allowing pyruvate produced by photosynthesis to be directly metabolised to ethanol [176]. Microalgae also rarely produce lignin or hemicellulose therefore their cell walls are a rich source of

cellulose and a two-step process where biodiesel is produced from the lipid fraction and bioethanol is produced from the carbohydrate fraction has been suggested [174]. Patents for rapid growing multicellular algae (such as kelp) containing microalgae transgenes for the production of hydrocarbons have also been issued [177].

Terrestrial plants such as poplar, which rapidly accumulates biomass, have also been proposed as a biofuel feedstock, however it is generally accepted that the C_4 grasses such as sugarcane, maize, switchgrass, miscanthus and sorghum are the optimal terrestrial crops for bioethanol production [178]. Sugarcane and certain sweet sorghum varieties uniquely accumulate high concentrations of sucrose in their stems at maturity and also produce high yields of lignocellulosic biomass [179, 180]. In fact, when combining fermentable sugar yields from both the soluble and lignocellulosic fractions of bioenergy crops, sorghum was shown to be the highest yielding with low cost, energy and nutrient inputs [1]. Sweet sorghum varieties have had little optimisation through breeding therefore there is a greater potential for crop improvement than in sugarcane. Adding Sorghums potential for improvements via breeding is the fact that it has a small, diploid genome and is easily crossed, compared with Sugarcane and Miscanthus which are both massively polyploidy, hard to cross, and have infrequent flowering making breeding very difficult. Sweet sorghums also produce a large seed head adding to the value of each crop and making sowing of seed much easier than sugarcane, which is propagated vegetatively from buds on stem billets. Sweet sorghums also grow optimally in mid to northern Australia and therefore strategies for crop improvement as an Australian biofuel feedstock will be further explored here.

2.3.3.1 Crop improvement strategies

Most food crops, especially grasses such as wheat, corn and rice, have been domesticated for thousands of years. Over time a range of breeding strategies have been employed, from primitive phenotype selection through to advanced molecular marker and transgenic techniques, to optimise these food crops for yield, quality and resistance to pathogens. Food crops including grain sorghum therefore produce reliable high yielding crops. Sweet sorghums, however have had

very little selection pressure placed upon them with respect to their use as a biofuel feedstock. For this reason, advanced crop improvement strategies capable of improving biofuel quality and yield must be developed and deployed. If plant derived bioethanol is to contribute to mitigating the effects of climate change, these strategies must be capable of condensing thousands of years of selection pressure into a commercial, high quality, high yielding biofuel feedstock in decades.

To achieve such a feat, multiple approaches will be required, including:

- Cataloguing and utilising natural variation in sweet sorghum genotypes
- Plant breeding including classic observable phenotype-based breeding strategies and marker assisted breeding
- Mutagenesis of existing lines and mapping of quantitative trait loci (QTL) to improve marker assisted breeding
- Improvements in agricultural practises
- Awareness of these issues within the agricultural community.

Clearly rapid deployment of such strategies will require a simple high-throughput method of assessing phenotypes in relation to biofuel feedstock quality and yield and is perhaps the greatest hurdle in generating rapid progress, especially due to the complex nature of the cell wall and the complexities of its deconstruction into fermentable sugars which currently require expensive, time-consuming methods to assess.

2.3.3.1.1 Utilising natural variation

There is a great diversity among wild and domesticated sweet sorghum varieties. The largest most diverse collection of sweet sorghum genotypes has been collected by the United States Department of Agriculture (USDA) and contains 2180 variable genotypes [181], however over 40, 000 accessions are thought to exist worldwide [182]. Screening of small subsets of the USDA Sorghum diversity collection, requiring large-scale expensive destructive analysis, have revealed great diversity in cell wall composition and hydrolysis potential. 381 field grown accessions were screened for cellulose hydrolysis yields [183]. A selection of 20

accessions from the same panel were screened for hydrolysis yield (fermentation products were quantified), cellulose crystallinity, and amounts of cellulose, hemicellulose and lignin in stem and leaf tissues [181]. Similarly, a collection of 152 biomass, forage, Sudangrass-like and sweet accessions were screened for biomass composition using the standard National Renewable Energy Laboratory (NREL) chemical hydrolysis method [184]. This method quantifies relative amounts of lignin and component monosaccharide sugars of the cellulose and hemicellulose fractions of the cell wall [185]. None of these studies accounted for biomass or the soluble fraction of the plants.

To understand and utilise this great natural resource, clearly there is a requirement for more high throughput methods that are capable of screening the full diversity that lies within the 40,000 accessions that are available around the world.

2.3.3.1.2 Breeding

Breeding programs found in the literature can be categorised as crossing of existing lines for optimisation of traits which can be marker and/or phenotype assisted, recombinant inbreeding for the identification of quantitative trait loci (QTL), and chemical mutagenesis of existing lines. Generally crossing is used as a tool for transferring beneficial traits of one line, such as disease resistance, with beneficial traits of another line, such as high yield, to generate progeny with both beneficial traits. This approach can also be used to generate hybrids with more vigour (thus yield) and produce male sterile lines. There are very few scientific reports of this form of crop improvement in Sorghum since it is not generally of interest to the scientific community and is often performed by commercial seed companies who desire specific traits for commercialisation under specific agricultural conditions. Well defined genetic markers, however can be used in marker assisted breeding with this approach and this relies on the identification of QTL for desirable traits.

Recombinant inbreeding of lines that differ greatly in a trait of interest generate progeny that inherit the trait in a mendelian or mendelian-like manner. The trait variability within these lines can be used to genetically map QTL thus facilitating

marker-assisted breeding. As of 2011, 50 QTL studies had been published mapping sorghum phenotypes to genomic regions [186]. Since then, as of August 2015, at least 6 more reports have been published [187-192] focussing on bioenergy traits including stem sugar concentrations and biomass composition and saccharification efficiency. QTL associated with: grain germination, morphology, composition and weight; leaf senescence, composition, morphology and yield; maturity; panicle yield and architecture; stress resistance to cold and stay-green phenotype; resistance to 16 pathogens; stem yield, morphology, biomass composition, sugar accumulation, tillering, height lodging tolerance, regrowth and 'rhizomatous-ness' have all been mapped numerous times [186]. There is, however, often disagreement on the location of genomic QTL between studies [186], likely due to different growth conditions, parental lines used, and methods of measurement and data analysis. Despite this, integration of 50 QTL studies was performed to generate confidence intervals for each QTL and collapse overlapping QTL into metaQTL which in some cases contain multiple traits [186]. The best understood QTL is probably that of plant height (PHT) [193]. In sorghum PHT has been mapped to four genetic loci contained on chromosomes 6 and 7: *Dw1*, *Dw2*, *Dw3* and *Dw4* [194]. Only *Dw3* however, has been linked to the polar auxin regulating gene responsible for the trait [195]. The QTL for PHT has been shown to correlate with numerous traits including high biomass, delayed flowering time, high soluble sugar accumulation in sweet lines and increased lodging [193]. Besides the increased lodging associated with PHT QTL, the increased biomass and sugar accumulation are desirable for bioenergy feedstocks and therefore of interest in future breeding programs.

Whilst many QTL have been accurately mapped and the PHT QTL are a well characterised example, studies identifying QTL for biomass hydrolysis and stem sugar accumulation are small in number and none have integrated bioenergy related traits to look at synergistic or negative correlations between QTL. For example, to understand whether a QTL for cell wall yield is positively or negatively correlated with soluble sugar accumulation it would be ideal to examine both traits simultaneously within one population of recombinant inbred lines (RILs) grown

under the same conditions. High throughput screening methods would also increase the efficiency of such studies and allow measurements of numerous bioenergy traits simultaneously and allow the effects of environmental conditions to be studied without a huge requirement for resources.

Chemical mutagenesis has also been used in many species as an efficient way of generating variable traits and has been instrumental in the discovery of novel phenotypes and mapping those traits to single nucleotide polymorphisms (SNPs) of single genes [196, 197]. With respect to the plant cell wall the first induced alteration of the lignin biosynthesis pathway was by chemical mutagenesis of sorghum which identified 28 brown midrib mutants (*bmr*) mutant lines. These mutants have provided and continue to provide information regarding the nature of lignin biosynthesis and an experimental system for studying the effects of reduced and altered lignin on plant fitness [198, 199]. *Bmr* mutants, due to their increased saccharification efficiency have also been of great agricultural importance as a forage crop and now as a potential biofuel feedstock [200, 201].
Interestingly, despite such success with these *bmr* mutants, very few reports of genetic mapping of sorghum mutants exist. Whilst numerous Sorghum mutagenesis populations exist and large repositories of chemically induced phenotypes are available [202-204], most phenotyping is purely observational and no reports of mapped chemically induced mutations relating to bioenergy traits have been reported. The maturation of next-generation sequencing has allowed much more rapid detection of genome SNPs, therefore will facilitate the more rapid characterisation of chemically induced mutation. This will be an integral step in identifying the missing genetic links in cell wall synthesis of grasses and lead toward a better understanding of the genetic control of bioenergy traits.

2.3.3.1.3 Transgenic

Transgenic manipulation of plant cell walls provides an opportunity to fine-tune cell wall composition to the requirements of the biofuel industry however, in practise this requires a transformable model species and a comprehensive understanding of the fundamental molecular machinery that underlies cell wall construction and how that relates to its physiological function in a growing plant.

Transgenic manipulation of cell wall components by down-regulation or overexpression of genes is currently limited by the transformation efficiency of biofuel feedstock species. While there are reports that C_4 grasses maize [205], switchgrass [206], sugarcane [207] and sorghum [208] are all transformable at acceptable levels, tissue regeneration of C_4 grasses is difficult and in species with polyploid genomes such as sugarcane, transgene stability can be an issue [207]. In addition, crop species have long life cycles therefore obtaining homozygous transgenes and performing crosses to introduce multiple transgenes is time consuming. A model species that is transformable with fast regeneration times will be required to allow rapid progress. *Brachypodium distachyon* is a C_3 grass with type II cell walls that is transformable and has rapid regeneration times [209]. It is currently being used as a model species for cell wall research [210], however, considering the deleterious effects of reduced cell wall lignin in *Bmr* plants that is specific to C_4 grasses [201], we suggest that the model species *Setaria viridis* is a more appropriate choice. *S. viridis* is a C_4 grass that is genetically more closely related to maize, switchgrass, sugarcane and sorghum than *B. distachyon* (**Figure 3**) [16, 17, 211, 212] and has recently been adopted as a model species for the study of C_4 photosynthesis [5, 210, 212].

Despite issues with transformation of C_4 grasses, a collection of gene knockdown and overexpression studies have been performed, generally in C_3 dicots such as *Arabidopsis thaliana*, alfalfa, poplar and tobacco. Whilst many cell wall synthesis genes, specifically those of the lignin biosynthesis pathway have been transgenically manipulated in these species giving great insight into cell wall deposition and outcomes for biofuel feedstocks, almost all studies involve single gene knockdowns or overexpression and there have been very few attempts at altering multiple genes for positive synergy towards improved feedstocks. New 'omics' platforms including proteomics, transcriptomics, genomics and metabolomics will facilitate understanding of whole systems at a molecular level and this systems approach will inform coordinated genetic manipulation of multiple genes in the future.

Numerous gene knockdown and overexpression lines have been mentioned throughout this review, however in general, gene effects on biofuel characteristics (i.e. secondary cell walls) can be summarised as follows:

- Single gene knockdowns that reduce cellulose deposition consistently result in brittle stems in rice and maize
- Very few genetic manipulations of hemicellulose synthesis genes in grasses have been performed and therefore no consistent phenotype can be proposed
- No genetic manipulations of cell wall MLG, pectin, silica or hydroxycinnamates have been performed
- Knockdowns of lignin biosynthesis genes result in either a reduction in lignin content, usually with adverse affects on yield, or altered lignin structure which generally increases digestibility of cell wall polysaccharides.

Of great interest is the genetic manipulation of transcription factors that regulate cell wall deposition. Two MYB family genes have been shown to control lignin deposition in maize, however of most interest is the SHINE/WAX INDUCER 1 (SHN1) which has proven to be a global regulator of cell wall synthesis in rice [213]. SHN1 increases cellulose and hemicellulose proportions and decreases lignin universally in cell walls. This greatly increased fermentable sugar yields upon digestion. Cell walls of these mutants are thickened which may compensate for the lack of lignin as plant biomass is increased and show no growth defects [213]. Another interesting transgenic line, whilst not in a grass species, is the overexpression of sucrose synthase in poplar which produced an increase in cellulose without any detrimental effects on plant health [214]. An ideal transgenic biofuel feedstock will likely contain multiple gene manipulations that will hopefully act synergistically to produce the large yield increases required to make industrial bioethanol production economically viable.

2.3.4. An improved breeding strategy for *Sorghum bicolor*

Fourier Transform infrared spectroscopy (FTIR) has been used to characterize molecules for ~100 years and were first used in the analysis of plant cell walls more than 30 years ago [215]. More recently, when coupled with multivariate data analysis techniques, FTIR spectroscopy has proven to be a robust and accurate method for high-throughput screening of cell wall mutations in experimental plant tissues such as the model species *A. thaliana* [216, 217] and *Z. mays* coleoptiles [60, 218] where even slight variations in the molecular structure of cell walls were detected. Recently, FTIR spectra have been used as a predictor for enzymatic hydrolysis of pre-treated biomass [219] and similarly, in the food industry to rapidly quantify sucrose, glucose and fructose in juice including mango and apple [220, 221] as well as other foodstuffs such as honey [222]. To date however, there have been no studies to incorporate this technology into high-throughput screening methods for biofuel feedstocks. To this end, a holistic high throughput screening strategy has been developed based upon stalk geometry, and PLS predictive models of FTIR spectra collected from the soluble sugar and cell wall fractions. This methodology enabled the identification of both high sugar accumulating and highly digestible lines and an assessment of total fermentable sugars in sorghum. Furthermore, we were able to characterise variations in soluble sugars and identify cell wall compositional changes irrespective of the effect on digestibility in a large variable population.

This novel methodology will be an ideal tool for thorough characterisation of natural sorghum germplasm collections, chemical mutagenesis populations and recombinant inbred lines and will hopefully help to realise the collective goal of generating high quality, high yielding bioenergy feedstocks within decades rather than millennia.

CHAPTER THREE

A high-throughput screening strategy for *Sorghum bicolor*

3.1 Summary

This chapter consists of two publications and outlines a high-throughput screening approach that utilises fourier transform infrared (FTIR) spectroscopy and mathematical modelling to rapidly screen large populations of *Sorghum bicolor* plants. Both publications were submitted to *Biotechnology for biofuels*. The first, which outlines the method has been published, while the second, which shows the use of the method to detect cell wall variants, was initially accepted but with additional data requested by reviewers. Additional data is currently being collected and the manuscript is being prepared for re-submission. It is presented in its current form here.

3.2 Publication 2: A holistic high-throughput screening framework

Martin AP, Palmer MP, Byrt CS, Furbank RT, Grof CPL **A holistic high-throughput screening framework for biofuel feedstock assessment that characterises variations in soluble sugars and cell wall composition in *Sorghum bicolor*** *Biotechnology for Biofuels* **6**, 1-13, doi:10.1186/1754-6834-6-186 (2013)

Statements of contribution of others: Publication 2

I attest that Research higher degree candidate ***Antony Martin*** *was the primary contributor to the development of this publication. The author contributions were as follows:*

APM theorised the sugar and cell wall FTIR models. WMP theorised the internode correlation study. APM designed all experiments, collected all data, analysed all data, produced Figures 1, 2, 3, 4, 5, 6, 7, and 8, and drafted the manuscript. WMP designed and collaboratively conducted the internode correlation, sugar model data collection and analysis and assisted in producing Figures 1-5. CB reviewed experimental design and provided feedback on the digestibility model. CPLG conceived and established the research program and contributed to overall experimental design. RTF contributed to experimental design and reviewed the manuscript. All authors read and approved the final manuscript.

_____

Antony Martin

_____

William Palmer

_____

Caitlin Byrt

_____

Robert Furbank

_____

Christopher Grof

Biotechnology for Biofuels

METHODOLOGY Open Access

A holistic high-throughput screening framework for biofuel feedstock assessment that characterises variations in soluble sugars and cell wall composition in *Sorghum bicolor*

Antony P Martin[1], William M Palmer[1], Caitlin S Byrt[1,2], Robert T Furbank[3] and Christopher PL Grof[1*]

Abstract

Background: A major hindrance to the development of high yielding biofuel feedstocks is the ability to rapidly assess large populations for fermentable sugar yields. Whilst recent advances have outlined methods for the rapid assessment of biomass saccharification efficiency, none take into account the total biomass, or the soluble sugar fraction of the plant. Here we present a holistic high-throughput methodology for assessing sweet *Sorghum bicolor* feedstocks at 10 days post-anthesis for total fermentable sugar yields including stalk biomass, soluble sugar concentrations, and cell wall saccharification efficiency.

Results: A mathematical method for assessing whole *S. bicolor* stalks using the fourth internode from the base of the plant proved to be an effective high-throughput strategy for assessing stalk biomass, soluble sugar concentrations, and cell wall composition and allowed calculation of total stalk fermentable sugars. A high-throughput method for measuring soluble sucrose, glucose, and fructose using partial least squares (PLS) modelling of juice Fourier transform infrared (FTIR) spectra was developed. The PLS prediction was shown to be highly accurate with each sugar attaining a coefficient of determination (R^2) of 0.99 with a root mean squared error of prediction (RMSEP) of 11.93, 5.52, and 3.23 mM for sucrose, glucose, and fructose, respectively, which constitutes an error of <4% in each case. The sugar PLS model correlated well with gas chromatography–mass spectrometry (GC-MS) and brix measures. Similarly, a high-throughput method for predicting enzymatic cell wall digestibility using PLS modelling of FTIR spectra obtained from *S. bicolor* bagasse was developed. The PLS prediction was shown to be accurate with an R^2 of 0.94 and RMSEP of 0.64 $\mu g.mgDW^{-1}.h^{-1}$.

Conclusions: This methodology has been demonstrated as an efficient and effective way to screen large biofuel feedstock populations for biomass, soluble sugar concentrations, and cell wall digestibility simultaneously allowing a total fermentable yield calculation. It unifies and simplifies previous screening methodologies to produce a holistic assessment of biofuel feedstock potential.

Keywords: FTIR, High-throughput, PLS modelling, Cell wall, Soluble sugar, Digestibility, *Sorghum*, Biofuel

* Correspondence: chris.grof@newcastle.edu.au
[1]School of Environmental and Life Sciences, University of Newcastle, University Drive, Callaghan NSW 2308, Australia
Full list of author information is available at the end of the article

Background

The photosynthetic fixation of atmospheric CO_2 to produce sugars by C_4 plants is an efficient conversion of sunlight into stored chemical energy. In grasses a large proportion of these sugars are stored either as soluble sugars, plant cell wall polymers, or starch. Maximising sugar yields from these three fractions for a broad spectrum of uses including animal feedstock, human nutrition, and second generation biofuels is imperative for efficient utilisation of plant productivity in the face of a burgeoning global population.

This study focused on the soluble sugar and cell wall fractions as a biofuel feedstock. Considering these two fractions, there are four main contributors that maximise total fermentable sugar yields of biofuel feedstocks: juice volume (V), sugar concentration ($[S]$), biomass (B), and digestibility (D). These attributes can be formulated as:

$$V.[S] + B.D = \textit{total fermentable sugar yield}$$

A European Union EPOBIO project [1], and a more recent report [2], stipulate that there is a need for a high-throughput method for assessing cell wall digestibility [1,2]. Since these reports, a number of high-throughput digestibility assays have been developed [3-5]. These methodologies present feasible high-throughput platforms for screening biomass, however, do not present strategies for high-throughput quantification of total plant biomass (B), or of the soluble sugar fraction (V and $[S]$). In addition, analysis requires automated robotics which is costly, and uses wet chemistry assays that are, again, costly, use hazardous chemicals, and even when automated, can be time consuming. This study focused on quantifying all of these attributes (V, $[S]$, B, and D) simultaneously, whilst maximising efficiency and reducing cost.

The principal components of cell walls are cellulose, hemicellulose, and lignin. The composition and arrangement of these carbohydrates and phenolic polymers is very diverse across plant species, organs, tissues, and even cell types [6]. With a strong focus on renewable energy being derived from plant biomass, the highly efficient C_4 plants and more specifically monocot grasses have garnered significant interest in recent times as potential biofuel feedstocks [7]. A promising saccharification approach for plant biomass requires reduction to a uniform size, pretreatment with weak acid or alkali, followed by enzymatic hydrolysis to release sugars for fermentation. The commelinoid group of monocotyledons which includes the grasses, hence principal biofuel feedstock targets such as sugarcane, *Sorghum*, and *Miscanthus*, possess cell walls containing ferulic acid and glucuronoarabinoxylan as the major non-cellulosic polysaccharide. Furthermore, the grasses differ from other commelinoids in possessing (1,3;1,4)-β-glucans [8]. Cell walls of non-commelinoid monocotyledons and most dicotyledonous plants do not contain ferulic acid and the non-cellulosic polysaccharides are pectins and xyloglucans [6,9]. These differences in cell wall composition and architecture as well as the bonds between the principal components strongly influence 'digestibility' of the plant biomass, whether this be undertaken in a biofuels processing context or by rumen/gut microorganisms.

Analysis techniques based upon infrared (IR) spectroscopy are proving to be highly effective tools for the detailed resolution of plant cell wall composition. IR spectroscopy is based on the absorption of IR light by specific quantized vibrational energy states of bonds between atoms and within molecules [10]. Mid-infrared (MIR) absorption spectra (4,000 to 400 cm^{-1}) define vibrational modes directly, generating sharp peaks that are more readily interpreted than near-infrared (NIR; 13,000 to 4,000 cm^{-1}) or far-infrared (FIR; 10 to 400 cm^{-1}) spectra [10]. Fourier transform MIR (FT-MIR or FTIR) spectra have been used to characterise molecules for approximately 100 years and were first used in the analysis of plant cell walls more than 30 years ago [11]. More recently, when coupled with multivariate data analysis techniques, FTIR spectroscopy has proven to be a robust and accurate method for high-throughput screening of cell wall mutations in experimental plant tissues, such as the model species *Arabidopsis thaliana* [12,13] and *Zea mays* coleoptiles [14,15], where even slight variations in the molecular structure of cell walls were detected. Recently, FTIR spectra have been used as a predictor for enzymatic hydrolysis of pre-treated biomass [16] and similarly, in the food industry to rapidly quantify sucrose, glucose, and fructose in juice including mango, apple, and sugarcane [17-19] as well as other foodstuffs such as honey [20]. Similarly, NIR has been used for rapid prediction of soluble sugars in sugarcane [21] and of biomass composition in *Miscanthus* and *Sorghum* [22,23].

Whilst these studies are suggestive of an application in high-throughput screening of biofuel feedstocks, there have been no reports incorporating this technology into a holistic, high-throughput assessment of biofuel feedstock potential. To this end, a screening strategy has been developed based upon stalk geometry, and partial least squares (PLS) predictive models of FT-MIR spectra collected from the soluble sugar and cell wall fractions. The methodology (Figure 1 and Additional file 1 for detailed protocol) was shown to be accurate for plants harvested at 10 days post-anthesis and enabled the identification of both high sugar accumulating and highly digestible lines leading to an assessment of total fermentable sugars in the model C_4 monocot *Sorghum bicolor*. The use of FT-MIR for both soluble sugar

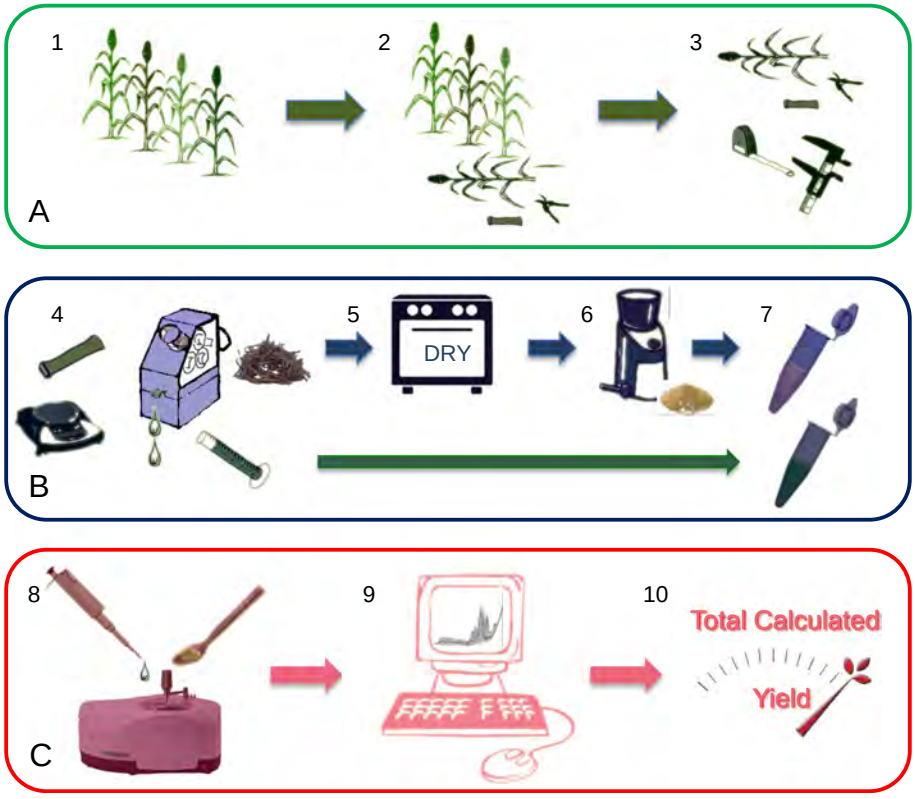


Figure 1 High-throughput methodology flow chart. (A) (1) Diverse plant population; (2) harvest the fourth internode; and (3) record in-field geometric measurements. **(B)** (4) Process into juice and bagasse; (5) dry bagasse; (6) grind dried tissue; and (7) process juice and bagasse. **(C)** (8) Collect FTIR spectra; (9) PLS prediction of FTIR spectra; and (10) calculate combined total fermentable sugar yield incorporating juice volume (*V*), sugar concentration ([S]), biomass (*B*), and digestibility (*D*) (see Additional file 1 for full protocol). FTIR, Fourier transform infrared; PLS, partial least squares.

concentration and cell wall digestibility prediction simplified the process and reduced set-up costs. Furthermore, the collection of FT-MIR, rather than NIR, spectra provides an opportunity for downstream spectral interpretation of cell wall modifications that lead to altered digestibility.

Results

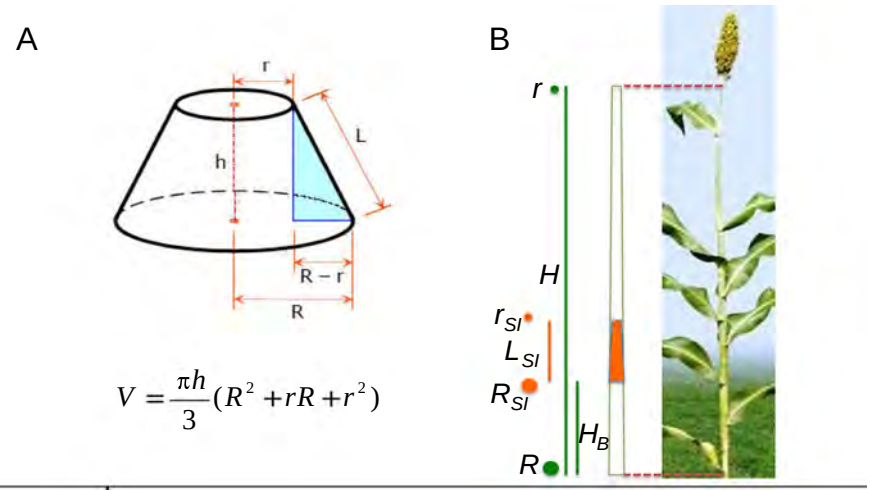
A detailed step-by-step protocol for undertaking the proposed screening method is available in Additional file 1. Briefly, the main stalk from each plant, harvested at approximately 10 days post-anthesis, was cut at the base of the first elongated above ground internode and laid along the length of a 4-meter ruler. Two heights and two diameters were recorded in a personal digital assistant (PDA) and the fourth internode from the base of the plant that had expanded more than 2 cm was sampled by cutting at the node with secateurs. Using this method, one person was able to sample approximately 100 internodes per hour. Samples were then pressed and the released juice was placed on the attenuated total reflectance (ATR) crystal of a portable FTIR spectrometer for spectral acquisition. The remaining bagasse was simultaneously placed in a drying oven followed by grinding, washing with water, and placing on the FTIR ATR crystal for spectral acquisition. Data was then batch processed by applying predictive models to FTIR spectra and incorporating data into yield calculations. An overview of this process is displayed in Figure 1 and the results describing the calibration and accuracy of each prediction process are presented below.

High-throughput calculation of whole stalk biomass, sugar content, and cell wall hydrolysis yield

For the purpose of calculating the whole stalk volume it was assumed that a *S. bicolor* stalk approximates a conical frustum (Figure 2). By simply measuring the radius at the top (r) and bottom (R) of the stalk and the stalk height (H), the volume of the whole stalk (V_{WS}) could be derived using the equation for the volume of a conical frustum:

$$V_{WS} = \frac{\pi H}{3}\left(r^2 + rR + R^2\right)$$

Using only two additional measurements, length from the base of the stalk to the base of the sampled



C	FW (g)	DW (g)	Sucrose (mM)	Glucose (mM)	Fructose (mM)	Digestibility (μg.mgDW^{-1}.h^{-1})
IN/WS ratio	0.91 ± 0.02	1.00 ± 0.02	0.97 ± 0.02	1.05 ± 0.02	0.97 ± 0.02	1.29 ± 0.05
Correlation (r)	0.97	0.98	0.94	0.88	0.87	0.85
RMSEP (%)	5.2	4.9	5.6	8.2	5.8	10.8

Figure 2 Extrapolating sampled biomass characteristics to whole stalks. (A) *S. bicolor* stalks were assumed to approximate a conical frustum allowing a simple volumetric calculation. **(B)** The fourth internode from the base of the plant that had expanded more than 2 cm was harvested and the height (*H*), stalk radius at the base (*R*) and top of the stalk (*r*), the height to the bottom of the fourth internode (H_B), and the length of the fourth internode (*L*) were recorded. Fresh weight (FW), dry weight (DW), sucrose, glucose, and fructose concentrations of pressed juice and digestibility of the lignocellulose were measured in the fourth internode at approximately 10 days post-anthesis. Using the volumetric ratio (V_{ratio}), these measurements were extrapolated to whole stalks and compared to equivalent measures taken from the same whole stalk. **(C)** The ratio of extrapolated fourth internode measures to whole stalks (IN/WS ratio), their correlation (r), and the error in prediction (RMSEP, %) are tabulated. DW, dry weight; FW, fresh weight; RMSEP, root mean squared error of prediction.

internode(s) (H_B) and length of the sampled internode (L_{SI}), the volume of the sampled internode (V_{SI}) was calculated. Both measurements were acquired simply by using a ruler during stalk height measurement, whilst the base radius (R_I) and top radius (r_I) of the sampled internode(s) were derived using simple geometry to obtain the following equation:

$$V_{SI} = \frac{\pi L_{SI}}{3}\left(R_I^2 + r_I R_I + r_I^2\right)$$

where $R_I = R - H_B\frac{R-r}{H}$ and $r_I = R - (H_B + L_{SI})\frac{R-r}{H}$

See Additional file 2 for full derivation.

The volumetric ratio of the sampled internode to the whole stalk was then calculated as:

$$V_{ratio} = V_{SI}/V_{WS}$$

From this ratio, measurements taken on the sampled internode could be extrapolated to the whole stalk by simply dividing by V_{ratio}. This, of course, is under the assumption that the concentration, or density of the measured variable in the sampled internode(s) is either equivalent to, or a fixed ratio of that in the whole stalk. In our case, the fourth internode from the base of the stalk was used as the sample internode since a stable relationship between sugar concentrations in the fourth internode and the whole stalk had been repeatedly observed in previous field measurements (unpublished data) and it has been shown that lower internodes in maize stalks within a genotype are less variable [24]. The relationship between fresh weight (FW), dry weight (DW), sugar concentration, and cell wall digestibility of the fourth internode and that of the whole stalk was determined using 15 mature, field grown sweet *S. bicolor* genotypes harvested at 10 days post-anthesis with a height and FW range of 237 to 338 cm and 228 to 941 g (Additional file 3), respectively (Figure 2). Four replicate, glasshouse grown 'Rio' sweet sorghum plants, also harvested at approximately 10 days post-anthesis, were used in the cell wall digestibility correlation calculations to supplement lost samples.

Calculations using V_{ratio} and measurements from the fourth internode were compared to those taken from the whole stalk. FW and DW returned values that were 91 ± 2% and 100 ± 2% of the whole stalk with correlations of 0.97 and 0.98, respectively. Sucrose, glucose, and fructose returned values that were 97 ± 2%, 105 ± 2%, and 97 ± 2% of the whole stalk with correlations of 0.94, 0.88,

and 0.87, respectively. Digestibility as calculated from the fourth internode, however, gave values that were 129 ± 5% of the whole stalk with a correlation of 0.85 (Figure 2C). Whilst not all ratios approached 100%, all had small standard errors suggesting that these fixed ratios were consistent across genotypes, and could therefore be used in extrapolating measurements from the fourth internode to whole stalks.

Using these experimentally determined ratios, FW, DW, sucrose, glucose, fructose, and digestibility values were adjusted and the root mean squared error of prediction (RMSEP) was determined to be 5.2%, 4.9%, 5.6%, 8.2%, 5.8%, and 10.8%, respectively (Figure 2C). The calculation-based method proposed here is, therefore, accurate across the 15 examined genotypes and is suitable for assessing total biomass, juice volume, sugar content, and digestibility of whole stalks in a high-throughput screen of sweet *S. bicolor* populations. Whilst these correlations allow relatively accurate predictions to be made in a high-throughput manner, it should be noted that it is the nature of predictive modelling that cultivars or samples which do not follow the established rule, such as a lower stem-specific gene mutation, will not perform well in the model. Often these samples will be identified as outliers for more rigorous study; however, one must also accept that there is always a statistical probability that a phenotype of interest will not be detected.

To our knowledge, this is the first mathematical-based modelling strategy for assessing total stalk biomass in a high-throughput manner in *S. bicolor* or any other biofuel feedstock. Whilst high-throughput methods for assessing shoot biomass are being developed using two-dimensional images [25] this technology is not applicable to dense field plots with thick canopies and does not allow extrapolation of other measurements. The method presented here not only produced data on stalk geometry including height and stalk diameter which are traits of interest when breeding lines for increased biomass yields, but also greatly reduced processing time by eradicating the need to: strip whole stalks of their leaves and leaf sheaths; transport large stalks from the field to the laboratory; operate large balances; press juice from whole stalks; dry whole stalks in large drying ovens; and grind whole stalks in a large roller mill. In addition, since this approach relies on stalk geometry that resembles a conical frustum, it is suggested that it could be applied to other grasses with similar geometry such as sugarcane, maize, and millet; however, the fixed ratios would have to be separately determined in each species.

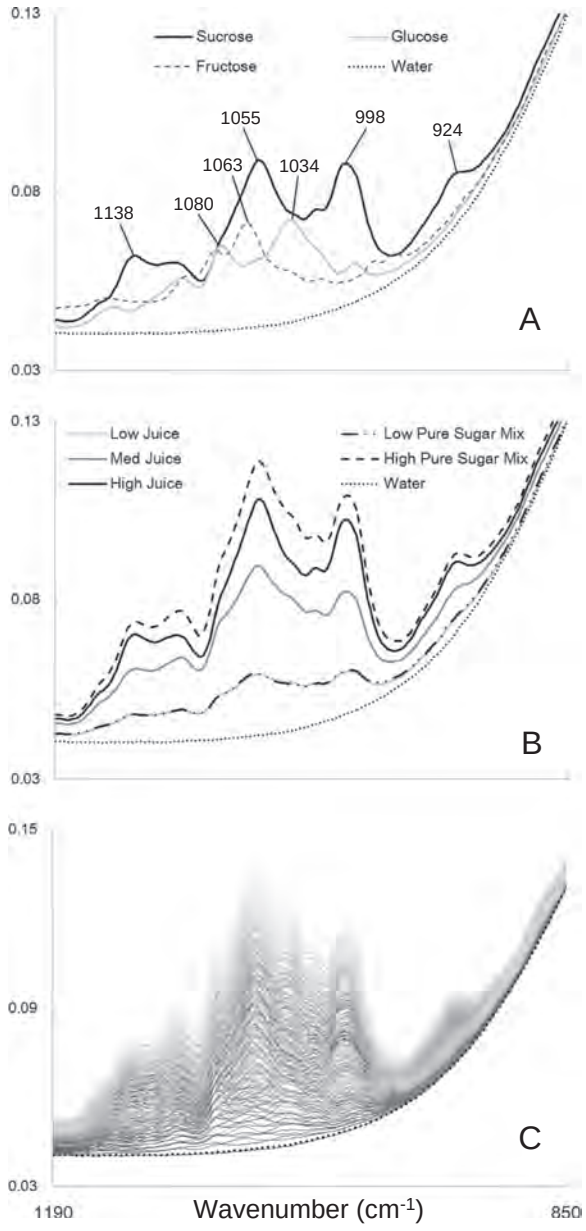


Figure 3 FTIR spectra of pure sugars, *S. bicolor* juice, and the PLS model calibration set. FTIR spectral fingerprint region collected from: **(A)** pure sucrose, glucose, fructose, and H_2O; and **(B)** *S. bicolor* juice containing low, medium, and high concentrations of sugars and a low and high mixture of pure sucrose, glucose, and fructose against a H_2O background. **(C)** Set of 216 samples consisting of varying mixtures of sucrose, glucose, and fructose making up the calibration set. Wavenumbers between 1,190 to 850 cm^{-1} were plotted. X-axis represents spectral wavenumber (cm^{-1}) and y-axis units are arbitrary. FTIR, Fourier transform infrared; PLS, partial least squares.

FTIR prediction of sucrose, glucose, and fructose concentrations in stalk juice

FTIR spectra were collected from pure sucrose, glucose, fructose, and H_2O samples (Figure 3A) for comparison with *S. bicolor* juice samples that were known to contain low, medium, and high mixtures of sucrose, glucose, and fructose (as determined initially by brix and then gas chromatography–mass spectrometry, GC-MS, data not

shown) and low/high concentration mixtures of pure sugars (Figure 3B). Pure sucrose, glucose, and fructose each exhibit unique spectra that were easily identifiable above the consistent H_2O background (Figure 3A); however, each had overlapping IR signatures making quantification by Beer's law inappropriate, since the variations of one sugar would affect the IR signature of the others. Multivariate PLS regression was therefore required to generate predictive models for each sugar that would model this accurately.

It was shown that pure sucrose, glucose, and fructose mixtures have almost identical IR signatures in the fingerprint region (specifically from 1,190 to 850 cm^{-1}) to that of juice pressed from *S. bicolor* (Figure 3B), not surprisingly, as sucrose, glucose, fructose, and H_2O are the major constituents of *S. bicolor* juice [26]. Hence, a PLS model for sucrose, glucose, and fructose could justifiably be calibrated using 216 mixtures of pure sucrose, glucose, and fructose as the calibration set (Figure 3C) (see Additional file 4 for concentrations).

Using these 216 calibration mixtures, PLS regression models were generated using the non-linear iterative partial least squares (NIPALS) algorithm (Unscrambler X, Camo, Woodbridge, NJ, USA) with mean centred data and full cross-validation. The FTIR spectral fingerprint region (1,800 to 800 cm^{-1}) was used as the x-predictor variables and sucrose, glucose, and fructose concentrations as the y-response variables. It was found that variable reduction (that is, reducing the number of spectral prediction variables) resulted in more robust models. Raw spectra truncated to 1,180 to 900 cm^{-1} wavenumbers were found to give the most accurate models and were used to generate the final PLS models for sucrose, glucose, and fructose. A model with six latent variables (LVs) minimised the root mean squared error of calibration (RMSEC) and maximised R^2 for sucrose and fructose giving a RMSEC of 5.62 and 2.46 mM, respectively, and both with R^2 of 0.99. A model with nine LVs was selected for glucose, which gave a RMSEC of 2.69 mM and R^2 of 0.99. It is not determined why glucose required three additional LVs, however, predictions with six LVs were also accurate. This was not thought to be over-fitting since independent validation samples performed slightly better with nine LVs in comparison to six LVs.

Each PLS model was offset by subtracting the y-intercept of the cross-validation predicted versus reference plot (effectively subtracting the water background) and a small slope adjustment factor was introduced for sucrose to minimise the slope deviation from one of the cross-validation samples. The PLS models, including adjustment factors, were then assessed with an independent test set consisting of three variable *S. bicolor* juice samples of known sugar concentrations, which were either spiked with pure sucrose, glucose, and fructose, separately and mixtures of the three, or diluted to achieve a large variable set of juice samples with known amounts of sugars. This was thought to be a more accurate method for generating a variable test set of juice samples than collecting a large variable set of *S. bicolor* juice samples and determining their sugar concentrations by GC-MS or HPLC. When the predicted values were plotted against the reference values for sucrose (Figure 4A), glucose (Figure 4B), and fructose (Figure 4C), the independent test set achieved a root mean squared error of validation (RMSEV) of 11.93, 5.52, and 3.23 mM, respectively, and a validation coefficient of determination (Q^2) of 0.99 for each (R^2 is used to denote the coefficient of determination of calibration samples and Q^2 is used for validation samples). According to Tamaki and Mazza (2011) [27] these indicate 'excellent' models since Q^2 >0.9 and $Q^2 - R^2$ <0.2, and are certainly adequate for a high-throughput screen. Furthermore, LV coefficient peaks for each sugar corresponded with the major sucrose, glucose, and fructose peaks (Figure 3A) (Additional file 5) indicating that the PLS regression models are indeed based on the chemistry of the sugars. The lower limit of detection (LoD) was conservatively determined to be approximately 25 mM, and spike/dilution tests were shown to be accurate (Additional file 6).

To further assess the accuracy of the generated PLS models, sucrose, glucose, and fructose were measured in three variable *S. bicolor* juice samples 'low' (120 mM sucrose, 52 mM glucose, 22 mM fructose), 'medium' (290 mM sucrose, 126 mM glucose, 70 mM fructose), and 'high' (461 mM sucrose, 56 mM glucose, 23 mM fructose) with six technical replicates using GC-MS and compared to results obtained with the PLS models (Figure 5A). Results were comparable (Figure 5A). Similarly 1,000 *S. bicolor* juice samples for which brix had been measured were assessed using the high-throughput FTIR-based PLS models. Total sugars (g/100 mL) were obtained by addition of sucrose, glucose, and fructose (g/100 mL) calculated from their respective PLS predictions using molar masses. A correlation of 0.94 between brix and total sugars (g/100 mL) was observed over the range of 5 to 20 g/L with an offset of −1.87 g/100 mL and a slope of 1.08 (Figure 5B). The observed negative offset indicated that brix readings were generally higher than those obtained from the FTIR PLS models. This was to be expected since brix measures not only sugars but all soluble solids in the sample. A slope greater than 1 suggests that this affect is less pronounced at higher sugar levels. This relationship to brix therefore supports the robustness of the PLS prediction models and it is suggested that this could be a more accurate method for widespread adoption by the sugar industry, specifically since it facilitates high-throughput quantification of

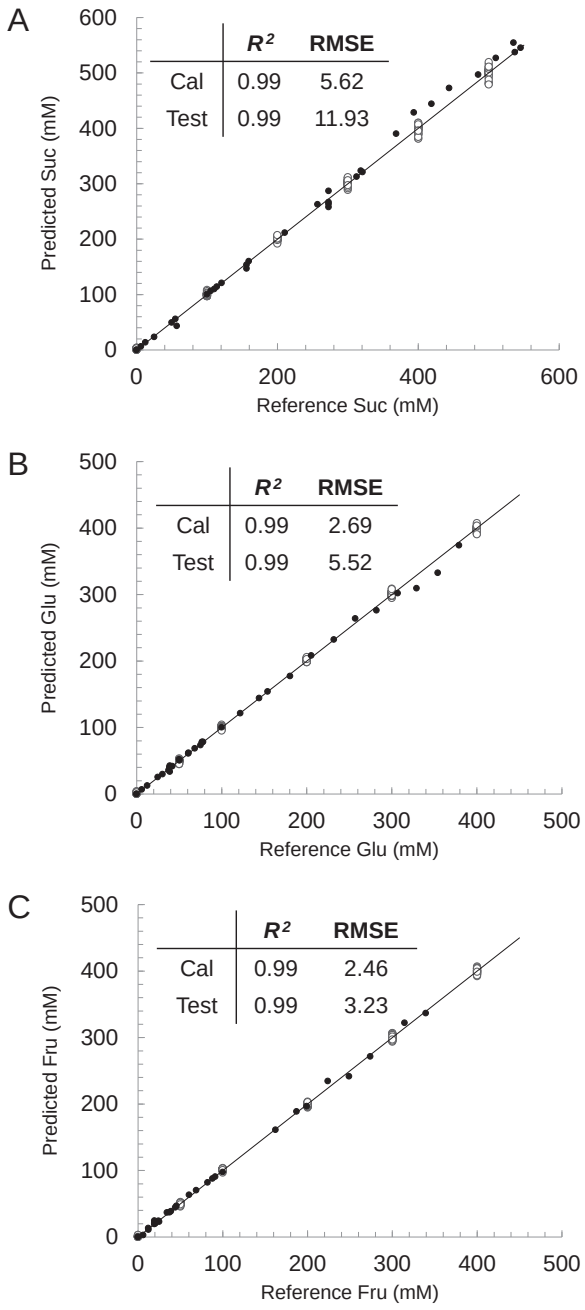


Figure 4 Soluble sugar FTIR PLS prediction models. PLS prediction of **(A)** sucrose (mM), **(B)** glucose (mM), and **(C)** fructose (mM) in an independent test set of *S. bicolor* juice mixtures. Predictions are plotted against the known reference values for each sample. Calibration samples (○) and independent validation samples (•) were overlaid on a target line ($y = x$). The coefficient of determination (R^2) and root mean square error (RMSE) for each sample set are displayed for each sugar. FTIR, Fourier transform infrared; PLS, partial least squares; RMSE, root mean squared error.

sucrose, glucose, and fructose separately. In addition, the FTIR method (30 seconds per sample) is much more rapid than GC-MS (1 hour per sample) or HPLC (20 minutes per sample) and requires no sample preparation. *S. bicolor* juice samples have been shown to degrade rapidly at room temperature [26] which means that samples must be treated to remove bacterial contamination and enzyme activity or HPLC devices must have a refrigerated autosampler if long sample runs are to be performed. Since the FTIR PLS prediction method is rapid there are no such issues with degradation of sugars. Whilst FTIR and NIR prediction of juice sugars have been reported previously [17,19,21], to our knowledge, this is the first report of accurate predictions of sorghum juice sugars. Although, based on previous reports, it was expected that models would be accurate, developing them was integral to the implementation of this high-throughput methodology.

To convert the measured sugar concentrations into sugar yield per stalk, V_{ratio} and the calculated whole stalk FW and DW (thus relative water content) from Section 1 could be used.

FTIR prediction of cell wall digestibility

Biomass, or more specifically, the plant cell wall, is a complex matrix of polymers; therefore, it was not possible to create a variable calibration set of cell walls in the laboratory, as was done with the soluble sugar PLS models. A selection of 90 highly variable *S. bicolor* ecotypes at varying developmental stages, grown under a range of different conditions (see Additional file 7 for sample descriptions) was, therefore, used to calibrate and validate an FTIR-based cell wall digestibility model.

Digestibility (D) was quantified by enzymatic hydrolysis of ball-milled cell walls using a *Trichoderma reesei* cellulase mixture (Sigma-Aldrich, Castle Hill, Australia) and the released reducing sugars were measured by the 3,5-dinitrosalicylic acid (DNS) method [2]. A large range of digestibility was observed from 0.95 to 12.1 μg.mgDW^{-1}.h^{-1} with the majority of samples lying between 1 and 10 μg.mgDW^{-1}.h^{-1} (Figure 6A). From this variable set of samples, 52 were selected as the calibration set and 38 as the independent validation set so as to approach a uniform distribution between 1 and 10 μg.mgDW^{-1}.h^{-1} as closely as possible for each set.

FTIR spectra were collected from the selected calibration set and a PLS regression model for cell wall digestibility was generated using second derivative spectra with an extended multiplicative scatter correction (EMSC) applied (spectra available in Additional file 8). A PLS model with spectra truncated to 1,800 to 850 cm^{-1} was calibrated and regression coefficients with 90% confidence intervals that spanned 0 were removed (that is, insignificant variables). The final PLS prediction model was calibrated from 147 wavenumbers with four LVs minimising the RMSEC (0.52 μg.mgDW^{-1}.h^{-1}) and maximising the R^2 (0.94) (Figure 6B). The model was then

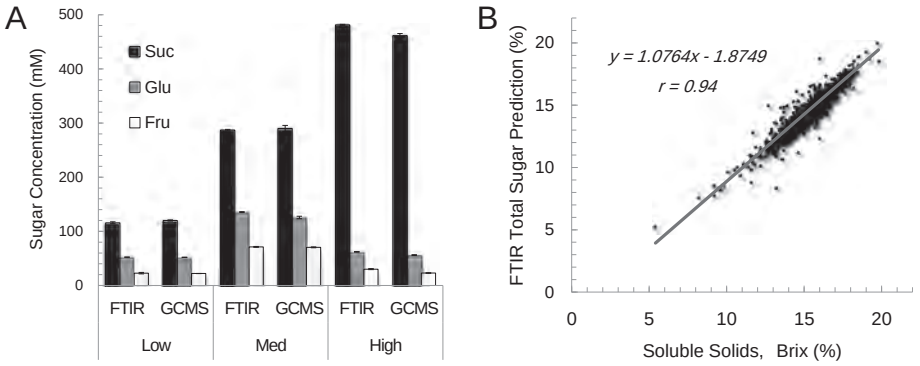


Figure 5 Comparison of the FTIR PLS model to GC-MS and brix. PLS prediction of sucrose (mM), glucose (mM), and fructose (mM) compared to values obtained using **(A)** GC-MS and **(B)** total sugars (g/100 mL) calculated using sucrose, glucose, and fructose PLS prediction models compared with total soluble solids (°brix). A linear regression line is fitted to the data and the equation and correlation (*r*) is displayed. FTIR, Fourier transform infrared; GC-MS, gas chromatography–mass spectrometry; PLS, partial least squares.

used to predict the digestibility of an independent test set.

When the predicted values of the independent test set were plotted against the reference (as determined by DNS assay) a RMSEV of 0.64 $\mu g.mgDW^{-1}.h^{-1}$ and Q^2 of 0.94 were observed (Figure 6B). This indicated an 'excellent' model since Q^2 >0.9 and $Q^2 - R^2$ <0.2 [27], and is adequate for a high-throughput screen. In addition, the PLS model regression coefficients resembled cell wall spectral peaks (Additional file 9) indicating that the model was based on cell wall chemistry. Analysis of the FTIR spectral peaks contributing to the prediction of cell wall digestibility in the PLS model could elucidate novel cell wall chemistry that is related to digestibility, however, this was outside the scope of this methods paper.

To further assess the PLS digestibility model, the digestibility of *S. bicolor* mutants, *bmr6*, *bmr12*, and their parent line was measured (six biological replicates) using both DNS assay and the FTIR-based PLS model (Figure 7). An increase in digestibility in both *bmr6* and *bmr12* was observed, as previously reported in the literature [28-30], and both methods produced comparable values (Figure 7). The observed biological variation (standard error), was also comparable between the two methods (Figure 7) providing further validation of the PLS prediction of digestibility.

To convert the measured digestibility into cell wall hydrolysis yield per stalk, the calculated stalk DW (using V_{ratio} from Section 1) could be used, assuming a 24-hour cellulase digestion.

FTIR prediction of digestibility by this method is much more rapid (approximately 45 seconds per sample) than cellulase digestion, which requires a 24-hour incubation and an average total processing time (excluding incubations) of approximately 15 minutes per sample. FTIR prediction of enzymatic cell wall hydrolysis has previously been demonstrated [16], as has biomass analysis by NIR spectroscopy in *Miscanthus giganteus* [22], switchgrass [31], and corn stover [32]; however, this is the first example in *S. bicolor* and the first time these predictive models have been incorporated into a holistic high-throughput screen in a biofuels context. FT-MIR was advantageous in our high-throughput screen since MIR spectra were used in the prediction of soluble sugars, thus only one Portable IR machine was required. MIR cell wall spectral peaks are also more readily interpreted than NIR spectral peaks [10], which allowed downstream interpretation of cell wall chemistry in variants identified from the high-throughput screen, but is beyond the scope of this publication.

It should also be noted that each PLS predictive model is specific for the wet chemistry method employed. In our case, a *T. reesei* cellulase mixture was used on biomass subjected to milling followed by a weak acid pre-treatment under 100 KPa pressure, a 24-hour incubation at pH 4.8/55°C, and the resulting reducing sugars measured by the DNS assay. Currently, biomass composition data as determined by the National Renewable Energy Laboratory (NREL) method [33] is being used to calibrate PLS models from the exact same spectral and biomass sample set. This highlights the fact that numerous wet chemistry and hydrolysis methodologies can be applied to the same variable sample set to calibrate numerous PLS predictive models. Prediction of biomass composition and enzymatic hydrolysis under varying conditions could then be made simultaneously from a single spectra. Once plants are being grown specifically for lignocellulosic ethanol production, this technique could be applied in an agricultural context to dictate price per ton based on a predicted hydrolysis yield that is unique to the methods employed by each processing plant.

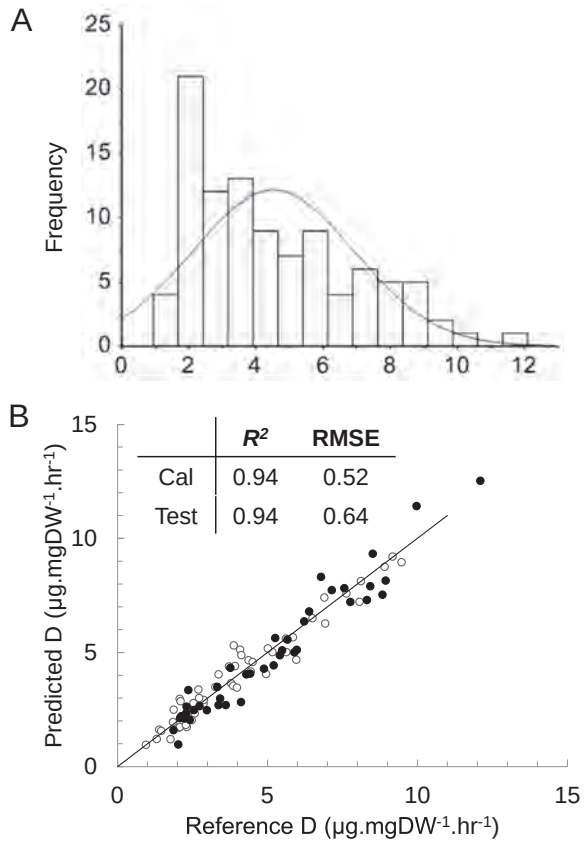


Figure 6 FTIR prediction of cell wall digestibility. (A) Distribution of cell wall digestibility (µg.mgDW^{-1}.h^{-1}) from a diverse collection of *S. bicolor* ecotypes grown under a range of different conditions and harvested at varying developmental stages. A fitted normal distribution has been overlaid. **(B)** PLS prediction of cell wall digestibility (*D*; µg.mgDW^{-1}.h^{-1}) in an independent test set. Predictions are plotted against the known reference values for each sample. Calibration samples (○) and independent validation samples (•) were overlaid on a target line ($y = x$). The coefficient of determination (R^2) and root mean square error (RMSE) for each sample set are inset. DW, dry weight; FTIR, Fourier transform infrared; PLS, partial least squares; RMSE, root mean squared error.

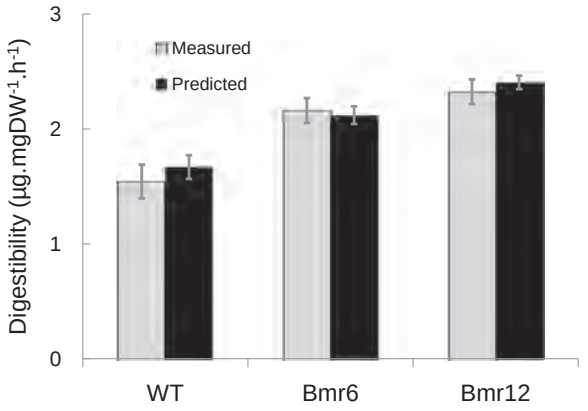


Figure 7 FTIR PLS prediction of cell wall digestibility in *S. bicolor* mutants. PLS prediction of cell wall digestibility (µg.mgDW^{-1}.h^{-1}) in *S. bicolor* mutants *bmr6*, *bmr12*, and their parent line (WT) harvested at 10 days post-anthesis, compared with DNS determination of digestibility. Data are represented as means of six biological replicates ± SE. DNS, 3,5-dinitrosalicylic acid; DW, dry weight; FTIR, Fourier transform infrared; PLS, partial least squares; SE, standard error.

Calculation of total fermentable sugars

Using the equation:

$$V.[S] + B.D = \textit{total fermentable sugar yield}$$

where whole stalk juice volume (V) and dry weight of biomass (B) were calculated in Section 1, total sugar concentration ([S]) was determined in Section 2, and digestibility (D) was determined in Section 3, the total fermentable sugar content of *bmr6*, *bmr12*, and their parent line (WT) were calculated (Figure 8). The high-throughput methodology was compared to values obtained using traditional methods (that is, brix and DNS cellulase assay) on whole stalks (see Additional file 10 for whole stalk yield equations). Both methods produced comparable results with comparable standard errors (Figure 8) showing the accuracy and robustness of the holistic high-throughput screening methodology. In addition, *bmr6* and *bmr12* mutations have been reported numerous times to result in a yield reduction of 15% to 30% [30,34], which is consistent with the observed total fermentable yield reduction of 19% and 16% in *bmr6* and *bmr12*, respectively (Figure 8).

Using this holistic high-throughput methodology for screening large variable biofuel feedstock populations, once samples were oven dried, three workers with very minimal training were able to process approximately 200 samples per day which encompassed harvesting through to a total fermentable sugar yield calculation. Data on cell wall composition, stalk sucrose, glucose, and fructose concentrations, relative water content, and stalk geometry including plant height was collected in the process. In addition, the majority of the processing time was consumed in the pressing and grinding of internodes, thus automation of these processes has the potential to dramatically increase throughput. The methodology was, however, designed to be transportable to any field site or laboratory, thus the use of non-portable, specialised, or expensive equipment was avoided. In comparison, using a traditional wet chemistry approach, and once samples were accumulated through multiple processes including oven drying and numerous incubation steps, three workers with moderate laboratory skills could process only approximately ten samples per day without collecting any information on cell wall composition or plant geometry. The traditional method required a laboratory set-up and was not easily portable to field sites.

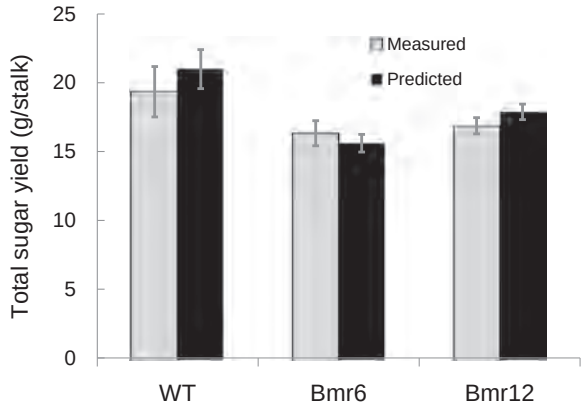


Figure 8 Calculated total fermentable sugar yield in *S. bicolor* mutants. Combined total sugar yield from saccharification of the cell wall fraction and soluble sugar fraction in *S. bicolor bmr6* and *bmr12* mutants harvested at 10 days post-anthesis as determined by wet chemistry of whole stalks (measured) and FTIR predictive models extrapolated from the fourth internode (predicted). Data are represented as means of six biological replicates ± SE. FTIR, Fourier transform infrared; SE, standard error.

Conclusion

FTIR spectral prediction models have been used in the food industry to predict sucrose, glucose, and fructose in fruit and sugarcane juice [17-19] and, more recently, to predict cellulase digestibility of lignocellulose [16]. Here we present a high-throughput screening methodology that unifies and simplifies previous methods to produce a holistic assessment of biofuel feedstock potential via a total fermentable sugar yield calculation. In the process, a wealth of information on cell wall composition, the soluble sugar fraction, and plant geometry was generated rapidly using equipment that is relatively cheap, accessible, easily operated, and transportable to any field site.

Materials and methods

Plant growth conditions

One hundred sweet *S. bicolor* ecotypes were field grown at the Pacific Seeds Gatton Research Centre (Queensland, Australia) in September 2009 and were harvested in February 2010 approximately 10 days post-anthesis. Fifteen genotypes were selected as possible cell wall digestibility variants based on physical characteristics such as stalk strength, stiffness, and leaf midrib colour which were determined subjectively. Subsequently, seed from these 15 selected genotypes, along with the sweet sorghum variety 'Rio,' were sown in a glasshouse in Newcastle, Australia, in May 2010 in 25 cm diameter pots. Plants in Newcastle were supplemented with 300 $\mu mol.m^{-2}.s^{-1}$ photosynthetically active radiation (PAR; 400 to 700 nm) from 4:00 pm to 6:30 pm each day and natural sunlight throughout the winter growth period ranged from 200 to 1,100 $\mu mol.m^{-2}.s^{-1}$ PAR with 28°C/15°C day/night temperatures. Plants were harvested in August 2010. This provided a set of phenotypically variable plants from which predictive models were built. The correlation studies were performed on 15 proprietary sweet sorghum cultivars that were field grown during the 2009 planting at the Pacific Seeds Gatton Research Centre. Descriptive data for these 15 lines is available in Additional file 3. Cell wall digestibility correlations were supplemented with four replicate glasshouse grown 'Rio' plants to replace lost samples.

Biomass assessment

A new high-throughput methodology was developed for screening large variable populations of *S. bicolor* for biomass, soluble sugar concentrations, and cell wall composition using a volumetric correlation between the fourth internode from the base of the stalk that had expanded more than 2 cm, to that of the whole stalk (Figure 1). During harvest of the fourth internode four field measurements were taken: stalk height (H), stalk base diameter (R), stalk top diameter (r), and length from the base of stalk to the base of the fourth internode (H_B). Each internode was cooled then transported to the field station where internode length (L_{SI}) and internode FW were recorded. Each internode was pressed using a Sukra Sugarcane Crusher (S.A. Ivy Multi Pumps Limited, Tamil Nadu, India) and FTIR spectra acquired. Two mL of juice was frozen and stored at –20°C. Pressed *S. bicolor* bagasse was bagged, dried, weighed, and ground. For correlations analysis, whole stalks were stripped of leaves, weighed, pressed, dried, and ground. See Additional file 1 for detailed protocol.

Fourier transform infrared (FTIR) spectroscopy of soluble sugars

Juice samples were centrifuged at 10,000 *g* for 30 seconds before 50 μL was placed onto a Spectrum II (PerkinElmer, Waltham, MA, USA) with a universal diamond ATR attachment operated by Spectrum One software package (PerkinElmer). Absorbance spectra were collected at a resolution of 4 cm^{-1} over two scans from 4,000 to 400 cm^{-1} wavenumbers. Spectra were imported into Unscrambler X 10.0.1 (Camo) and truncated to the fingerprint region. Predictive models were optimised when variable reduction was employed to further truncate spectra to 1,180 to 900 cm^{-1} wavenumbers. A 9-point Savitzky–Golay smoothing algorithm was applied to juice spectra from both the calibration and validation sets (see Additional file 4 for sample catalogues) before model calibration and prediction.

Gas chromatography–mass spectrometry (GC-MS) of soluble sugars

Eleven μL of sorghum juice, plus 5 μL of a 2 mg mL^{-1} ribitol internal standard was lyophilised. Derivatisation/

trimethylsilylation and quantification of sugars by GC-MS were performed following the published procedure [35] with the following modifications. The GC-MS system used was a 5973A MSD (Agilent Technologies, Santa Clara, CA, USA), GC was performed on a 30 m BPX5 column (SGE, Victoria, Australia), and both chromatograms and mass spectra were evaluated using the Enhanced MSD ChemStation D.01.02 software (Agilent Technologies). Mass spectra of eluting sugars were identified using the Wiley Registry of Mass Spectral Data (Wiley-Blackwell, Hoboken, NJ, USA), and all compounds were verified by subsequent analysis of pure standards. Absolute quantification of sugars was achieved by equating normalised areas to a set of standard curves for sucrose, glucose, and fructose.

Cell wall preparation

Dried *S. bicolor* bagasse (pressed internode) was ground to a fine powder (<100 μm) with a TissueLyser II (Qiagen, Venlo, Netherlands) at a frequency of 1/30 for 2 minutes using 1.5 cm diameter stainless steel ball bearings. A total of 150 mg of powder was then washed with 1.7 mL of water and dried in a vacuum concentrator centrifugal evaporator (Jouan RC 10.10, St. Herblain, France) before acquiring FTIR spectra. For model calibration and validation, a series of extraction steps following the published procedure [36] was performed on 120 mg of washed powder before determination of cell wall saccharification efficiency by enzymatic hydrolysis.

FTIR spectroscopy of cell walls

Approximately 50 mg of washed and dried cell wall material was mounted onto a Spectrum II (PerkinElmer) with a universal diamond ATR attachment operated by Spectrum One software package. The cell wall powder was forced onto the ATR crystal with a press at a constant force. Spectra were collected at a resolution of 4 cm^{-1} and co-added over three scans. Once all spectra were collected, data was imported into Unscrambler X 10.0.1 (Camo) for pre-treatment. Spectra were truncated to the cell wall fingerprint region (1,800 to 850 cm^{-1} wavenumbers), second derivative spectra were calculated by the Savitzky–Golay method with nine smoothing points and an EMSC was applied to correct for light scattering by cell wall particles.

3,5-Dinitrosalicylic acid (DNS) determination of cell wall saccharification

Fifteen mg of isolated cell wall was vacuum infiltrated under 100 KPa pressure for 5 minutes in 1.25 mL of 50 mM sodium citrate buffer (pH 4.8) followed by addition of 50 μL of 1 mg/mL *T. reesei* cellulase mixture (Sigma-Aldrich) and a control was included for each sample where cellulase was replaced with buffer. Samples were vortexed briefly and incubated at 55°C with gyrorotation at 250 rpm for 24 hours. Cellulase digested material was centrifuged at 3,500 *g* for 3 minutes. One mL of supernatant was added to 3 mL of DNS reagent and incubated for 15 minutes in a 100°C water bath. One mL of 40% (w/v) potassium tartrate and 20 mL of dH_2O was immediately added to dilute and stabilise the colour. Absorbance was measured at 540 nm and compared to a D-glucose standard curve to determine the total amount of reducing sugars.

Partial least squares (PLS) model calibration and validation

Pre-treated spectra were imported into the Unscrambler X 10.0.1 (Camo) for data analysis. The soluble sugar predictive model used 216 sugar standards to calibrate a PLS regression using the NIPALS algorithm, and the cell wall digestibility predictive model used 58 hand-picked calibration samples (that were representative of the range in digestibility) with the 'wide-kernel PLS' algorithm [37]. All models used mean centred data, computed a default of 15 factors, and used full cross-validation. Outliers were identified using residual and leverage plots and were manually examined to ensure that they were outliers based on technical errors in spectral collection or wet chemistry before removal, and then PLS algorithms were re-calculated. The optimal number of factors was chosen as that which minimised the RMSEV and maximised the amount of sample variation that was described by the model. The performance of models must be continually checked against wet chemistry and continually updated with biological extremes to increase the range of samples accurately predicted by the model, therefore samples that lie outside of the range of the model should be analysed and incorporated into the calibration set. The models presented here are, therefore constantly being updated and improved as more samples are screened and this is an integral part of the methodology of chemometric model development.

PLS models were validated by prediction on an independent test set of 33 soluble sugar samples and 25 cell wall samples (see Additional files 4 and 7 for calibration and validation sample catalogues). Predicted values were plotted against reference values and the RMSEP and validation coefficient of determination (Q^2) were used to assess the accuracy of the models. The analysis of *bmr6*, *bmr12*, and their parent line (WT) was performed independently using the methods outlined above.

Additional files

Additional file 1: High-throughput screening protocol. A detailed step-by-step protocol for performing the described high-throughput screening platform.

Additional file 2: Volumetric ratio derivation. A mathematical derivation of the volumetric ratio used to relate measurements on the fourth internode to whole stalks.

Additional file 3: Sweet sorghum lines used for correlation analysis. Fresh weight (FW), dry weight (DW), and height data for the 15 sweet sorghum lines used in the correlations analysis. DW, dry weight; FW, fresh weight.

Additional file 4: Calibration and validation samples used for the sugar PLS models. Sucrose, glucose, and fructose content of the calibration and validation sample sets used for calibrating and validating the sugar FTIR PLS models. FTIR, Fourier transform infrared; PLS, partial least squares.

Additional file 5: Sugar model regression coefficients. Weighted regression coefficients from the sucrose, glucose, and fructose PLS models showing that regression coefficients resemble pure standard spectra of each sugar. PLS, partial least squares.

Additional file 6: Spike/dilution recovery and limit of detection of the FTIR PLS sugar models. Limit of detection and spike/dilution recovery of sucrose, glucose, and fructose using the FTIR PLS models for each sugar. FTIR, Fourier transform infrared; PLS, partial least squares.

Additional file 7: Calibration and validation samples used for the digestibility PLS model. Sorghum lines and sampled tissue used for calibrating and validating the digestibility PLS model. PLS, partial least squares.

Additional file 8: Processed spectra used to calibrate the digestibility PLS model. Second derivative spectra with an EMSC applied, which were used to calibrate the digestibility PLS model. EMSC, extended multiplicative scatter correction; PLS, partial least squares.

Additional file 9: PLS digestibility model diagnostics and band assignment chart. Model diagnostics for the PLS digestibility model showing clear separation of digestibility in the scores plot and a representation of cell wall peaks in the regression coefficients. A band assignment chart is displayed for reference. PLS, partial least squares.

Additional file 10: Whole stalk fermentable sugar calculations. Whole stalk calculations for fermentable sugars in the soluble sugar fraction and cell wall fraction resulting in a total fermentable sugar yield calculation.

Abbreviations

ATR: Attenuated total reflectance; *Bmr*: Brown midrib; DNS: 3,5-Dinitrosalicylic acid; DW: Dry weight; EMSC: Extended multiplicative scatter correction; FIR: Far-infrared; FTIR: Fourier transform infrared; FT-MIR: Fourier transform mid-infrared; FW: Fresh weight; GC-MS: Gas chromatography–mass spectrometry; HPLC: High-performance liquid chromatography; IR: Infrared; LoD: Limit of detection; LV: Latent variable; MIR: Mid-infrared; NIPALS: Non-linear iterative partial least squares; NIR: Near-infrared; NREL: National renewable energy laboratory; PAR: Photosynthetically active radiation; PDA: Personal digital assistant; PLS: Partial least squares; RMSEC: Root mean squared error of calibration; RMSEP: Root mean squared error of prediction; RMSEV: Root mean squared error of validation; WT: Wild type.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

APM theorised the sugar and cell wall FTIR models, and collaboratively theorised the internode correlation study with WP. APM designed all experiments, collected all data, analysed all data, produced Figures 1, 2, 3, 4, 5, 6, 7, and 8, and drafted the manuscript. WP designed and collaborated on the sugar model data collection and analysis, including volumetric internode correlation experiments, and produced Figures 1, 2, 3, 4, and 5. CB reviewed experimental design and provided feedback on the digestibility model. CPLG conceived and established the research program and contributed to overall experimental design. RTF contributed to experimental design and reviewed the manuscript. All authors read and approved the final manuscript.

Acknowledgments

AM is the recipient of an Office of the Chief Executive (OCE) PhD scholarship from the Commonwealth Scientific and Industrial Research Organisation (CSIRO). We thank the Australian Research Council (ARC) and Pacific Seeds for financial support under ARC's Linkage Projects funding scheme (project number: LP0883808). The authors would like to thank Simon Wheeler, Amalia Belgeri, and Gretel Linich for assistance in collection of samples and Joseph Enright for maintenance of plants at the Plant Growth Facility, University of Newcastle (Callaghan, Australia). The authors would also like to thank Dr Rosemary White and the High Resolution Plant Phenomics Centre (HRPPC; Canberra, Australia) for operational assistance and access to the FTIR instrument.

Author details

[1]School of Environmental and Life Sciences, University of Newcastle, University Drive, Callaghan NSW 2308, Australia. [2]Australian Research Council Centre of Excellence in Plant Cell Walls, Waite Research Institute, University of Adelaide, Glen Osmond, SA 5064, Australia. [3]CSIRO Plant Industry, High Resolution Plant Phenomics Centre, GPO Box 1600, Canberra ACT 2601, Australia.

Received: 12 August 2013 Accepted: 10 December 2013
Published: 23 December 2013

References

1. Möller R: *Cell Wall Saccharification. Outputs from the EPOBIO Project, November 2006.* Newbury: CPL Press; 2006.
2. Decker S, Brunecky R, Tucker M, Himmel M, Selig M: **High-throughput screening techniques for biomass conversion.** *BioEnergy Res* 2009, **2:**179–192.
3. Chundawat SP, Balan V, Dale BE: **High-throughput microplate technique for enzymatic hydrolysis of lignocellulosic biomass.** *Biotechnol Bioeng* 2008, **99:**1281–1294.
4. Gomez L, Whitehead C, Barakate A, Halpin C, McQueen-Mason S: **Automated saccharification assay for determination of digestibility in plant materials.** *Biotechnol Biofuels* 2010, **3:**23.
5. Santoro N, Cantu SL, Tornqvist C-E, Falbel TG, Bolivar JL, Patterson SE, Pauly M, Walton JD: **A High-throughput platform for screening milligram quantities of plant biomass for lignocellulose digestibility.** *Bioenerg Res* 2010, **3:**93–102.
6. Carpita NC: **Structure and biogenesis of the cell walls of grasses.** *Annu Rev Plant Physiol Plant Mol Biol* 1996, **47:**445–476.
7. Byrt CS, Grof CPL, Furbank RT: **C4 plants as biofuel feedstocks: optimising biomass production and feedstock quality from a lignocellulosic perspective.** *J Integr Plant Biol* 2011, **53:**120–135.
8. Doblin MS, Pettolino F, Bacic A: **Plant cell walls: the skeleton of the plant world.** *Funct Plant Biol* 2010, **37:**357–381.
9. Carpita NC, Gibeaut DM: **Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth.** *Plant J* 1993, **3:**1–30.
10. Griffiths PR, de Haseth JA: *Fourier Transform Infrared Spectrometry.* 2nd edition. John Wiley & Sons Inc: Hoboken, NJ; 2007.
11. Morikawa H, Senda M: **Infrared analysis of oat coleoptile cell walls and oriented structure of matrix polysaccharides in the walls.** *Plant and Cell Physiology* 1978, **19:**327–336.
12. Brown DM, Zeef LA, Ellis J, Goodacre R, Turner SR: **Identification of novel genes in arabidopsis involved in secondary cell wall formation using expression profiling and reverse genetics.** *Plant Cell* 2005, **17:**2281–2295.
13. Mouille G, Robin S, Lecomte M, Pagant S, Hofte H: **Classification and identification of *Arabidopsis* cell wall mutations using Fourier-Transform InfraRed (FT-IR) microscopy.** *Plant J* 2003, **35:**393–404.
14. Carpita NC, Defernez M, Findlay K, Wells B, Shoue DA, Catchpole G, Wilson RH, McCann MC: **Cell wall architecture of the elongating maize coleoptile.** *Plant Physiol* 2001, **127:**551–565.
15. McCann MC, Defernez M, Urbanowicz BR, Tewari JC, Langewisch T, Olek A, Wells B, Wilson RH, Carpita NC: **Neural network analyses of infrared spectra for classifying cell wall architectures.** *Plant Physiol* 2007, **143:**1314–1326.
16. Sills DL, Gossett JM: **Using FTIR to predict saccharification from enzymatic hydrolysis of alkali-pretreated biomasses.** *Biotechnol Bioeng* 2012, **109:**353–362.

17. Duarte IF, Barros A, Delgadillo I, Almeida C, Gil AM: **Application of FTIR spectroscopy for the quantification of sugars in mango juice as a function of ripening.** *J Agric Food Chem* 2002, **50:**3104–3111.
18. Irudayaraj J, Tewari J: **Simultaneous monitoring of organic acids and sugars in fresh and processed apple juice by Fourier transform infrared attenuated total reflection spectroscopy.** *Appl Spectrosc* 2003, **57:**1599–1604.
19. Cadet F: **Measurement of sugar content by multidimensional analysis and mid-infrared spectroscopy.** *Talanta* 1999, **48:**867–875.
20. Wang J, Kliks MM, Jun S, Jackson M, Li QX: **Rapid analysis of glucose, fructose, sucrose, and maltose in honeys from different geographic regions using Fourier transform infrared spectroscopy and multivariate analysis.** *J Food Sci* 2010, **75:**C208–C214.
21. Rodriguez-Saona LE, Fry FS, McLaughlin MA, Calvey EM: **Rapid analysis of sugars in fruit juices by FT-NIR spectroscopy.** *Carbohydr Res* 2001, **336:**63–74.
22. Haffner F, Mitchell V, Arundale R, Bauer S: **Compositional analysis of *Miscanthus giganteus* by near infrared spectroscopy.** *Cellulose* 2013, **20:**1629–1637.
23. Wolfrum E, Payne C, Stefaniak T, Rooney W, Dighe N, Bean B, Dahlberg J: *Multivariate Calibration Models for Sorghum Composition Using Near-Infrared Spectroscopy*, Technical Report NREL/TP-510056838. Golden, CO: National Renewable Energy Laboratory (NREL); 2013.
24. Muttoni G, Johnson J, Santoro N, Rhiner C, von Mogel KJ, Kaeppler S, de Leon N: **A high-throughput core sampling device for the evaluation of maize stalk composition.** *Biotechnol Biofuels* 2012, **5:**27.
25. Golzarian M, Frick R, Rajendran K, Berger B, Roy S, Tester M, Lun D: **Accurate inference of shoot biomass from high-throughput images of cereal plants.** *Plant Methods* 2011, **7:**2.
26. Wu X, Staggenborg S, Propheter JL, Rooney WL, Yu J, Wang D: **Features of sweet sorghum juice and their performance in ethanol fermentation.** *Ind Crop Prod* 2010, **31:**164–170.
27. Tamaki Y, Mazza G: **Rapid determination of carbohydrates, ash, and extractives contents of straw using attenuated total reflectance Fourier transform mid-infrared spectroscopy.** *J Agric Food Chem* 2011, **59:**6346–6352.
28. Barrière Y, Ralph J, Méchin V, Guillaumie S, Grabber JH, Argillier O, Chabbert B, Lapierre C: **Genetic and molecular basis of grass cell wall biosynthesis and degradability. II. Lessons from brown-midrib mutants.** *C R Biol* 2004, **327:**847–860.
29. Dien BS, Sarath G, Pedersen JF, Sattler SE, Chen H, Funnell-Harris DL, Nichols NN, Cotta MA: **Improved sugar conversion and ethanol yield for forage sorghum (Sorghum bicolor L. Moench) lines with reduced lignin contents.** *BioEnergy Research* 2009, **2:**153–164.
30. Sattler SE, Funnell-Harris DL, Pedersen JF: **Brown midrib mutations and their importance to the utilization of maize, sorghum, and pearl millet lignocellulosic tissues.** *Plant Sci* 2010, **178:**229–238.
31. Liu L, Ye XP, Womac AR, Sokhansanj S: **Variability of biomass chemical composition and rapid analysis using FT-NIR techniques.** *Carbohydr Polym* 2010, **81:**820–829.
32. Wolfrum E, Sluiter A: **Improved multivariate calibration models for corn stover feedstock and dilute-acid pretreated corn stover.** *Cellulose* 2009, **16:**567–576.
33. Sluiter A, Hames B, Ruiz R, Scarlata C, Sluiter J, Templeton D, Crocker D: *Determination of Structural Carbohydrates and Lignin in Biomass. Technical Report NREL/TP-510-42618.* Golden, CO: National Renewable Energy Laboratory (NREL); 2011.
34. Pedersen JF, Vogel KP, Funnell DL: **Impact of reduced lignin on plant fitness.** *Crop Sci* 2005, **45:**812–819.
35. Roessner U, Patterson JH, Forbes MG, Fincher GB, Langridge P, Bacic A: **An investigation of boron toxicity in barley using metabolomics.** *Plant Physiol* 2006, **142:**1087–1101.
36. Foster CE, Martin MT, Pauly M: **Comprehensive compositional analysis of plant cell walls (lignocellulosic biomass) part I: Lignin.** *J Vis Exp* 2010, **37:**36–41.
37. Rännar S, Lindgren F, Geladi P, Wold S: **A PLS kernel algorithm for data sets with many variables and fewer objects. Part 1: theory and algorithm.** *J Chemometr* 1994, **8:**111–125.

doi:10.1186/1754-6834-6-186
Cite this article as: Martin *et al.*: A holistic high-throughput screening framework for biofuel feedstock assessment that characterises variations in soluble sugars and cell wall composition in *Sorghum bicolor*. *Biotechnology for Biofuels* 2013 **6**:186.

3.3 Publication 3 (draft manuscript): Detection of cell wall variants using the high-throughput screening framework

Martin AP, Palmer WM, Brown C, Byrt CS, Furbank RT, Grof CPL **Detection and characterisation of *Sorghum bicolor* cell wall variants using FTIR spectra generated from a high-throughput screening framework** *Biotechnology for Biofuels* Under review for re-submission (2015)

Statements of contribution of others: Publication 3

I attest that Research higher degree candidate ***Antony Martin*** *was the primary contributor to the development of this publication. The author contributions were as follows:*

APM theorised the cell wall cluster model, designed all experiments, collected all data, analysed all data, produced all figures and tables (excluding figure 4), and drafted the manuscript. WMP also collected sorghum samples and FTIR spectra, produced figure 4 and assisted in wet chemistry analysis of cell walls. CB assisted in wet chemistry analysis of cell walls. CSB and CPLG designed and performed the generation of the EMS mutant population. CPLG established the Sorghum research program. CPLG, CSB and RTF reviewed the manuscript. All authors read and approved the final manuscript.

_____

Antony Martin

_____

William Palmer

_____

Christopher Brown

_____

Caitlin Byrt

_____

Robert Furbank

_____

Christopher Grof

Detection and characterisation of *Sorghum bicolor* cell wall variants using FTIR spectra generated from a high-throughput screening framework

Antony P. Martin[1] (antony.paul.martin@gmail.com)

William M. Palmer[1] (william.moreau.palmer@gmail.com)

Christopher Brown[1] (christopher.brown@uon.edu.au)

Caitlin S. Byrt[1,2] (caitlin.byrt@adelaide.edu.au)

Robert T. Furbank[3] (robert.furbank@csiro.au)

Christopher P.L. Grof[1*] (chris.grof@newcastle.edu.au)

[1]School of Environmental and Life Sciences, University of Newcastle, University Drive, Callaghan, NSW 2308, Australia

[2]Australian Research Council Centre of Excellence in Plant Cell Walls, Waite Research Institute, University of Adelaide, Glen Osmond, SA 5064, Australia

[3]CSIRO Plant Industry and High Resolution Plant Phenomics Centre, GPO Box 1600, Canberra, ACT 2601, Australia

***indicates corresponding author**

Abstract

Background: The identification of cell wall variants in genetically modified, mutant, and breeding populations will be integral to furthering our understanding of the relationship between cell wall recalcitrance, cell wall composition, and crop yields in a lignocellulosic biofuel context. A holistic high-throughput (HTP) screen was developed which, in part, used FTIR spectra to quantify the digestibility of the cell wall fraction of *S. bicolor* stalks. Here we used the spectra collected from this HTP screen to identify and qualitatively characterise cell wall variants using *Bmr6* and *Bmr12* mutants with altered cell wall lignin as a proof of concept and then applying this to a selection of 100 EMS mutant sorghum lines from a segregating M3 population.

Results: Principle component analysis (PCA) of FTIR spectra collected from a HTP screen showed that type I and type II cell walls formed distinct clusters as did different plant tissues from *S. bicolor* and that this was based on changes in FTIR spectral peaks. This led to the development of a hierarchical classification model which accurately classified both type I & II cell walls and repeatedly classified the well established *Bmr6* and *Bmr12* mutants with altered lignin into a 'lignin deficient' clade. Similar hierarchical classification models were built on spectra collected from a selection of sorghum mutant lines and 6/100 lines formed distinct clusters away from the parent cluster. Cell wall composition was shown to be altered in all 6 lines and confirmed that the FTIR spectral classification model was able to cluster into low and high lignin clades, and an altered glucan monomer clade, likely to be indicative of a reduction in cellulose content.

Conclusions: FTIR spectra collected from a HTP screening framework designed for assessing biofuel feedstocks were used to accurately classify and partially characterise *S. bicolor* cell wall mutants without the need for calibrating quantitative prediction models. The fact that FTIR spectra could also distinguish cell wall differences in a range of species and plant organs/tissues suggests that this approach could be applied to cell wall variant detection in any plant species or tissue.

Keywords: FTIR, cell wall, hierarchical modelling, high-throughput screening, sorghum, biofuel, digestibility

Background

A typical plant cell maintains a primary cell wall throughout cell expansion followed by the deposition of a secondary cell wall (often in three distinct layers: S1, S2 & S3), which is usually thickened by lignification and/or suberisation [1]. The secondary cell wall contains the greatest proportion of biomass [2], thus it is of most interest as a renewable source of carbon for bioenergy. However, an in-depth understanding of the synthesis, structure and composition of the cell wall as a whole will be integral to engineering its deconstruction for biofuel production.

The basic principles of cell wall synthesis and deposition are relatively well conserved between type I and II cell walls. The composition and molecular interactions between components, however, is different, particularly in the hemicellulose fraction [3, 4]. Type II walls are of most interest to lignocellulosic bioethanol production with monocot grass species such as maize, wheat, rice, sugarcane and sorghum accounting for the majority of world biomass production. Type II primary cell walls are typically composed of 20-30% cellulose in the form of cellulose microfibrils, 35-70% hemicelluloses, ~5% pectin, ~1% structural glycoproteins, 1-5% phenolics and a small percentage of cell wall enzymes [5]. This differs most prominently from the type I cell wall in the amount of pectin (reduced in type II cell walls) and the amount and type of hemicellulose [3, 5, 6]. Most significant is the type of hemicellulose, which in type I cell walls is predominantly xyloglucan, but in type II cell walls is predominantly glucuronoarabinoxylans (GAX) [3, 5]. As opposed to the primary cell wall, secondary cell wall composition is designed for strength and rigidity, and its composition reflects this. A typical grass secondary cell wall is composed of 35-45% cellulose embedded in a matrix consisting of 40-50% GAX, 20% lignin, 5-15% silica, 0.5-1.5% phenolics (i.e. ferulic acid & ρ-coumaric acid) and 0.1% pectin [5, 7]. Since secondary cell walls are not required to expand (a very complex process), much of the complexity of the primary cell wall is absent. This can be seen with the absence of mixed linkage glucans, reduced amount of pectin, and the fact that the GAX is mostly present as low substituted GAX (as opposed to a mixture of low and high substitution), in the secondary cell wall of grasses [5].

To generate bioethanol from lignocellulosic material, the cell wall must first be treated to

increase porosity and break cross-links between components, thus increasing accessibility of enzymes to cellulose [1, 8]. During secondary cell wall deposition, lignin content, composition, and structural integrity have been strongly implicated in cell wall digestibility due to the ability of lignin to 'shield' cellulose from enzyme degradation by forming cross-links to polysaccharides and forming a hydrophobic layer around cellulose microfibrils [1, 4, 9-11]. The *S. bicolor* brown midrib (*Bmr*) mutants *Bmr6* and *Bmr12* are a well characterised example where reduced lignin content results in increased digestibility [12]. Here we show that FTIR spectra generated from a high throughput (HTP) screening framework [13] was able to detect cell wall variants within a variable *S. bicolor* population and could classify them based on cell wall chemistry. Classification of *A. thaliana* cell wall variants has previously been shown [14], but in largely undifferentiated tissue of young hypocotyls prior to substantial lignification, as has classification of young developing *Z. mays* coleoptiles [15]. This is the first example, to our knowledge, of classification using FTIR spectra in mature secondary cell walls of a C_4 grass population in a HTP screening context for biofuel application.

Results and discussion

1. FTIR spectra differentiate type I and type II cell walls from different species and plant tissues

To demonstrate that cell wall FTIR spectra generated from a HTP screening framework [13] were able to detect variations in cell wall chemistry, a variety of cell walls with well established structural differences were analysed. These included type I cell walls from tomato (*Solanum lycopersicum*), tobacco (*Nicotiana tobaccum*), and *Arabidopsis thaliana;* type II cell walls from *S. bicolor* and *Setaria viridis*; and the unique cotton fibre cell wall which consists of ~95% cellulose and no lignin [16]. An extended multiplicative scatter correction (EMSC) was applied to adjust for light scattering due to particle size, and the second derivate was calculated and plotted (**Figure 1A**). As expected, clear spectral differences were observed. Since FTIR spectra are multivariate in nature, principal component analysis (PCA) was used to reduce the data to 2 principal components (PC's) that described 80% of the spectral variability. Type I and type II cell walls from multiple species

separated along PC1 and PC2 distinctly as did the unique cotton fibre cell wall (**Figure 1C**) indicating that the FTIR spectrum is suitably robust for differentiating cell wall types.

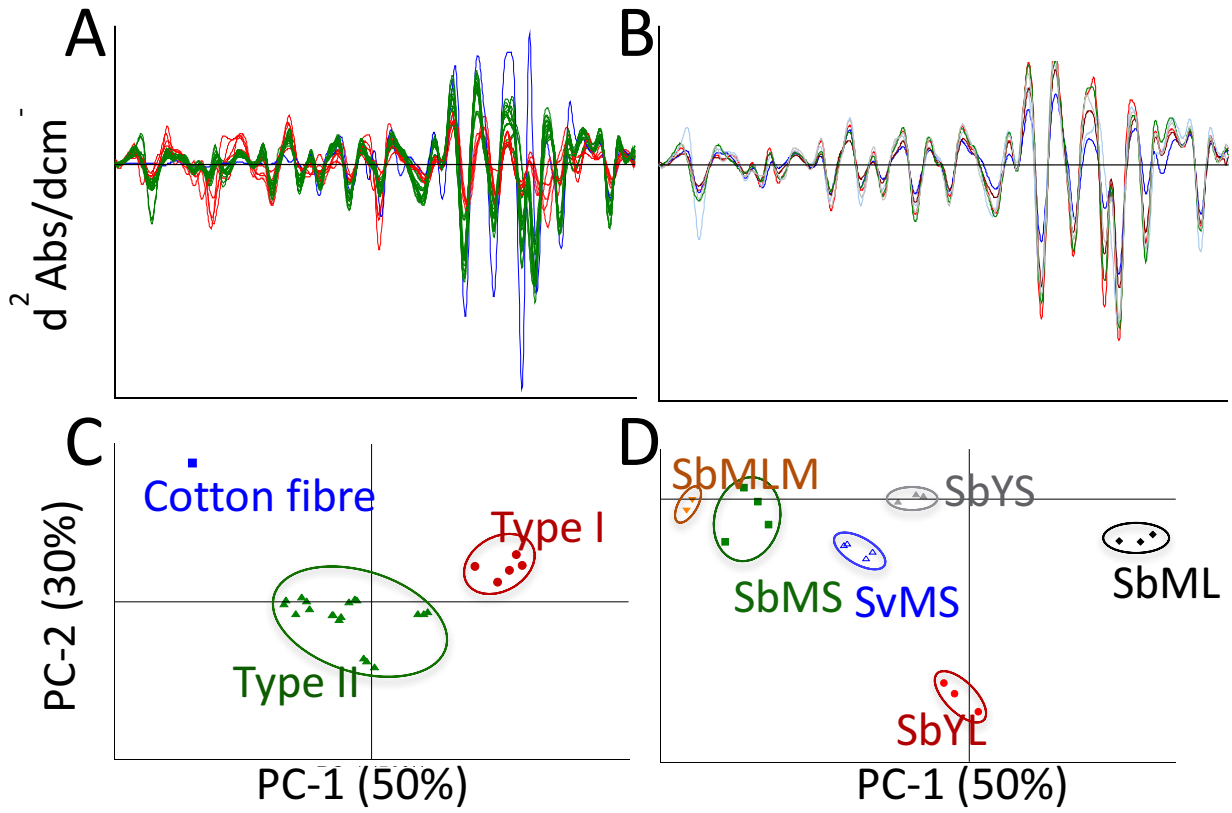


Figure 1. Spectral differentiation of type I and type II cell walls from a variety of species and plant organs/tissues. (**A**) Second derivate spectra of type I cell walls (red; *Solanum lycopersicum* (young stem, young leaf, and mature leaf), *Nicotiana tobaccum* and *Arabidopsis thaliana* young whole plant), type 2 cell walls (green) and the unique cell wall of mature cotton fibre (blue). (**B**) PCA of cell wall spectra showing separation between type I (red), type II (green) and cotton fibre (blue) cell walls. (**C**) Second derivative spectra of type II cell walls (*S. bicolor* (mature leaf midrib (orange; SbMLM), mature stem (green; SbMS), young stem (grey; SbYS), young leaf (red; SbYL) and mature leaf (black; SbML)) and *Setaria viridis* (mature stem (blue; SvMS)). (**D**) PCA of type II cell walls showing separation between species and plant tissues.

To further explore the capability of FTIR spectra to distinguish cell wall differences in organs, type II cell walls from young and mature stems and leaves were analysed. Second derivative spectra with an EMSC applied were plotted (**Figure 1B**) and PCA was performed (**Figure 1D**). Clear spectral differences were observed and this was reflected in the PCA plot which, once again, showed distinct separation into clusters along PC1 and PC2 for each plant tissue suggesting that the FTIR spectra were able to distinguish more subtle cell wall differences of different organ/tissue types within *S. bicolor*. Interestingly, spectra derived from stem samples from both *S. bicolor* and *S. viridis* clustered together suggesting that these genetically similar species [17] have very similar cell wall composition.

2. FTIR cell wall cluster model produced a lignin deficient clade

To further explore the classification ability of cell wall FTIR spectra generated from a HTP screening framework of a variable collection of *S. bicolor* genotypes, an unsupervised hierarchical classification model based on the squared Euclidean distance between samples was generated. Typical type I cell walls (*N. tabacum*, *S. lycopersicum* and *A. thaliana* cultivar Col-0) were included as controls to demonstrate the classification ability of the cluster model. Similarly, an *A. thaliana* T-DNA insertional mutant line *ref8* (SALK_048706C; AT2G40890) which has been reported to have reductions in, and an altered form of lignin [18] was included as a control since clustering of *A. thaliana* cell wall mutants using FTIR spectra has been demonstrated previously [14].

The cluster model firstly classified type I and type II cell walls in two distinct clades whilst the unique cotton fibre cell wall clustered separately. In addition, all 3 biological replicates of the *ref8* mutant clustered separately from the Col-0 *A. thaliana* cell walls. The reduced lignin phenotype of the *Ref8* T-DNA insertional line was confirmed by a simple stain before FTIR spectra were collected (**Additional file 1**).

Within the collection of variable *S. bicolor* genotypes, cell walls clustered accurately with biological replicates of *Rio*, *Bmr12*, *Bmr6*, and their parent line, WT *AF805*, clustering into distinct groups. Previous studies have shown that *S. bicolor Bmr6* and *Bmr12* mutants have reduced cinnamyl alcohol dehydrogenase (CAD) and caffeic acid/5-hydroxyferulic acid O-methyl transferase (COMT) activity respectively, resulting in reduced and altered forms of cell wall lignin [12, 19, 20]. The hierarchical cluster model was able to cluster *Bmr6* and

Bmr12 separately from their parent line (WT *AF805*) in a 'lignin-deficient' clade. Interestingly, *Bmr6* clustered together with *OMR-SC1201*. The uncharacterized mutant *OMR-SC1201* was observed to have brown colored midribs (**Additional file 2**) suggesting it harbors an equivalent mutation to that of *Bmr6*.

The cluster model was therefore able to cluster a range of cell wall types into distinct clades, and within a set of variable *S. bicolor* cell walls was able to form a clade of lignin deficient mutants that separated from their parent line. This provided a simple analytical method for detection of cell wall variants from a HTP screening framework which did not require the calibration of quantitative predictive models and could be applied to any plant species or tissue. Here we have used the approach to identify a lignin deficient clade, and as more genotypes are screened, samples that fall into this clade will be putatively classed as 'lignin deficient' awaiting further analysis. Similarly, clades representing other cell wall modifications could be characterized by detailed chemical analysis of a select number of lines from that clade to infer cell wall modifications characteristic of the clade.

Before undertaking detailed chemical analysis however, the FTIR spectra could be used to attribute certain cell wall modifications to a particular clade. As a proof of concept, the FTIR spectra of the well established lignin modifications of *Bmr6*, and *Bmr12* cell walls were analysed (**Figure 3**).

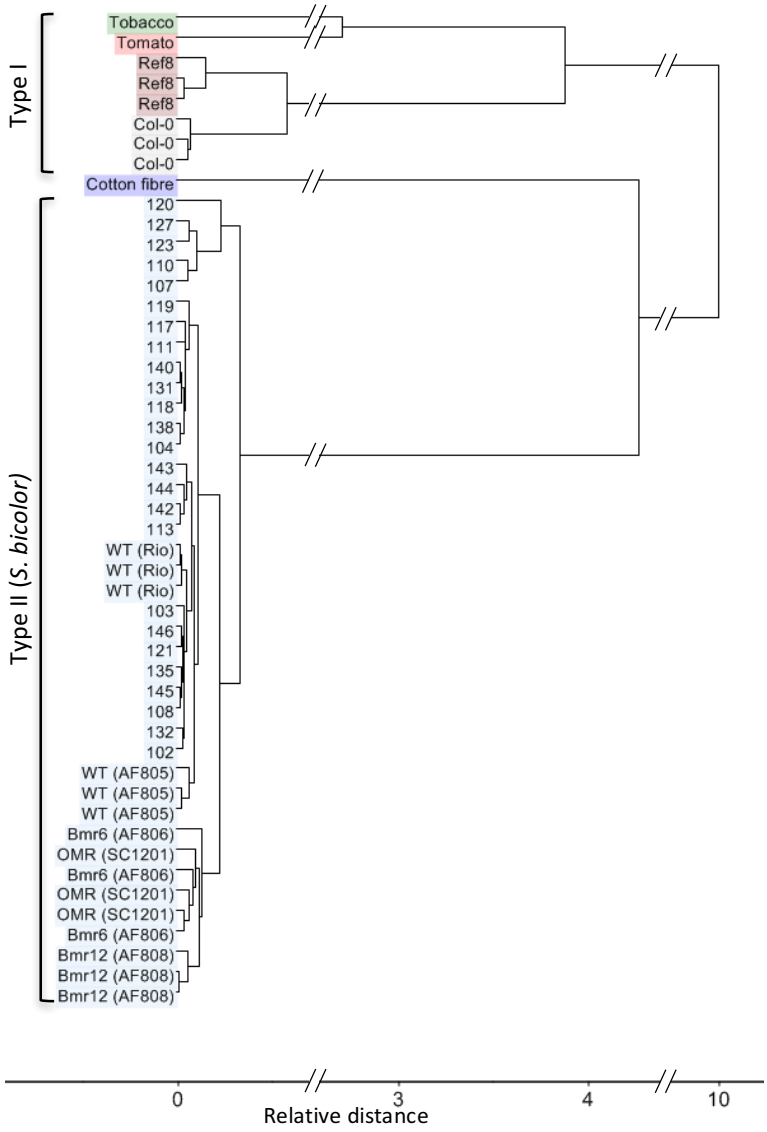


Figure 2. Cell wall hierarchical cluster dendrogram generated using the squared Euclidean distance between 2nd derivate FTIR spectra to determine relative distance between cell wall types. Six clusters were defined by relative distance to their branch point and colour coded. These clusters were *S. bicolor*, mature cotton fibre, *A. thaliana* Col-0, *A. thaliana Ref8* T-DNA insertional mutant of Col-0, *S. lycopersicum,* and *N. tabacum*. Type I and type II cell walls are marked, as is the 'lignin deficient' clade. Numbers 102-146 represent distinct proprietary *S. bicolor* genotypes.

3. Spectral analysis of Bmr6 and Bmr12 cell walls

It is well established that both *Bmr6* and *Bmr12* mutations result in reduced total cell wall lignin content [19, 20] with a small corresponding increase in cellulose and hemicellulose [21]. The *Bmr6* mutation results in a reduction in total lignin with a H:S:G lignin ratio similar to that of WT plants [12, 19, 20]. The *Bmr12* mutation results in a less severe reduction in total lignin, however S lignin units are particularly reduced compared to WT plant [12, 19, 20].

Three spectra were collected from each of three replicate plants for WT, *Bmr6* and *Bmr12* (**Figure 3A**) and a PCA was performed (**Figure 3B**). WT, *Bmr6*, and *Bmr12* formed tight clusters which clearly separated the three cell wall types (**Figure 3B**). Firstly, to examine the reduction in total lignin that is more severe in *Bmr6* than *Bmr12*, the peak at 1604 cm^{-1} that is attributed to the α-β double bond of the propanoid side group in lignin-like structures [22-24] was used, since this bond is present in all lignin structures. As expected, it was observed that both *Bmr6* and *Bmr12* had a reduction in absorbance at 1604 cm^{-1} with the reduced absorbance being more severe in *Bmr6* than in *Bmr12* (**Figure 3A**). To specifically examine the reduction in G lignin units that is more severe in *Bmr6* than in *Bmr12*, the peak at 1515 cm^{-1} associated with the C=C bond of the guaiacyl ring of lignin [22, 23] was used. As expected, both *Bmr6* and *Bmr12* had a reduction in absorbance at 1515 cm^{-1} with the reduced absorbance being more severe in *Bmr6* than in *Bmr12* (**Figure 3A**). To specifically examine the reduction in S lignin units that is more severe in *Bmr12* than in *Bmr6*, the peak at 1463 cm^{-1} attributed to the C-H deformation of syringyl lignin [22-24] was used. As expected, both *Bmr6* and *Bmr12* had a reduction in absorbance at 1463 cm^{-1} with the reduced absorbance being more severe in *Bmr12* than in *Bmr6* (**Figure 3A**). To our knowledge, no specific spectral peaks have been assigned to H lignin units, however, it is known that both *Bmr6* and *Bmr12* have equivalent reductions in H lignin units [19, 20]. It is speculated that the peak at 1587 cm^{-1} could be attributed to H lignin since both *Bmr6* and *Bmr12* show equivalent reductions in absorbance at 1587 cm^{-1}.

This analysis confirmed the clustering of *Bmr6* and *Bmr12* separately from each other and WT and that these clusters were based on cell wall chemistry, suggesting that a similar

approach could be taken to identify clusters with other cell wall modifications. This also provided experimental evidence for the peak at 1604 cm^{-1} being indicative of total lignin, the peak at 1515 cm^{-1} being indicative of G lignin units, and the peak at 1463 cm^{-1} being indicative of S lignin units in *S. bicolor* cell walls.

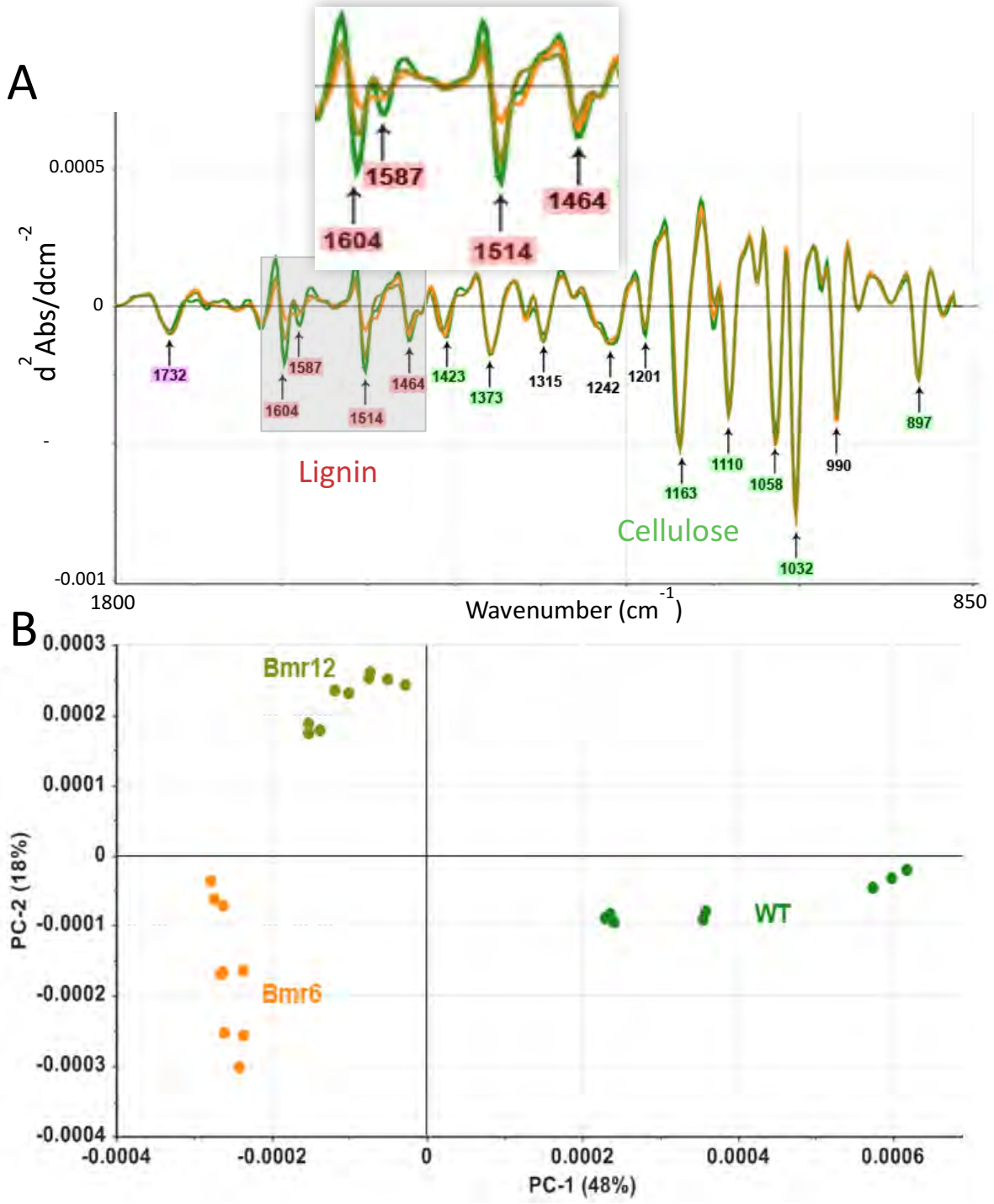


Figure 3. FTIR spectral analysis of *Bmr6* and *Bmr12* cell walls. (**A**) Average second derivative FTIR spectra collected from isolated cell walls of ground mature stem tissue from *Bmr6*,

Bmr12 and their parent line (WT) with the lignin absorption region of the spectra expanded and inset. (**B**) Principle component analysis (PCA) of the spectra. 3 spectra were collected from each of 3 replicate plants for PCA analysis, all units are arbitrary and percentages represent the amount of variability described by that PC. Peaks attributed to lignin, cellulose, and hemicellulose are coloured red, green, and pink respectively.

4. Screening of a segregating M3 EMS sweet sorghum mutant population

To show that this method of cell wall classification is applicable to novel, uncharacterised populations of *S. bicolor*, FTIR spectra from 100 EMS mutants were pre-treated as previously, and classified. A large cluster containing 8 parent lines and 85 EMS mutant lines formed, indicating a cell wall composition that was un-altered by the chemical mutagenesis. Four that differed from the parent cluster, containing a total of 7 *S. bicolor* EMS lines, were identified. Line M454 formed cluster 1, lines M865 and M744 formed cluster 2, lines M720 and M379 formed cluster 3, and lines M419 and M014 formed cluster 4 (**Figure 4**).

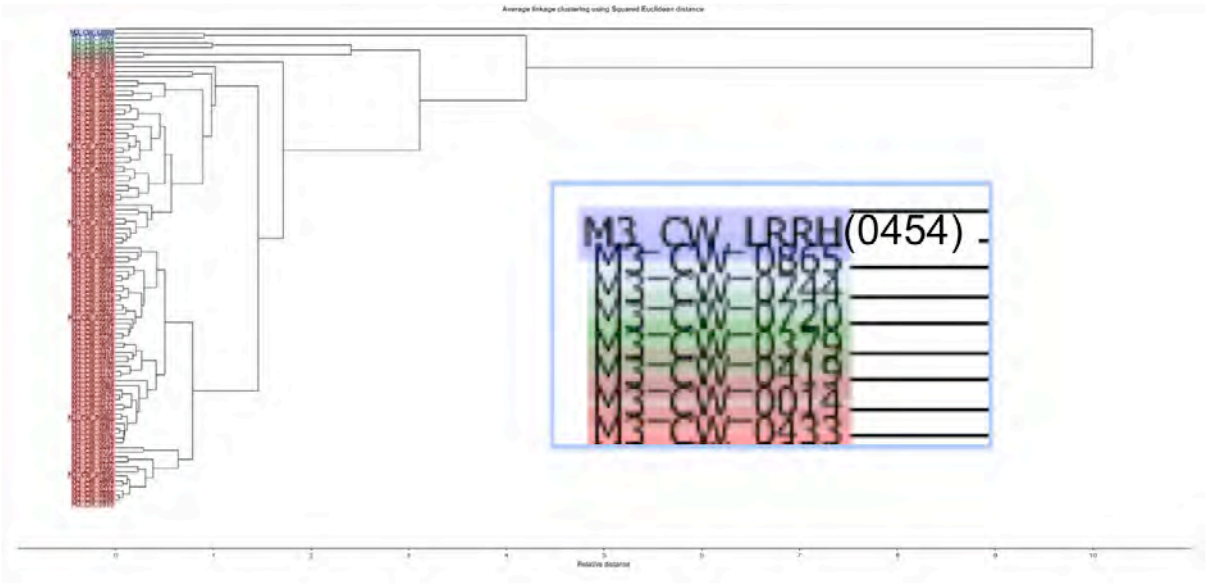


Figure 4. Cell wall hierarchical cluster dendrogram generated from 100 segregating M3 EMS sweet sorghum lines Five classification clusters were defined by their relative distance to the branch point and were color coded. Out-dented samples are parent lines and shades of blue, green and brown indicate clusters that differ from the parent clade (red). Lines that were classified differently from the parent cluster are indented.

To show that these lines which clustered separately from the parent lines were, in fact, cell

wall variants, and to characterise these variations, klason lignin and the monosaccharide composition of the cell walls were measured. In comparison to the parent line, cluster 1 (M454) showed a 48 % increase in klason lignin content, cluster 2 (M865 and M744) showed a 16 % reduction in glucan monosaccharides (likely indicative of cellulose), cluster 3 (M720 and M379) showed a 16 % reduction in klason lignin, and cluster 4 (M419 and M014) showed a 26 % increase in klason lignin (**table 1**). The consistent cell wall phenotypes observed in each clade supports the applicability of the FTIR based classification system presented here. Cell wall variants of lignin and cell wall glucan were detected from a M3 segregating population of uncharacterized *S. bicolor* EMS mutants.

Table 1. Monosaccharide and klaxon lignin content of cell walls in the mutagenised parent line (mean ± SE, n=7) and a selection of individual segregating M3 plants identified from their FTIR spectra.

	Glucan (%)	Xylan (%)	Galactan (%)	Arabinan (%)	Mannan (%)	Lignin (%)
Parent	36.81 ± 1.45	17.00 ± 1.15	1.35 ± 0.21	2.72 ± 0.12	3.10 ± 0.09	20.01 ± 0.45
M379	35.55	14.80	1.46	2.58	2.97	16.55
M720	29.89	14.91	1.59	2.85	2.77	16.87
M014	42.39	17.22	1.62	2.93	4.31	25.82
M419	31.73	14.29	1.15	2.18	2.50	24.75
M865	31.68	14.40	1.51	2.77	2.58	19.62
M744	30.08	14.77	2.03	2.98	3.10	20.09
M454	34.34	18.97	2.36	4.43	3.16	29.58

**All data is presented as % cell wall dry weight*

Conclusions

Using FTIR spectra generated from a HTP screening framework, a hierarchical model capable of robustly, and accurately clustering cell wall variants across a variety of cell wall types including both type I and type II cell walls was generated. *Bmr6* and *Bmr12* were identified in a 'lignin deficient' clade and both mutants were characterised in terms of their FTIR spectra in relation to reported cell wall modifications of these mutants. The approach was then applied to 100 lines from a segregating M3 EMS sorghum population and 6 lines were identified that clustered distinctly from the parent line. These lines were shown to have altered lignin and cell wall glucan content, using wet chemistry, as predicted by FTIR spectral analysis. This therefore, provides a blueprint for characterising cell wall mutant populations of any species without the need for laborious 'wet' chemical analysis or calibration of quantitative predictive models which may be species specific.

Materials and Methods

Plant growth conditions

One hundred sweet *Sorghum bicolor* ecotypes were field grown at the 'Pacific Seeds Gatton Research Centre' Queensland, in September 2009 and were harvested in February 2010 approximately 10 days post anthesis. Subsequently, seed from 15 selected genotypes, along with the sweet sorghum variety 'Rio', were sown in a glasshouse in Newcastle, Australia in May 2010 in 25cm diameter pots. Plants in Newcastle were supplemented with 300 $\mu mol.m^{-2}.s^{-1}$ PAR (400-700 nm) from 4-6:30pm each day and natural sunlight throughout the winter growth period ranged from 200-1100 $\mu mol.m^{-2}.s^{-1}$ PAR with 28/15°C day/night temperatures. Plants were harvested in August 2010.

Additionally, ~3300 sweet sorghum seed of the 'Rio' variety were soaked for 16 hrs with agitation in 200 mL of 0.2 % (v/v) ethyl methanesulfonate (EMS) to induce ~300 point mutations of G:C to A:T base pairs per seed. Seed were then washed in 400 mL of water for 5 hrs with changes every 30 min. Seed were planted at the 'Pacific Seeds Research Centre' Queensland, in September 2009. Plants were self fertilised and propagated to the M3 generation and planted at the same site in September 2011. 100 individual plants were sampled as per the high-throughput screening protocol outlined in [13].

FTIR spectroscopy of isolated cell walls

Once all spectra were collected, as outlined in [25], data was imported into Unscrambler X 10.0.1 (Camo Inc.) for pre-treatment. Spectra were truncated to the cell wall fingerprint region (1800-850 cm^{-1} wavenumbers), an extended multiplicative scatter correction (EMSC) was applied to correct for light scattering by cell wall particles and second derivative spectra were calculated by the Savitsky-Golay method with 9 smoothing points.

Statistical analysis

Principal component analysis (PCA) was performed with full cross-validation and a default of 15 principal components were computed using the 'Singular Value Decomposition' algorithm (The Unscrambler X 10.0.1, CAMO) with mean centred data. The hierarchical

cluster models were generated using unsupervised hierarchical classification model based on the squared Euclidean distance between samples where the number of cluster was set manually.

Cell wall chemistry

Klason lignin and monosaccharide content of cell walls was determined by the NREL method for determination of structural carbohydrates and lignin in biomass [26].

List of abbreviations used

FTIR, fourier transform infrared; HTP, high-throughput; *Bmr*, brown midrib; PCA, principal component analysis; WT, wild type; GAX, glucuronoarabinoxylan; EMSC, extended multiplicative scatter correction; PC, principal component; CAD, cinnamyl alcohol dehydrogenase; COMT, caffeic acid/5-hydroxyferulic acid O-methyl transferase; Col-0, Columbia-0; ANOVA, analysis of variance; PAR, photosynthetically active radiation; ATR, attenuated total reflection; HPLC, high performance liquid chromatography.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

APM theorised the cell wall cluster model, designed all experiments, collected all data, analysed all data, produced all figures and tables (excluding figure 4), and drafted the manuscript. WMP also collected sorghum samples and FTIR spectra and produced figure 4. CB assisted in wet chemistry analysis of cell walls. CSB and CPLG designed and performed the generation of the EMS mutant population. CPLG conceived of, and established the Sorghum biofuel research program. CPLG, CSB and RTF reviewed the manuscript. All authors read and approved the final manuscript.

Acknowledgments

Antony Martin is the recipient of an Office of the Chief Executive (OCE) PhD Scholarship from CSIRO. We thank the Australian Research Council (ARC) and Pacific Seeds for financial support under ARC's Linkage Projects funding scheme (Project Number LP0883808). The

authors would like to thank Joseph Enright for maintenance of plants in the University of Newcastle Plant Growth Facility and Geoff Doherty and Tony Banks from Ethanol Technologies Ltd for operation and access to the HPLC instrument.

References

1. Sticklen MB: **Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol.** *Nature Reviews Genetics* 2008, **9:**433-443.
2. Boudet A-M: **Towards an understanding of the supramolecular organization of the lignified wall.** In *The Plant Cell Wall. Volume* 8. Edited by Rose JKC. Oxford: Blackwell; 2003
3. Carpita NC, Gibeaut DM: **Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth.** *The Plant Journal* 1993, **3:**1-30.
4. Carpita NC, McCann MC: **Maize and sorghum: genetic resources for bioenergy grasses.** *Trends in Plant Science* 2008, **13:**415-420.
5. Vogel J: **Unique aspects of the grass cell wall.** *Current Opinion in Plant Biology* 2008, **11:**301-307.
6. Carpita NC: **Structure and Biogenesis of the Cell Walls of Grasses.** *Annual Review of Plant Physiology and Molecular Biology* 1996, **47:**445-476.
7. Hunt JW, Dean AP, Webster RE, Johnson GN, Ennos AR: **A Novel Mechanism by which Silica Defends Grasses Against Herbivory.** *Annals of Botany* 2008, **102:**653-656.
8. Maris AJAv, Abbott DA, Bellissimi E, Brink Jvd, Kuyper M, Luttik MAH, Wisselink HW, Scheffers WA, Dijken JPv, Pronk JT: **Alcoholic fermentation of carbon sources in biomass hydrolysates by Saccharomyces cerevisiae: current status.** *Antonie van Leeuwenhoek* 2006, **90:**391-418.
9. McCann MC, Carpita NC: **Designing the deconstruction of plant cell walls.** *Current Opinion in Plant Biology* 2008, **11:**314-320.
10. Pauly M, Keegstra K: **Cell-wall carbohydrates and their modification as a resource for biofuels.** *The Plant Journal* 2008, **54:**559-568.
11. Mansfield SD: **Solutions for dissolution—engineering cell walls for deconstruction.** *Current Opinion in Plant Biology* 2009, **20:**286-294.
12. Sattler SE, Funnell-Harris DL, Pedersen JF: **Brown midrib mutations and their importance to the utilization of maize, sorghum, and pearl millet lignocellulosic tissues.** *Plant Science* 2010, **178:**229-238.
13. Martin AP, Palmer WM, Byrt CS, Furbank RT, Grof CP: **A holistic high-throughput screening framework for biofuel feedstock assessment that characterises variations in soluble sugars and cell wall composition in Sorghum bicolor.** *Biotechnology for Biofuels* 2013, **6:**1-13.
14. Mouille G, Robin S, Lecomte M, Pagant S, Hofte H: **Classification and Identification of *Arabidopsis* cell wall mutations using Fourier-Transform InfraRed (FT-IR) microscopy.** *The Plant Journal* 2003, **35:**393-404.
15. McCann MC, Defernez M, Urbanowicz BR, Tewari JC, Langewisch T, Olek A, Wells B, Wilson RH, Carpita NC: **Neural Network Analyses of Infrared Spectra for Classifying Cell Wall Architectures1.** *Plant Physiology* 2007, **143:**1314-1326.
16. Kim HJ, Triplett BA: **Cotton Fiber Growth in Planta and in Vitro. Models for Plant Cell Elongation and Cell Wall Biogenesis.** *Plant Physiology* 2001, **127:**1361-1366.
17. Brutnell TP, Wang L, Swartwood K, Goldschmidt A, Jackson D, Zhu X-G, Kellogg E, Van Eck J: **Setaria viridis: A Model for C4 Photosynthesis.** *The Plant Cell Online* 2010, **22:**2537-2544.

18. Franke R, Hemm MR, Denault JW, Ruegger MO, Humphreys JM, Chapple C: **Changes in secondary metabolism and deposition of an unusual lignin in the ref8 mutant of Arabidopsis.** *The Plant Journal* 2002, **30:**47-59.
19. Barrière Y, Ralph J, Méchin V, Guillaumie S, Grabber JH, Argillier O, Chabbert B, Lapierre C: **Genetic and molecular basis of grass cell wall biosynthesis and degradability. II. Lessons from brown-midrib mutants.** *Plant Biology and Pathology* 2004, **327:**847-860.
20. Palmer NA, Sattler SE, Saathoff AJ, Funnell D, Pedersen JF, Sarath G: **Genetic background impacts soluble and cell wall-bound aromatics in brown midrib mutants of sorghum.** *Planta* 2008, **229:**115-127.
21. Vogler RK, Tesso TT, Johnson KD, Ejeta G: **The effect of allelic variation on forage quality of brown midrib sorghum mutants with reduced caffeic acid O-methyl transferase activity.** *African Journal of Biochemistry Research* 2009, **3:**70-76.
22. Pandey KK: **A study of chemical structure of soft and hardwood and wood polymers by FTIR spectroscopy.** *Journal of Applied Polymer Science* 1999, **71:**1969-1975.
23. Corredor DY, Salazar JM, Hohn KL, Bean S, Bean B, Wang D: **Evaluation and Characterization of Forage Sorghum as Feedstock for Fermentable Sugar Production.** *Applied Biochemical Technologies* 2009, **158:**164-179.
24. Theerarattananoon K, Wu X, Staggenborg S, Propheter R, Madl R, Wang D: **Evaluation and characterization of sorghum biomass as feedstock for sugar production.** *Transactions of the ASABE* 2010, **53:**509-525.
25. Martin A, Palmer W, Byrt C, Furbank R, Grof C: **A holistic high-throughput screening framework for biofuel feedstock assessment that characterises variations in soluble sugars and cell wall composition in Sorghum bicolor.** *Biotechnology for Biofuels* 2013, **6:**1-13.
26. Sluiter A, Hames B, Ruiz R, Scarlata C, Sluiter J, Templeton D, Crocker D: **Determination of Structural Carbohydrates and Lignin in Biomass.** *NREL technical report* 2011.

CHAPTER FOUR

Gene discovery in a developing *Setaria viridis* internode

4.1 Summary

This chapter consists of one publication that outlines the use of a developing *Setaria viridis* internode as an ideal experimental system for gene discovery in relation to the biogenesis of stem tissue, sugar accumulation in mature stems and, importantly for this thesis, cell wall biosynthesis. It provides a framework for identification of cell wall genes that are important in cell wall biosynthesis during stem development. The publication has been published in *Biotechnology for Biofuels*. Furthermore, specific genes for future investigation and genetic manipulation identified during the production of this publication have been outlined in the final discussion chapter of this dissertation.

4.2 Publication 4: A developing *Setaria viridis* internode: an experimental system for the study of biomass generation in a C_4 model species

Martin AP, Palmer WM, Brown C, Abel C, Lunn JE, Furbank RT, Grof CPL **A developing *Setaria viridis* internode: an experimental system for the study of biomass generation in a C_4 model species** *Biotechnology for Biofuels* **9**:45 doi:10.1186/s13068-016-0457-6 (2016).

Statements of contribution of others: Publication 4

*I attest that Research higher degree candidate **Antony Martin** was the primary contributor to the development of this publication. The author contributions were as follows:*

APM, WMP, CPLG and RTF conceived of the study. APM designed and conducted all experiments, analysed all data, produced all figures and drafted the manuscript. WMP contributed to data analysis, drafting of the manuscript, and production of figures. CB performed some preliminary RNAseq data analysis. CA optimised Setaria growth conditions. JEL provided intellectual input and resources for Setaria growth, RNA sequencing and metabolite measurements. RTF, JEL and CPLG reviewed the manuscript. All authors read and approved the final manuscript.

_____

Antony Martin

_____

William Palmer

_____

Christopher Brown

_____

Christin Abel

_____

John Lunn

_____

Robert Furbank

_____

Christopher Grof

Martin *et al. Biotechnol Biofuels (2016) 9:45*
DOI 10.1186/s13068-016-0457-6

Biotechnology for Biofuels

RESEARCH Open Access

A developing *Setaria viridis* internode: an experimental system for the study of biomass generation in a C_4 model species

Antony P. Martin[1], William M. Palmer[1], Christopher Brown[1], Christin Abel[2], John E. Lunn[2], Robert T. Furbank[3,4] and Christopher P. L. Grof[1*]

Abstract

Background: Recently, there has been interest in establishing a monocot C_4 model species with a small genome, short lifecycle, and capacity for genetic transformation. *Setaria viridis* has been adopted to fill this role, since reports of Agrobacterium-mediated transformation in 2010, and sequencing of its genome in 2012. To date, *S. viridis* has primarily been used to further our understanding of C_4 photosynthesis, but is also an ideal system for the study of biomass crops, which are almost exclusively C_4 panicoid grasses. Biogenesis of stem tissue, its cell wall composition, and soluble sugar content are important determinants of bioenergy crop yields. Here we show that a developing *S. viridis* internode is a valuable experimental system for gene discovery in relation to these important bioenergy feedstock traits.

Results: The rate of maximal stem biomass accumulation in *S. viridis* A10 under long day growth was at the half-head emergence developmental stage. At this stage of development, internode 5 (of 7) was found to be rapidly expanding with an active meristem, a zone of cell expansion (primary cell walls), a transitional zone where cell expansion ceased and secondary cell wall deposition was initiated, and a mature zone that was actively accumulating soluble sugars. A simple method for identifying these zones was established allowing rapid dissection and snap-freezing for RNAseq analysis. A transcriptome profile was generated for each zone showing a transition from cell division and nucleic acid synthesis/processing in the meristem, to metabolism, energy synthesis, and primary cell wall synthesis in the cell expansion zone, to secondary cell wall synthesis in the transitional zone, to sugar transport, and photosynthesis in the mature zone.

Conclusion: The identification of these zones has provided a valuable experimental system for investigating key bioenergy traits, including meristematic activity, cell wall biosynthesis, and soluble sugar accumulation, in a C4 panicoid grass that has genetic resources, a short life cycle, and small stature allowing controlled experimental conditions in growth cabinets. Here we have presented a comprehensive map of gene expression and metabolites in this experimental system to facilitate gene discovery and controlled hypothesis testing for bioenergy research in *S. viridis*.

Keywords: *Setaria viridis*, RNAseq, Stem, Bioenergy, Cell wall, Sugar, C_4 grass

Background

Setaria viridis, the wild relative of foxtail millet (*Setaria italica*), has recently been adopted as a model species for the study of C_4 panicoid grasses by the photosynthesis research community [1–6]. The support of the research community for this model is evidenced by the first *Setaria* conference held in Beijing in 2014 [7]. Reported uses of *Setaria viridis* to date have focused on the collection of natural diversity panels [8, 9], the development of mapping populations [10], transformation [11–14] and crossing [15] techniques, and generation of molecular resources such as transformation vectors [12] and

*Correspondence: chris.grof@newcastle.edu.au
[1] School of Environmental and Life Sciences, University of Newcastle, University Drive, Callaghan, NSW 2308, Australia
Full list of author information is available at the end of the article

notably a genome sequence [16]. Most research using these resources has been directed towards the study of C_4 photosynthesis and associated leaf development and cellular differentiation [17]; however, *S. viridis* is also a valuable model species for the study of biomass crops, which are almost exclusively C_4 panicoid grasses (e.g. giant miscanthus (*Miscanthus x giganteus*), switchgrass (*Panicum virgatum*), sugarcane (*Saccharum officinarum*), maize (*Zea mays*), and sorghum (*Sorghum bicolor*)) sharing a highly similar lineage to *Setaria*. Whilst photosynthesis is an important determinant of crop biomass yield, knowledge of the factors controlling biogenesis of stem sink tissue, its cell wall composition, and soluble sugar content are crucial for the improvement of forage and bioenergy crop quality and yields.

The main agricultural products of forage and bioenergy crops are soluble sugars and the cell wall fractions of culms, which are used as a source of feed or fuel. To maximise yields, larger culm volumes with high sugar content and easily hydrolysed cell wall polymers are required. The cell wall composition of *S. viridis* has been shown to be consistent with that of major bioenergy crops with equivalent conversion rates of cellulose to fermentable sugars [18]. No study of soluble sugar levels in *S. viridis* culms has been published; however, Brix concentrations of up to ten can be achieved (Martin et al. unpublished observations), which is equivalent to that of mid-range sweet sorghum [19] and sugarcane [20] crops. *Setaria viridis* culms therefore possess attributes that suggest they would be a valuable C_4 model system for the study of not only cell wall synthesis and stem development but also sugar accumulation in C_4 panicoid grass culms.

A grass culm consists of multiple repeating phytomeric units, comprised of a node and an internode. Stalks develop and elongate acropetally, whilst individual internodes develop and elongate basipetally with a sigmoidal elongation pattern [21]. At the midway stage in its development, an internode will contain an active intercalary meristem at its base, acropetally followed by a cell expansion zone, before transitioning into mature cells towards the top. Based on early physiological studies of oat culm development [22] and extensive literature in sugarcane and sweet sorghums, which show the accumulation of sugar in mature internodes [23–25], the following picture emerges for a developing internode in panicoid grass species. The meristem contains actively dividing cells undergoing mitosis and cell differentiation processes, and the expansion zone contains cells that are actively expanding under turgor whilst synthesising primary cell walls and cell membranes. As the cells mature, secondary cell wall deposition terminates cell expansion and the cells realise their capacity to import and store sugar.

Here we characterise a developing *S. viridis* internode that contains an active intercalary meristem, rapidly expanding cells with primary cell walls, a transitional zone where secondary cell wall synthesis is initiated, followed by a mature zone where sugar import and storage are occurring. RNA sequencing of tissues from each of these developmental zones provides a framework for gene discovery. Ultimately, genetic manipulation of genes discovered here has the potential to alter sink size by affecting cell division and expansion in the meristematic and cell expansion zones, and to influence the molecular composition of cell wall polymers and soluble sugars by affecting metabolic pathways in the transitional and maturation zones. Whilst there have been other transcript studies conducted in grass stems at various stages of development, including young and mature maize internodes [26], and in bamboo [27], this is the first study to characterise and utilise the developmental zones within an elongating internode, and the first whole RNAseq transcriptome dataset for *S. viridis* stem tissue. This experimental system and transcriptomic resource in the C_4 model species, *S. viridis*, will facilitate gene discovery with respect to grass culm development and allow controlled experimental exploration of traits relevant to bioenergy crops such as culm sink size, cell wall composition, and soluble sugar content. The genetic and physiological similarities between *Setaria viridis* and key bioenergy crops will increase the translatability of results into forage and bioenergy crop improvement programmes.

Results

To identify a rapidly expanding internode, *S. viridis* plants were grown under highly controlled conditions and seven developmental stages were examined. The fourth developmental stage, half-head emergence, was defined as the point when half of the seed head had emerged from the leaf sheath (Fig. 1a). This was chosen as the developmental stage with the most rapid rate of stem growth, which also corresponded with the most rapid rate of sucrose accumulation (Fig. 1b). At this developmental stage, the leaf sheath was carefully stripped to expose the stem and its constituent internodes. Internode five, counted from the bottom (considering only internodes $\geq$5 mm in length), or, interchangeably, the second internode from the top (not including the flag internode), was selected as a rapidly elongating internode containing well-defined elongation and mature zones (Fig. 1c). To ensure this internode was rapidly elongating, the same internode was also measured in mature plants and was shown to increase from an average length of 3.8 cm at harvest (half-head emergence) to 10.2 cm at maturity (Fig. 1d).

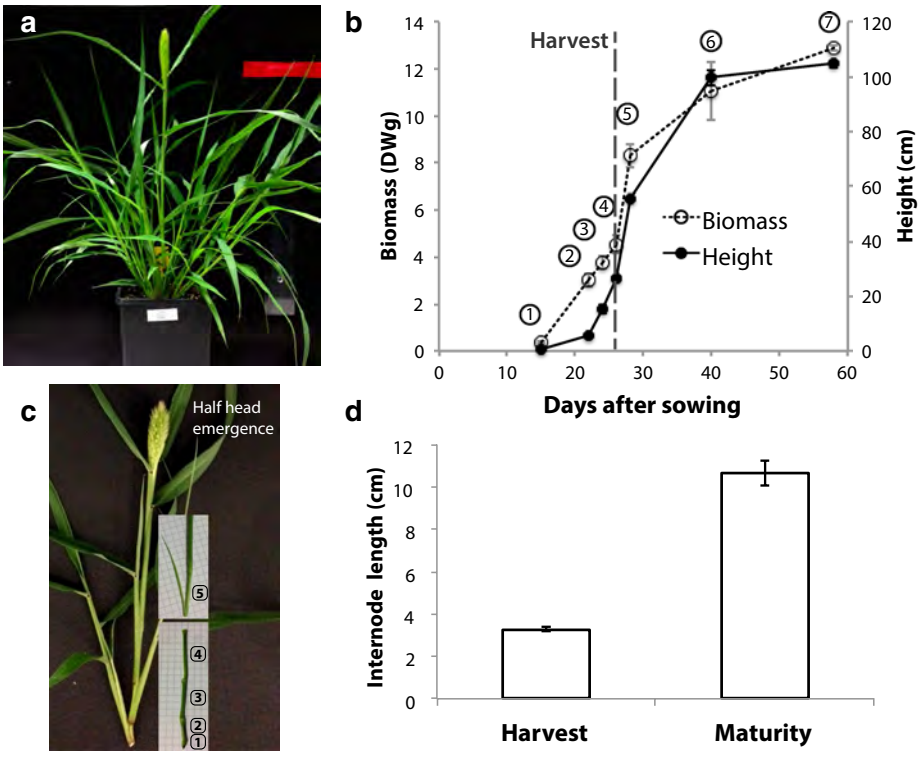


Fig. 1 Selection of a rapidly elongating internode. **a** *S. viridis* A10 grown under controlled conditions at half-head emergence developmental stage. **b** Main stalk height and sucrose concentration measured at seven developmental stages: 6-leaf (*1*), 9-leaf (*2*), booting (*3*), half-head emergence (*4*), anthesis (*5*), milky dough (*6*), and maturity/hard dough (*7*). The harvested stage (half-head emergence) is indicated with a *dashed line*. **c** The main stalk was stripped of its leaf sheath with internodes labelled 1–5 (counting only internodes ≥5 mm in length). **d** Internode 5 length measured at harvest and at maturity. All values are mean ± SE of five biological replicates

To define the meristematic, cell expansion, transitional, and mature zones of internode five, nuclei density and cell size were examined using fluorescence microscopy. It was also essential to develop a method for visually identifying these zones of internode five so that harvest for RNAseq analysis could be achieved rapidly, without any need for microscopy, and therefore without excessive tissue damage.

Internodes were photographed (Fig. 2a), whilst intact and then embedded in agarose and sectioned longitudinally. Nuclei were stained with DAPI, illuminated with UV light and viewed using a long-pass DAPI filter that allowed red chlorophyll fluorescence (Fig. 2b) and blue DAPI fluorescence of nuclei (Fig. 2c) to be visualised simultaneously. Processing DAPI images allowed nuclei density (a proxy for the number of cells) and also cell size to be plotted along the length of the internode (Fig. 2d). A spike in nuclei density mapped to a well-defined white band at the base of the internode, and this was defined as the meristematic zone (MsZ) (Fig. 2a) allowing for rapid identification of this zone. Directly above the meristematic zone, cell size increased and nuclei density decreased, both asymptotically. It was observed that this cell expansion zone was dark green in colour, which was also indicated by the increased red chlorophyll fluorescence observed in longitudinal sections of this region (Fig. 2b). The base of the internode was flexible (Fig. 2a), likely due to the lack of lignified secondary cell walls. The point where the internode became rigid corresponded with the end of the chlorophyll-enriched zone and the approach of nuclei density and cell size towards their asymptote. Both the end of the chlorophyll-enriched zone and the point where the internode became rigid were therefore used as markers to dissect the cell expansion zone (CEZ) (Fig. 2a) of the internode without the need to stain or section the tissue. The region of the internode directly above this point, cut to ~85 % the length of the CEZ, was dissected and defined as the transition zone (TZ) (Fig. 2a). A zone immediately below the node above internode five, cut at the same length as the CEZ, was dissected as the mature zone (MZ) (Fig. 2a).

Using brightfield microscopy, it was evident that there were many different cell types within these zones. The cell expansion zone contained immature vascular bundles with protoxylem and very little cell wall thickenings in any cell types (Fig. 2e). The transition zone contained

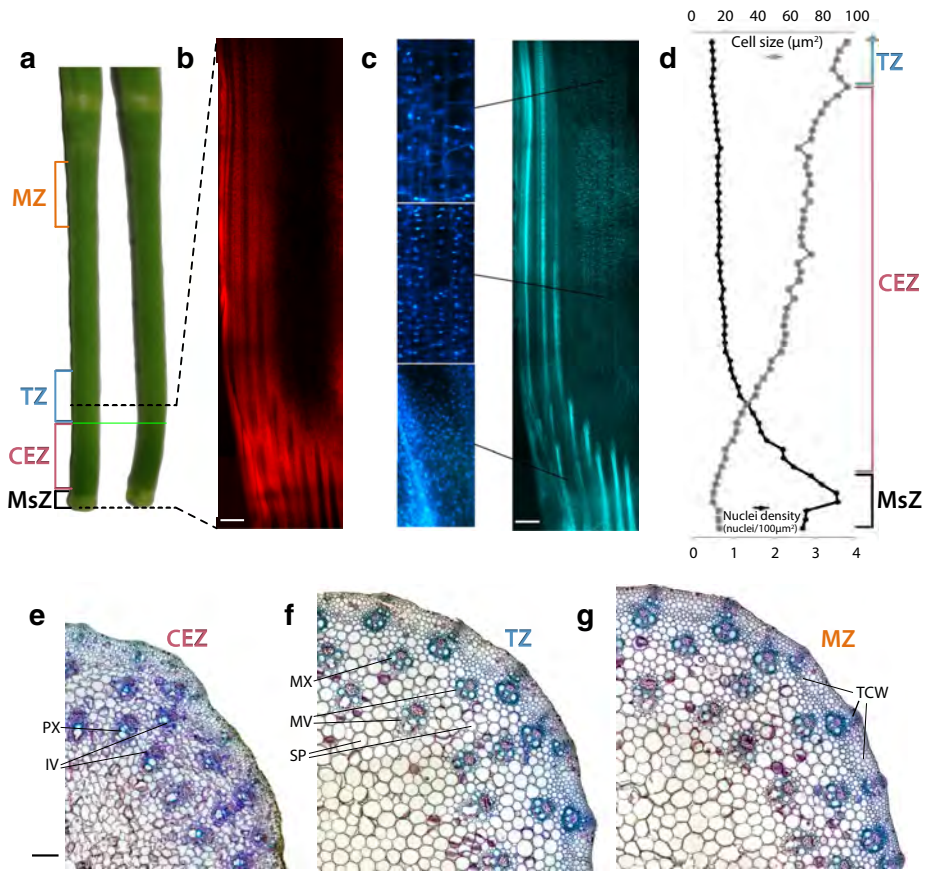


Fig. 2 Defining meristematic (MsZ), cell expansion (CEZ), transition (TZ), and cell maturation (MZ) zones of internode 5. **a** Internode 5 harvested at the half-head emergence developmental stage with its leaf sheath stripped, alongside an equivalent internode 5 where the lower, flexible zone has been bent to display the differentiation between the upper rigid zone and the lower flexible zone. The *green line* indicates the interface between the flexible and rigid zones of the internode. The harvested meristematic (MsZ—*black*), cell expansion (CEZ—*pink*), transitional (TZ—*blue*), and mature (MZ—*orange*) zones are indicated. **b** and **c** Longitudinal, vibratome cut, 50-µm thick section of the lower region of the internode (indicated by *dashed black lines*), stained with DAPI and viewed under UV illumination with **b**, the *red* chlorophyll emissions isolated, and **c**, the *blue* DAPI emissions isolated. Enlarged regions are offset in **c**, and *white scale bars* are 50 µm. **d** Nuclei density and cell size measured using image J were plotted at intervals along the lower region of the internode. **e**, **f** and **g**, Microtome cut 2-µm-thick cross-sections from e, the cell expansion, **f**, transitional, and **g**, mature zones, stained with Toluidine blue, counter stained with Lugol's iodine and imaged using brightfield illumination. Protoxylem (PX), immature vascular bundles (IV), metaxylem (MX), mature vascular bundles (MV), storage parenchyma (SP), and thickened cell walls (TCW) are labelled. *Black scale bars* in **e**, **f**, and **g** are 100 µm

fully developed vascular bundles with fully developed xylem, but only modest cell wall thickenings (Fig. 2f). The mature zone contained fully developed vascular bundles with fully developed xylem and more cell wall thickening than the transitional zone (Fig. 2g). Based on evidence from these images, the dissected zones contained a variety of cell types, but were sufficiently enriched in expanding, transitioning, and mature tissues to reveal differences in metabolite levels and gene expression between each developmental stage.

Primary metabolites were measured in the four zones of the developing internode to establish carbon pools (Fig. 3). Starch was found only in the developmentally young meristematic and cell expansion zones. The ratio of sucrose to hexoses was high in the meristematic zone, reduced in the growing cell expansion and transitional zones, and increased in the mature zone, which was likely at the beginning of a sugar storage phase as sugar levels rise towards the maximal concentration observed in stage 6 (Fig. 1b). High levels of the soluble C_4 acid malate were also measured and correlated strongly with sucrose levels. The role of malate in this developing C_4 internode is unclear; however, it may be involved in a functional C_4 pathway, used as a pH regulator, a storage molecule in vacuoles, or an osmolyte for turgor regulation [28].

The lignin content of cell walls was also measured in the cell expansion, transitional, and mature internode zones as an indicator of the presence of primary or secondary cell walls (Fig. 3). Very low levels of lignin were observed in the cell expansion zone; however, the cell walls rapidly became lignified in the transitional zone, and continued to accumulate lignin up to the mature zone (Fig. 3). This suggested that the cell expansion zone contained primary cell walls, whereas the transitional zone was rapidly synthesising secondary cell walls, and that this synthesis was maintained, up to higher levels in the mature zone where it has, likely, completed lignification.

Once the developmental status of the four internode zones was established, RNA was extracted, sequenced, mapped (see Additional file 1 for mapping details), and expression values calculated. Principal component analysis (PCA) of global gene expression confirmed that each internodal zone had a unique gene expression profile with the meristematic and cell expansion zones clustering separately from the transitional and mature zones (Fig. 4a). Very little variation between biological replicates was also observed confirming the reproducibility of the harvesting technique.

Comparing $\log_2$ fold-change expression of each gene to that obtained from a 'whole plant' background transcriptome [3], lists of genes that were up-regulated (>1 $\log_2$ fold) in each zone were generated and displayed as a Venn diagram (Fig. 4b). This also revealed a dichotomy between the two younger meristematic and cell expansion zones, and the two older transitional and mature zones. Thousand eight hundred and sixty genes were up-regulated in common in MsZ/CEZ and 983 in TZ/MZ. In comparison, only 52 MsZ/TZ, 157 MsZ/MZ, 213 CEZ/TZ, and 12 CEZ/MZ genes were up-regulated in common (Fig. 4b). The number of up-regulated genes and the number expressed above a threshold of one fragments per kilobase per million reads (FPKM) also showed dichotomy with more being found in younger than older tissues (Fig. 4c). This suggests that greater biological complexity is required for meristematic activity and cell expansion than for the established biological processes of mature tissues including secondary cell wall synthesis and carbon storage.

Using criteria of gene expression above a level of 80 FPKM (chosen by visually examining read coverage of genes with ~80 FPKM), and a $\log_2$ fold change >2.5 against the whole plant (WP) background transcriptome, 418 'stem-enriched', 13 'MsZ-enriched', 8 'CEZ-enriched', 84 'TZ-enriched', and 45 'MZ-enriched' genes were identified which can be used to identify promoter sequences for the construction of tissue-specific, and likely cell-specific, expression vectors ('enriched' genes are marked in Additional file 2).

To further characterise the segregation of biological processes within each zone of the developing internode, genes were categorised based on the most recent Mapman ontology, which was updated for this analysis (Additional file 3). The meristematic zone was enriched in processes related to cell division and cell cycle including DNA, RNA, and protein synthesis/regulation, confirming that this zone is actively undergoing cell division. The cell expansion zone was enriched in metabolism, energy

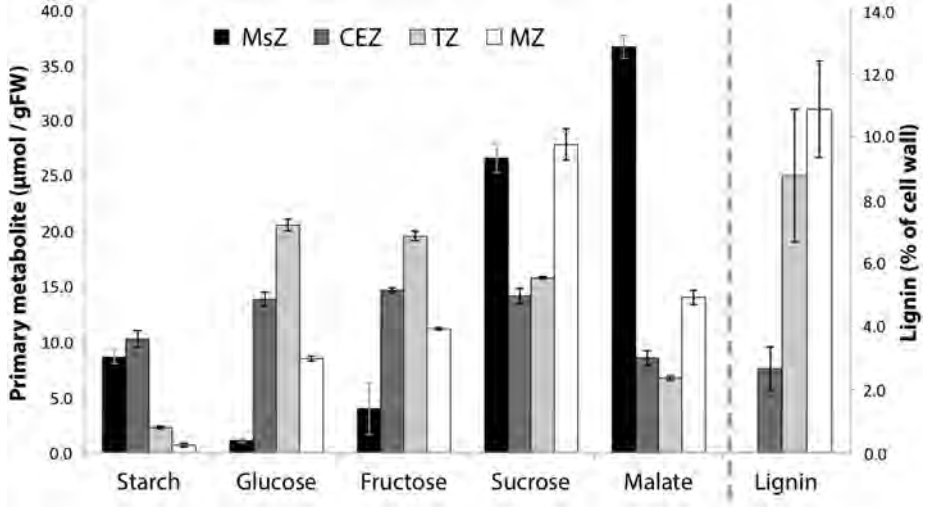
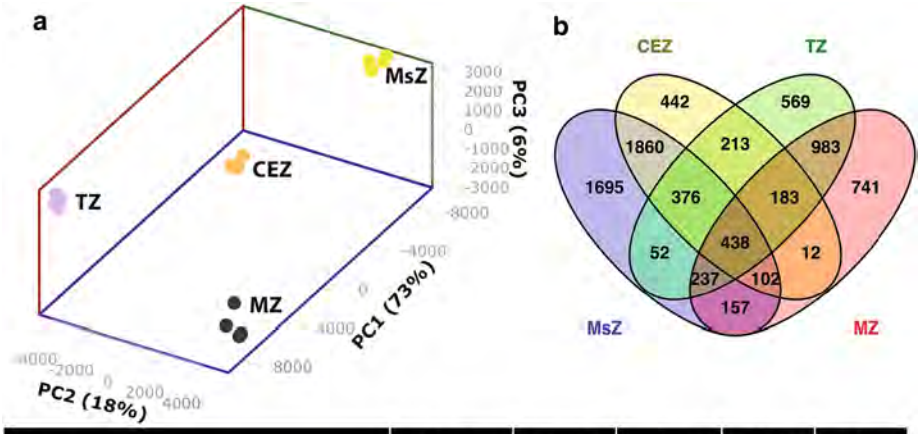


Fig. 3 Primary metabolites and cell wall lignin in the developing internode. Starch (measured as molecules of released glucose), glucose, fructose, sucrose, and malate were assayed enzymatically in extracts from the meristem, cell expansion, transition, and mature zones of the fifth internode. Lignin was measured using the acetyl bromide method. All values are mean ± SE of four biological replicates of pooled material from ten plants



c	MsZ	CEZ	TZ	MZ	WP
No. genes (>1FPKM)	18333	17329	16869	17206	21773
Up-regulated genes (from WP)	4917	3626	3051	2853	
'Stem specific'	418				
'Zone specific'	13	8	84	45	

Fig. 4 Internode zones have unique RNAseq profiles and contain stem- and zone-specific genes. **a** Principal component analysis of FPKM values obtained from RNA sequencing of the MsZ, CEZ, TZ, and MZ of internode 5. **b** *Venn diagram* of genes up-regulated greater than a $\log_2$ fold change of one against the background levels in the whole plant (WP) transcriptome. **c** *Table* showing the number of genes expressed >1 FPKM, number of genes up-regulated against the WP transcriptome with >1 $\log_2$ fold, number of identified 'stem-specific' genes with >2.5 $\log_2$ fold change against the WP background and >80 FPKM in any internode zone, and the number of 'zone-specific' genes with, in addition to being 'stem-specific', >2 $\log_2$ fold change against all other internodal zones (with some exceptions, see Additional file 2)

production, and lipid biosynthesis, which are required to drive cell expansion. The transitional zone was enriched in secondary cell wall biosynthesis-related gene categories including lignin biosynthesis, laccases, cell vesicle transport, MYB, and NAC transcription factors (which are maintained in the mature zone), and amino acid metabolism. The mature zone was enriched in sugar transporters, photosynthesis (which begins to operate, however, not to the levels observed in the whole plant), and flavonoid biosynthesis (Fig. 5). These transcript profiles confirm that the transcriptional machinery required for stem-specific intercalary meristem cell division and differentiation is active in the meristematic zone; energy-driven cell expansion is active in the cell expansion zone; secondary cell wall synthesis, including transcriptional switches, is active in the transitional zone; and sugar unloading, storage, and possibly synthesis (photosynthesis) is active in the mature zone.

To encapsulate the effectiveness of this experimental system for gene discovery and manipulation in a variety of fundamental biological processes, Mapman was used to visualise the expression of gene families relating to three major carbon demands of a developing grass culm (Fig. 6). In the meristematic zone, gene expression suggested predominantly symplasmic import of photoassimilate and utilisation of starch for cell metabolism and energy production (demand 1), since expression of sugar transporters, including *SWEET*s (sugar effluxers [29], *SUT*s (sucrose-H^+ symporters [30]), *HT*s (hexose transporters [31]), *PLT*s (polyol transporters [32]), and *CWI* (cell wall invertase [33]) was low (Fig. 6). Gene expression in the cell expansion zone also suggested that photoassimilate is imported predominantly symplasmically [i.e. low expression of *SWEET*s, *SUT*s, *HT*s, *PLT*s, and *CWI*s and higher expression of plasmodesmatal located proteins (*PDLP*s)] not only in this zone and is directed towards demand 1 (energy/metabolism), but also primary cell wall synthesis, including hemicellulose, pectin, and cell wall protein synthesis (demand 2) (Fig. 6). The transitional zone is a pivotal developmental stage switching from predominantly metabolism (demand 1) and symplasmic photoassimilate import to a highly significant up-regulation of genes associated with apoplasmic sugar transport, coupled with a major switch and a re-direction of carbon flow to secondary cell wall synthesis (demand 2) (Fig. 6). The mature zone showed down-regulation of genes linked to metabolism (demand 1) and cell wall synthesis (demand 2), whilst apoplasmic transport genes (*SWEET*s, *SUT*s, *HT*s, *PLT*s, and *CWI*s) and vacuolar sugar storage (demand 3) genes (tonoplast

		MsZ	CEZ	TZ	MZ	WP
MZ	Transport - sugars	14.2	16.0	19.3	23.6	15.5
	Hormone metabolism	13.4	15.9	26.1	32.0	16.6
	PS	67.7	114.0	165.5	222.9	451.3
	Development	13.3	10.4	12.9	17.0	11.7
	Redox	61.3	57.3	49.6	67.9	63.0
	DNA repair	13.6	9.9	14.1	18.4	10.0
	Flavanoids	12.5	14.2	43.3	45.5	8.4
	MYB & NAC TFs	5.2	3.7	9.7	10.8	9.6
TZ	Lignin	18.7	19.4	180.1	135.6	20.1
	Laccase	6.9	3.8	27.5	12.0	4.7
	Cell wall	19.3	30.5	54.2	30.7	14.4
	Transport - other	12.4	13.1	16.4	15.8	17.8
	Signalling-MAP kinases	19.7	19.5	37.0	31.9	17.5
	Cell vesicle transport	29.3	34.8	36.2	26.9	18.9
	Cell organisation	43.0	49.6	68.3	51.3	24.4
	Amino acid metabolism	44.4	47.5	81.8	72.1	37.3
CEZ	Sucrose metabolism	31.5	51.9	51.6	25.0	17.0
	Starch metabolism	59.1	68.3	30.8	22.9	47.3
	Lipid	21.5	25.9	21.0	22.9	29.5
	Transport - expansion	61.6	99.7	70.0	55.1	62.7
	TCA	78.5	105.2	88.4	60.5	168.0
	Mitochondrial energy	80.6	103.2	61.6	40.0	47.0
	Glycolysis	110.8	158.0	91.4	84.1	73.9
MsZ	Metal handling	160.3	33.6	69.9	127.9	86.9
	Stress	21.2	18.3	11.3	10.8	16.1
	Signalling-Gproteins	42.8	42.3	32.9	27.2	22.6
	TF Cell proliferation	14.8	10.8	9.5	12.0	11.0
	Protein	52.6	50.6	22.6	24.8	30.5
	RNA binding	38.5	33.1	19.5	24.1	24.6
	RNA processing	39.9	32.5	14.3	18.8	27.6
	RNA transcription	18.7	14.0	10.3	13.9	14.0
	DNA synthesis	37.0	27.0	12.5	15.5	24.8
	Cell division & cell cycle	28.4	25.0	15.5	17.9	15.8

Fig. 5 Meristematic, cell expansion, transition, and mature zones are enriched in unique gene families. Average expression of selected gene families, based on the Mapman annotations, arranged into categories that are enriched in the meristematic (MsZ), cell expansion (CEZ), transition (TZ), and mature (MZ) zones of the internode. Mean expression is tabulated for MsZ, CEZ, TZ, MZ, and whole plant (WP) and the heat map is generated from the $\log_2$ fold change across the four internode zones. *Blue* indicates higher expression, whilst *red* indicates lower expression, and values are given as mean FPKM for each gene category

monosaccharide transporters—*TMT*s [34]) showed further up-regulation. This suggested that the mature zone had ceased most of its transcriptional activity related to secondary cell wall synthesis and is actively transcribing genes that are associated with apoplasmic unloading of photoassimilate for storage.

Discussion

Setaria viridis is becoming more widely accepted as a C_4 grass model species, especially within the photosynthesis research community. Here we have presented the first transcriptomic analysis of this model system in tissue relevant to bioenergy research, at a gene discovery level. We have identified and shown that a developing *S. viridis* internode contains (i) an active meristematic zone that utilises starch and likely symplasmically imported sucrose and malate as a carbon source to drive metabolism, energy production, and synthesis of DNA, RNA, and protein, (ii) a zone of cell expansion with predominantly non-lignified primary cell walls which also utilises starch and symplasmically imported sugars, in the form of hexoses to produce energy, driving cell expansion, (iii) a transitional zone where cell expansion ceases, secondary cell wall synthesis is initiated and sugar import

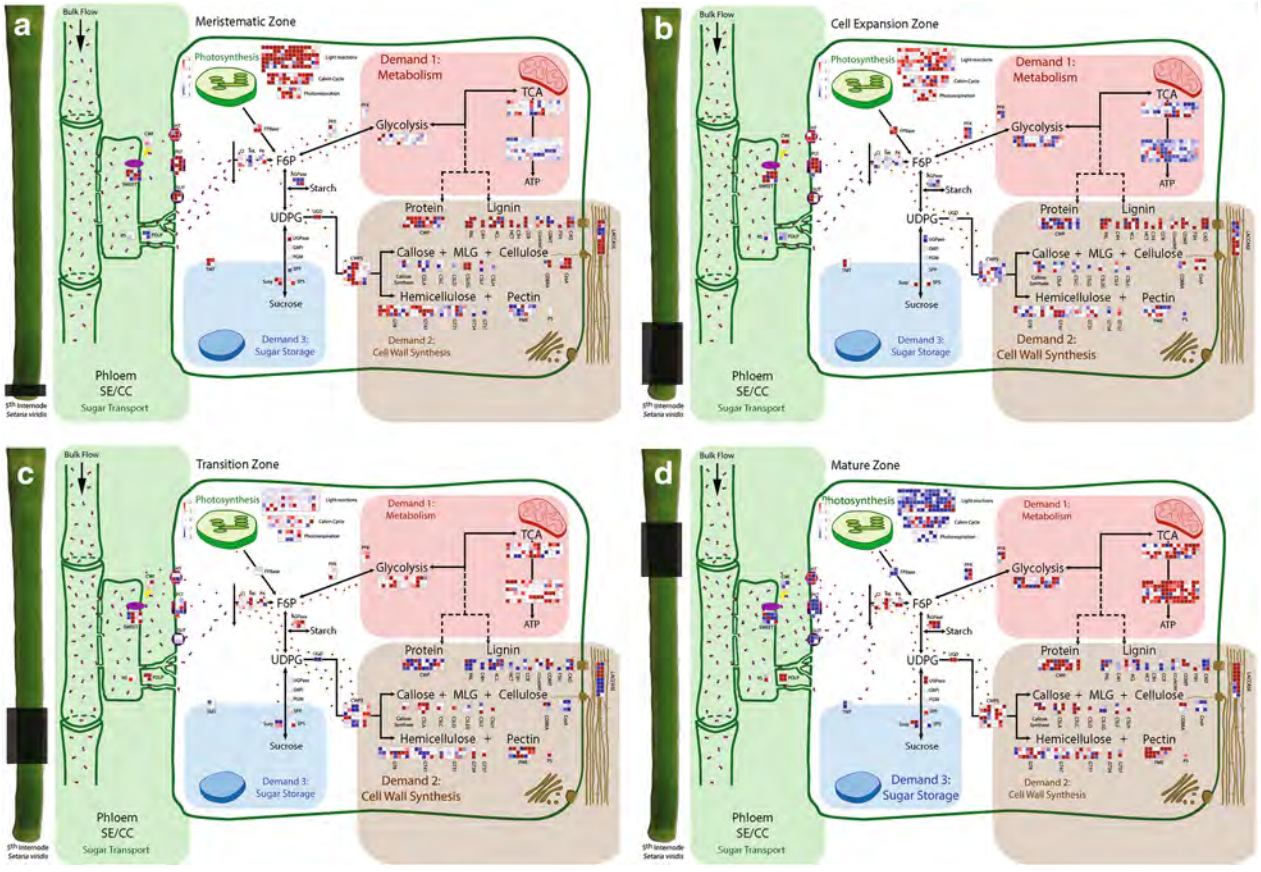


Fig. 6 Schematic model of carbon demands during internode development based on transcriptomic analysis. A schematic pathway of carbon import into the developing *S. viridis* fifth internode, showing the three major demands on carbon supply. A Mapman pathway file (Additional file 4 and Additional file 5) was created to generate this visualisation of gene expression with regard to each carbon demand (also using the updated *S. viridis* mapping file, Additional file 3, and the experiment file, Additional file 2). Genes were assigned into categories (as labelled on the diagram) and their expression values were displayed as squared $\log_2$ fold change from the squared average of the four internode zones for each gene. Up-regulated genes are shown in *blue*, whilst down-regulated genes are in *red*. The.mp4 animation (Additional file 6) changes from meristematic zone (**a**), to cell expansion zone (**b**), to transition zone (**c**), to mature zone (**d**) indicating the internode region with a transparent *grey box* on the left and highlighting the greatest demand for carbon in each region. *MLG* mixed linkage glucans, *RS* raffinose synthase, *PDLP* plasmodesmata localised proteins, *SWEET* sugars will eventually be exported transporters, *CWI* cell wall invertase, *HT* hexose transporters, *PLT* polyol transporters, *SUT* sucrose transporters, *CI* cytosolic invertases, *HK* hexokinase, *FK* fructokinase, *FPBase* fructose-1,6-bisphosphatase, *F6P* fructose-6-phosphate, *PFK* phospho-fructokinase, *AGPase* glucose-1-phosphate adenylyltransferase, *UGD* UDP-glucose 6-dehydrogenase, *UGPase* UDP-glucose phosphorylase, *G6PI* glucose-6-phosphate isomerase, *PGM* phosphoglucomutase, *SPP* sucrose phosphate phosphatase, *SPS* sucrose phosphate synthase, *Susy* sucrose synthase, *TMT* tonoplast membrane transporter, *CWPS* cell wall precursor synthesis, *CWP* cell wall protein, *PAL* phenylalanine ammonia-lyase, *C4H* cinnamate-4-hydroxylase, *4CL* 4-hydroxycinnamate CoA ligase, *HCT* hydroxycinnamoyl transferase, *C3H* coumarate 3-hydroxylase, *CCR* cinnamoyl-CoA reductase, *CCoAOMT* caffeoyl CoA 3-*O*-methyltransferase, *COMT* caffeic acid *O*-methyltransferase, *F5H* ferulate 5-hydroxylase, *CAD* cinnamyl alcohol dehydrogenase, *CSLA-H* cellulose synthase-like A-H, *CesA* cellulose synthase complex, *GT* glycosyltransferase, *PME* pectin methylesterase, *PS* pectin synthesis, *SE* sieve element, *CC* companion cell. *Blue* and *purple dots* are representative of sugar flow

switches to a, likely, apoplasmic step, and (iv) a mature zone containing thickened, lignified cell walls, that is actively accumulating and storing carbon in the form of sucrose and malate via the expression of energy-dependent sugar transport genes, suggesting a, likely, apoplasmic step during the unloading of photoassimilate into this tissue.

Dye feeding experiments in sugarcane [35] and sorghum [36] have shown that 5(6)-carboxyfluorescein (CF) moves symplasmically from vascular phloem to storage parenchyma cells in mature stem sink tissues. Interestingly, high expression of energy-dependent sugar transport genes within mature internodes has also been observed, in sorghum [37]; however, the protein cellular

localisation, and therefore precise role in sugar accumulation within grass stems has not yet been elucidated. The strong expression of *SWEET*, *SUT*, *HT*, *PLT*, and *CWI* genes in mature sink tissue of the *S. viridis* developing internode provides evidence that an apoplasmic step may exist in *Setaria*. Localisation and characterisation of these sugar transport proteins, however, will determine the cellular localisation of a possible apoplasmic unloading step, which may be vascular in origin or located on storage parenchyma cells to facilitate unloading of sugar into the apoplasm for storage and regulation of turgor. Whilst the role of these transporters in phloem unloading and assimilate storage in maturing internodes is yet to be investigated, this experimental system has provided gene candidates and a biological system for protein localisation and phenotyping of gene knockdown lines which will further unravel this biological process that is integral to controlling soluble sugar levels in C_4 panicoid grasses.

Similarly, the clear transition from primary cell wall synthesis to secondary cell wall synthesis in this system, coupled with gene expression data, has uncovered numerous *S. viridis* gene candidates which can now be manipulated and studied under controlled conditions to further our understanding of cell wall biosynthesis in C_4 grass stems. Cellulose synthase-like (CSL) families A, C, D, F, and H were all switched on in the meristematic and cell expansion zones of the internode, implicating a potential role in cell division and/or primary cell wall synthesis of hemicelluloses. As an example of gene discovery in this system, we will examine the proteins encoded by the CSLF and CSLH gene families. Homologues of CSLF and CSLH have been shown to produce mixed linkage glucans (MLGs) in primary cell walls of barley [38, 39] during wall expansion, and in secondary cell walls of rice stems [40]. The dataset presented here has identified a *CSLF* and a *CSLH* gene expressed in primary cell wall tissues, and a *CSLF* gene expressed in mature tissues (Fig. 6) differentiating transcriptional control over MLGs in primary and secondary cell walls of *S. viridis*. Characterisation in this biological system by protein localisation, promoter reporter constructs, and genetic knockdown using available *S. viridis* transformation techniques [12, 13] is now possible to further our understanding of the role of each of these genes in MLG deposition. Similarly, a number of gene isoforms can now be identified within this dataset for the manipulation and characterisation of lignin precursor synthesis, lignin polymerisation, cellulose synthesis, and hemicellulose synthesis, examining genes associated with both primary and secondary cell wall synthesis (Fig. 6). Expression of transcription factors in this experimental system was not characterised here but would also be of great interest when manipulating cell wall deposition and composition in this experimental C_4 panicoid grass system.

Conclusions

The developing internode characterised here is a valuable experimental system for gene discovery related to a variety of fundamental biological processes, especially for traits important in determining bioenergy crop yields. These include meristematic control of sink size, biosynthesis of cell walls, and accumulation of soluble sugars. Coupling this resource with genetic transformation techniques in *S. viridis* [12–14], generating controlled experimental data for the genetic manipulation of sink size, cell wall composition, and soluble sugar content in a C_4 grass model system is now possible.

Methods

Plant growth conditions

Setaria viridis A10 was grown in 2 L of 7:3:5:5 top soil:sand:perlite:crushed clay with 1 g/L of Plantacote depot 4M© (14:9:15 NPK) inside a growth cabinet with 16/8 h, 28/20C, and 55/70 %rH day/nights. Plants were illuminated with ~600 μmol s^{-1} m^{-2} PAR from fluorescent lamps during light hours. After germination, plants were thinned to leave one plant per pot and were given 150 mL of water (with 1.5 mL/L of Previcur® on the first watering) approximately 3 h into the light cycle every day.

Identification of a developing internode

Plants were harvested at seven developmental stages. These were defined as the 6-leaf stage (emerging but still straight), 9-leaf (emerging but still straight), booting, half emerged head, anthesis (3/4 visible pollen), milky dough, and hard dough. Dry weights of whole plants and internodes were measured by freeze drying for ~5 days to determine biomass. Plant heights and internode lengths were also measured. Whole stems were frozen slowly at −20 °C and thawed before centrifuging through glass wool filters to obtain the soluble fraction of the stem. Sucrose was measured in the soluble fraction by FTIR spectroscopy as previously described [19].

Histological determination of developmental zones within the internode

The fifth internode from the base of the main stalk harvested at the half emerged stage was photographed, dissected, and fixed in 4 % (*w/v*) paraformaldehyde/1.5 % (*w/v*) glutaraldehyde. Fixed tissue was halved longitudinally, embedded in 6 % agarose and sectioned longitudinally on a vibratome along the entire length of the internode at a thickness of 50 μm. These sections were stained with 10 μg/mL DAPI and viewed under

UV illumination on a Zeiss Axiophot microscope (Carl Zeiss Pty. Ltd., Oberkochen, Germany, http://www.zeiss.com). Nuclei, and therefore cells, were counted along the length of the internode using the ITCN plugin for imageJ (http://imagej.nih.gov/ij/plugins/itcn.html). Nuclei density and cell size were calculated in a sliding window by dividing the number of each within the window by the width of the window at 5-µm interval. Cell size and nuclei density were immediately matched with defining features of the unsectioned half of the internode. Other internodes were quartered transversely, dehydrated through an ethanol series, and infiltrated with Technovit© (Heraeus Kulzer, Hanau, Germany, http://www.heraeus-kulzer.com). Technovit© blocks were polymerised and 4-µm transverse sections were cut on a microtome. Sections were stained with 0.05 % Toluidine blue and counterstained with Lugol's iodine before viewing under brightfield illumination.

Harvest of a developing internode

Eighty plants were harvested at the half emerged head stage (see "Results" section for details) from two crops grown under the same conditions (40 plants per crop) between 8 and 11 h into the light cycle. Harvest was staggered as plants reached the half emerged stage from 25 to 28 days after sowing (DAS). Plants were photographed; height and fresh weight were measured to ensure uniformity, and the fifth internode was dissected immediately after removal from the growth cabinet. The leaf sheath was removed from the fifth internode, and the intercalary meristem, cell expansion, transition, and mature zones were dissected and snap frozen in liquid N_2. Since each dissected segment was only 5–15 mg, tissue was pooled from ten plants for each replicate to generate four biological replicate pools for each developmental zone.

Acetyl bromide lignin

Twenty mg of fresh, ground tissue was weighed and washed sequentially with water, ethanol, and acetone before being freeze dried overnight. Lignin was solubilised by incubating in a thermomixer at 55 °C and 600 rpm for 150 min in 25 % (*v/v*) acetyl bromide prepared with glacial acetic acid. After cooling, acetic acid: 2 M NaOH mixture (50:9 *v/v*) and 0.5 M hydroxylamine chlorhydrate were added and the absorbance was read at 280 nm. Acetyl bromide lignin content was calculated using an extinction coefficient of 20.48 that was previously determined for corn stover [41].

Primary metabolites

Soluble sugars and malate were measured enzymatically in ethanolic extracts from 20 mg fresh weight of frozen, ground tissue powder, and starch was measured enzymatically in the ethanol-insoluble residue, as previously described [42].

RNA extraction, library preparation, and sequencing

Samples were ground to a powder at −60 °C using a cryogenic grinding robot (Labman Automation, Middlesborough, UK; http://www.labman.co.uk). RNA was extracted from aliquots (80–100 mg FW) of frozen tissue powder using a standard TRIzol© (Life technologies, Carlsbad, USA, http://www.lifetechnologies.com) extraction method and precipitated with EtOH. DNase treatment was performed using Turbo DNA-*free*™ kit (Life technologies, Carlsbad, USA, http://www.lifetechnologies.com) as per the manufacturer's instruction. RNA quality was assessed by analysis with an RNA 6000 Nano kit and Bioanalyzer 2100 (Agilent Technologies; Santa Clara, CA, USA; http://www.agilent.com). RIN scores of above eight were considered acceptable. RNAseq libraries were generated with polyA enrichment and rRNA depletion using NEBNext® Ultra™ Directional RNA Library Prep Kit (NEB, Ipswich, USA, http://www.neb.com) for Illumina® and were sequenced on an Illumina® HiSeq 2500 at the Max Planck Genome Centre, Cologne, Germany.

RNAseq analysis

Illumina 100 bp paired-end reads from four biological replicates of the four internode zones were imported into CLC genomics workbench 8 (CLCbio, Aarhus, Denmark, http://www.clcbio.com) for adapter removal, trimming, mapping, and read counting. A single-end Illumina read library, generated from a mixture of leaf, stem, node, crown, root, spikelet, floret, and seed tissue across three developmental stages (seed germination, vegetative growth, and reproduction), was downloaded from the NCBI short read archive (SRA784127), imported into CLC genomics workbench 8, and analysed in parallel with the internode data. This 'whole plant' sample was used as an indication of background expression for the assessment of stem- and zone-specific gene expression.

Adapter sequences were trimmed from all reads. Low-quality reads were trimmed using a modified Mott trimming algorithm within CLC genomics 8. Reads with a quality score limit below 0.05 and regions containing more than two base ambiguities were trimmed. Reads were then mapped to the *S. italica* genome version 2.1 with gene annotations obtained from Phytozome 10.1 on the 29th April, 2015 [16], since an annotated *Setaria viridis* transcriptome and genome were not publically available. Mapping was performed using the principles of Mortazavi et al. [43] using a mismatch cost of 2, insertion cost of 3, deletion cost of 3, a length fraction of 0.8, and a similarity fraction of 0.8. Paired read distances were

set at 180–660 bp, based on Bioanalyser runs of the RNA libraries, and the maximum number of hits for a read was set at 10. Fragments per kilobase per million (FPKM) was calculated for each gene, and EDGE test [44] and ANOVA false discovery rates were calculated. Log_2 fold change between internode regions was calculated against the average of the four zones for each gene using x^2 transformed data. Mapman bins were assigned to genes using the Mapman Phytozome v9.0 mapping file for *S. italica* downloaded on the 12th May, 2015. Genes encoding sugar transporters and cell wall synthesis related enzymes/proteins were manually annotated by BLAST against maize, rice, and sorghum proteome sequences and Mapman bins were assigned accordingly. A file containing FPKM, log_2 fold change, Mapman bin numbers, and gene annotations can be found in Additional file 2. This file was used, along with our updated Mapman mapping file (Additional file 3) to generate Mapman visualisations of the RNAseq dataset.

Principal component analysis was performed in 'The Unscrambler® X 10.2' (CAMO) with standard deviation-weighted FPKM values using the NIPALS algorithm and full cross validation [45]. Venn diagrams were generated using 'Venny 2.0' [46]. Gene categories of interest were selected using the Mapman binning system in Additional file 3 and average FPKM was calculated for each category.

Additional files

Additional file 1. RNAseq mapping statistics. Number of reads, reads after trimming, mapped reads (%) and unique reads (%) for each RNAseq sample described in the manuscript.

Additional file 2. RNAseq data. A spreadsheet containing FPKM, log_2 fold change, Mapman bin numbers, and gene annotations for the RNAseq data generated for this manuscript.

Additional file 3. Updated Mapman mapping file. A mapping file that can be imported into Mapman containing some manual annotations that were used for this manuscript.

Additional file 4. Mapman pathway image. A schematic overview of the biological processes occurring in a developing *S. viridis* internode that can be imported into Mapman to map genes and metabolite levels onto this image.

Additional file 5. Mapman pathway file. A Mapman pathway file that maps relevant genes onto the image in Additional file 4.

Additional file 6. Schematic model of carbon demands during internode development based on transcriptomic analysis. Figure 6 from the manuscript compiled as an .mp4 video.

Abbreviations

DAPI: 4′,6-diamidino-2-phenylindole; MsZ: meristematic zone; CEZ: cell expansion zone; TZ: transition zone; MZ: mature zone; PCA: principle component analysis; FPKM: fragments per kilobase per million reads; SWEET: sugar efflux proteins; SUT: sucrose transporters; HT: hexose transporters; PLT: polyol transporters; CWI: cell wall invertase; TMT: tonoplast monosaccharide transporters; PDLP: plasmodesmatal localised proteins; CSL: cellulose synthase-like proteins.

Authors' contributions

APM, WMP, CPLG, and RTF conceived the study. APM designed and conducted all experiments, analysed all data, produced all figures, and drafted the manuscript. WMP contributed to data analysis, drafting of the manuscript, and production of figures. CB performed some preliminary RNAseq data analysis and reviewed the manuscript. CA optimised Setaria growth conditions and reviewed the manuscript. JEL provided intellectual input and resources for Setaria growth, RNA sequencing, and metabolite measurements. RTF, JEL, and CPLG revised the manuscript. All authors read and approved the final manuscript.

Author details

[1] School of Environmental and Life Sciences, University of Newcastle, University Drive, Callaghan, NSW 2308, Australia. [2] Department of Metabolic Networks, Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam, Germany. [3] ARC Centre of Excellence for Translational Photosynthesis, Research School of Biology, Australian National University, Canberra, ACT 2601, Australia. [4] CSIRO Agriculture Flagship High Resolution Plant Phenomics Centre, GPO Box 1600, Canberra, ACT 2601, Australia.

Acknowledgements

The authors would like to thank Marc Lohse for initial discussion about sequencing strategies, Armin Schlereth for assistance with preparation of samples for RNA sequencing, Eugenia Maximova for assistance with histology, Gretel Linich for retrieving datasets, Carlos Figueroa for assistance with starch measurements, Beatrice Encke and Regina Feil for assistance with metabolite measurements, and Grégory Mouille, Sylvie Citerne and Stéphanie Boutet for assistance in the measurement of cell wall components. The authors acknowledge the use of the Australian Plant Phenomics Facility, for plant growth and sample preparation in this work.

Competing interests

The authors declare that they have no competing interests.

Received: 28 October 2015 Accepted: 9 February 2016
Published online: 24 February 2016

References

1. Brutnell TP, Bennetzen JL, Vogel JP. *Brachypodium distachyon* and *Setaria viridis*: model genetic systems for the grasses. Annu Rev Plant Biol. 2015;66:465–85.
2. Brutnell TP, Wang L, Swartwood K, Goldschmidt A, Jackson D, Zhu X-G, Kellogg E, Van Eck J. *Setaria viridis*: a model for C4 photosynthesis. Plant Cell. 2010;22:2537–44.
3. Xu J, Li Y, Ma X, Ding J, Wang K, Wang S, Tian Y, Zhang H, Zhu X-G. Whole transcriptome analysis using next-generation sequencing of model species *Setaria viridis* to support C4 photosynthesis research. Plant Mol Biol. 2013;83:77–87.
4. Huang P, Feldman M, Schroder S, Bahri BA, Diao X, Zhi H, Estep M, Baxter I, Devos KM, Kellogg EA. Population genetics of *Setaria viridis*, a new model system. Mol Ecol. 2014;23:4912–25.
5. Rizal G, Acebron K, Mogul R, Karki S, Larazo N, Quick WP. Study of flowering pattern in *Setaria viridis*, a proposed model species for C4 photosynthesis research. J Botany. 2013;2013:7.
6. Li P, Brutnell TP. *Setaria viridis* and *Setaria italica*, model genetic systems for the panicoid grasses. J Exp Bot. 2011;62:3031–7.
7. Bennetzen J. First international Setaria genetics conference. In: Book First international Setaria genetics conference (Editor ed. eds.). City; 2014.
8. Jia G, Shi S, Wang C, Niu Z, Chai Y, Zhi H, Diao X. Molecular diversity and population structure of Chinese green foxtail [*Setaria viridis* (L.) Beauv.] revealed by microsatellite analysis. J Exp Bot. 2013;64:3645–56.
9. Layton DJ, Kellogg EA. Morphological, phylogenetic, and ecological diversity of the new model species *Setaria viridis* (Poaceae: Paniceae) and its close relatives. Am J Bot. 2014;101:539–57.

10. Qie L, Jia G, Zhang W, Schnable J, Shang Z, Li W, Liu B, Li M, Chai Y, Zhi H, Diao X. Mapping of quantitative trait locus (QTLs) that contribute to germination and early seedling drought tolerance in the interspecific cross *Setaria italica* × *Setaria viridis*. PLoS One. 2014;9:e101868.
11. Brutnell TP, Wang L, Swartwood K, Goldschmidt A, Jackson D, Zhu X-G, Kellogg E, Van Eck J. *Setaria viridis:* a model for C4 photosynthesis. Plant Cell Online. 2010;22:2537–44.
12. Van Eck J, Swartwood K. *Setaria viridis*. In: Wang K editor. Agrobacterium Protocols. Volume 1223. New York: Springer; 2015. 57–67: Methods in Molecular Biology].
13. Martins PK, Ribeiro AP, Cunha BADBd, Kobayashi AK, Molinari HBC. A simple and highly efficient Agrobacterium-mediated transformation protocol for *Setaria viridis*. Biotechnol Rep. 2015;6:41–4.
14. Martins PK, Nakayama TJ, Ribeiro AP, Cunha BADBd, Nepomuceno AL, Harmon FG, Kobayashi AK, Molinari HBC. *Setaria viridis* floral-dip: a simple and rapid Agrobacterium-mediated transformation method. Biotechnol Rep. 2015;6:61–3.
15. Jiang H, Barbier H, Brutnell T. Methods for performing crosses in *Setaria viridis*, a new model system for the grasses. J Vis Exp: JoVE. 2013:50527.
16. Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, Estep M, Feng L, Vaughn JN, Grimwood J, et al. Reference genome sequence of the model plant Setaria. Nat Biotechnol. 2012;30:555–61.
17. John CR, Smith-Unna RD, Woodfield H, Covshoff S, Hibberd JM. Evolutionary convergence of cell-specific gene expression in independent lineages of C4 grasses. Plant Physiol. 2014;165:62–75.
18. Petti C, Shearer A, Tateno M, Ruwaya M, Nokes S, Brutnell T, DeBolt S. Comparative feedstock analysis in *Setaria viridis* L. as a model for C4 bioenergy grasses and panicoid crop species. Front Plant Sci. 2013;4:181.
19. Martin AP, Palmer WM, Byrt CS, Furbank RT, Grof CP. A holistic high-throughput screening framework for biofuel feedstock assessment that characterises variations in soluble sugars and cell wall composition in *Sorghum bicolor*. Biotechnol Biofuels. 2013;6:1–13.
20. Jackson PA. Breeding for improved sugar content in sugarcane. Field Crops Res. 2005;92:277–90.
21. Morrison TA, Kessler JR, Buxton DR. Maize internode elongation patterns. Crop Sci. 1994;34:1055–60.
22. Kaufman PB, Brock TG. Structural Development of the Oat Plant. 677 S. Segoe Rd., Madison, WI 53711, USA: American Society of Agronomy and Crop Science Society of America; 1992.
23. Moore PH. Temporal and spatial regulation of sucrose accumulation in the sugarcane stem. Funct Plant Biol. 1995;22:661–79.
24. Moore PH, Paterson AH, Tew T. Sugarcane: the crop, the plant, and domestication. Sugarcane: physiology, biochemistry and functional biology .Wiley Blackwell, Oxford. 2014:1–17.
25. Hoffmann-Thoma G, Hinkel K, Nicolay P, Willenbrink J. Sucrose accumulation in sweet sorghum stem internodes in relation to growth. Physiol Plant. 1996;97:277–84.
26. Bosch M, Mayer C-D, Cookson A, Donnison IS. Identification of genes involved in cell wall biogenesis in grasses by differential gene expression profiling of elongating and non-elongating maize internodes. J Exp Bot. 2011;62:3545–61.
27. He C-y, Cui K, Zhang J-g, Duan A-g, Zeng Y-f. Next-generation sequencing-based mRNA and microRNA expression profiling analysis revealed pathways involved in the rapid growth of developing culms in Moso bamboo. BMC Plant Biol. 2013;13:119.
28. Fernie AR, Martinoia E. Malate. Jack of all trades or master of a few? Phytochemistry. 2009;70:828–32.
29. Chen L-Q, Qu X-Q, Hou B-H, Sosso D, Osorio S, Fernie AR, Frommer WB. Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. Science. 2012;335:207–11.
30. Kühn C, Grof CPL. Sucrose transporters of higher plants. Curr Opin Plant Biol. 2010;13:287–97.
31. Slewinski TL. Diverse functional roles of monosaccharide transporters and their homologs in vascular plants: a physiological perspective. Mol Plant. 2011;4:641–62.
32. Reuscher S, Akiyama M, Yasuda T, Makino H, Aoki K, Shibata D, Shiratake K. The sugar transporter inventory of tomato: genome-wide identification and expression analysis. Plant Cell Physiol. 2014;55:1123–41.
33. Ruan Y-L. Sucrose Metabolism: gateway to Diverse Carbon Use and Sugar Signaling. Annu Rev Plant Biol. 2014;65:33–67.
34. Hedrich R, Sauer N, Neuhaus HE. Sugar transport across the plant vacuolar membrane: nature and regulation of carrier proteins. Curr Opin Plant Biol. 2015;25:63–70.
35. Rae AL, Perroux JM, Grof CP. Sucrose partitioning between vascular bundles and storage parenchyma in the sugarcane stem: a potential role for the ShSUT1 sucrose transporter. Planta. 2005;220:817–25.
36. Milne RJ, Offler CE, Patrick JW, Grof CP. Cellular pathways of source leaf phloem loading and of phloem unloading in developing stems of Sorghum bicolor in relation to stem sucrose storage. Funct Plant Biol. 2015.
37. Milne RJ, Byrt CS, Patrick JW, Grof CPL. Are sucrose transporter expression profiles linked with patterns of biomass partitioning in Sorghum phenotypes? Front Plant Sci. 2013;4:223.
38. Lockhart J. Uncovering the unexpected site of biosynthesis of a major cell wall component in grasses. Plant Cell. 2015;27:483.
39. Doblin MS, Pettolino FA, Wilson SM, Campbell R, Burton RA, Fincher GB, Newbigin E, Bacic A. A barley cellulose synthase-like CSLH gene mediates (1,3;1,4)-β-D-glucan synthesis in transgenic Arabidopsis. Proc Natl Acad Sci. 2009;106:5996–6001.
40. Vega-Sánchez ME, Verhertbruggen Y, Christensen U, Chen X, Sharma V, Varanasi P, Jobling SA, Talbot M, White RG, Joo M, et al. Loss of cellulose synthase-like F6 function affects mixed-linkage glucan deposition, cell wall mechanical properties, and defense responses in vegetative tissues of rice. Plant Physiol. 2012;159:56–69.
41. Fukushima RS, Hatfield RD. Extraction and isolation of lignin for utilization as a standard to determine lignin concentration using the acetyl bromide spectrophotometric method. J Agric Food Chem. 2001;49:3133–9.
42. Stitt M, Lilley RM, Gerhardt R, Heldt HW. Metabolite levels in specific cells and subcellular compartments of plant leaves. Methods Enzymol. 1989:518–52.
43. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008;5:621–8.
44. Robinson MD, McCarthy DJ, Smyth GK. EdgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.
45. Martens H, Naes T. Multivariate calibration. Wiley; 1992.
46. An interactive tool for comparing lists with Venn's diagrams [http://bioinfogp.cnb.csic.es/tools/venny/index.html].

CHAPTER FIVE

In situ imaging techniques for understanding structure, composition and function of plant tissues

5.1 Summary

This chapter outlines two imaging techniques that will greatly enhance the study of genetic manipulations in *S. viridis* and our understanding of the molecular and structural workings of plants. The first is a technique for imaging *in situ* proteins by immunohistochemistry and expressed fluorescent proteins within whole plant organs in 3D. This work is published in *Scientific Reports (Nature)*. The second is an adaption of a published technique for Raman confocal imaging of cell wall composition. The published method was applied to *S. viridis* for the first time to show its applicability in assessing cell specific, *in situ* cell wall compositional changes in *S. viridis* plants. This work was performed by myself and is unpublished, however, a description of the results is described in the body of this thesis.

5.2 Publication 5: PEA-CLARITY: 3D molecular imaging of whole plant organs

Palmer WM, Martin AP, Flynn JR, Reed SL, White RG, Furbank RT, Grof CPL **PEA-CLARITY: 3D molecular imaging of whole plant organs** *Scientific Reports (Nature)* **5**, Article number: 13492 doi:10.1038/srep13492 (2015).

Statements of contribution of others: Publication 5

I attest that Research higher degree candidate ***Antony Martin*** *contributed extensively to the development of this publication as co-primary author. The author contributions were as follows:*

WMP, APM and JRF conceived of the study. WP conceived of the enzymatic treatment and initial protocol development. WMP and APM planned and conducted experiments, produced all figures and wrote the draft manuscript. SLR processed GFP and CFP A. thaliana samples and aided in imaging. RGW, RTF and CPLG provided resources, helpful feedback and reviewed the manuscript.

_____

William Palmer

_____

Antony Martin

_____

Jamie Flynn

Stephanie Reed

_____

Rosemary White

_____

Robert Furbank

_____

Christopher Grof

SCIENTIFIC REPORTS

OPEN

PEA-CLARITY: 3D molecular imaging of whole plant organs

William M. Palmer[1], Antony P. Martin[1], Jamie R. Flynn[2], Stephanie L. Reed[1], Rosemary G. White[3], Robert T. Furbank[4] & Christopher P. L. Grof[1]

Received: 17 July 2015
Accepted: 27 July 2015
Published: 02 September 2015

Here we report the adaptation of the CLARITY technique to plant tissues with addition of enzymatic degradation to improve optical clearing and facilitate antibody probe penetration. Plant-Enzyme-Assisted (PEA)-CLARITY, has allowed deep optical visualisation of stains, expressed fluorescent proteins and IgG-antibodies in Tobacco and Arabidopsis leaves. Enzyme treatment enabled penetration of antibodies into whole tissues without the need for any sectioning of the material, thus facilitating protein localisation of intact tissue in 3D whilst retaining cellular structure.

Fixation and embedding of plant tissue for molecular interrogation using techniques such as histological staining, immunohistochemistry or *in situ* hybridisation has been the foundation of cell biology studies for decades. Applying these techniques for 3D tissue analysis is seriously limited by the need to section the tissue, image each section, and then reassemble the images into a 3D representation of the structures of interest. Here we present a fundamental shift from the two dimensional plane to that of three dimensions whilst retaining molecular structures of interest without the need to section the plant tissue. Recent advances in fixation and 'clearing' techniques such as SeeDB[1], Sca*l*eA2[2], 3DISCO[3], CLARITY[4] and its recent variant PACT[5] enabled intact imaging of whole embryos, brains and other organs in mouse and rat models. The new CLARITY system fixes and binds tissues within an acrylamide mesh structure. Proteins and nucleic acids are covalently linked to the acrylamide mesh by formaldehyde, then optically interfering lipid structures of animal cell membranes are removed using detergent (SDS). This renders such tissue optically transparent and suitable for deep imaging of up to ~5 mm using confocal microscopy[4].

Three dimensional imaging of plants using confocal microscopy has been limited to already semi-transparent tissue types such as root tips or meristems but resolution becomes limiting in cells deeper within tissues[6]. Other plant specific imaging techniques including modified pseudo-Schiff propidium iodide (mPS-PI) staining do allow for deep optical penetration, although the clearing steps also remove proteins and nucleic acids[7,8]. A major hindrance to applying techniques such as CLARITY to plant material is the cell wall, comprised mainly of cellulose, hemicellulose, lignin and pectin, which is permeable only to molecules under 60 kDa[9]. This creates a significant permeability barrier as common IgG antibodies used in immunohistochemistry are ~150 kDa in size and therefore unable to penetrate cell walls. PEA-CLARITY overcomes this limitation by using cell wall degrading enzymes to increase wall permeability, together with starch hydrolysing enzymes to reduce optical interference from starch grains. Cell wall degrading enzymes have been used to achieve 3D immunofluorescence within *A. thaliana* apical meristems however with harsh enzymatic degradation, the tissue lost structural integrity[10]. A contrasting method using urea as clearing agent, together with enzyme treatment, generated 3D structural images of plant tissues to localise nuclei and cell walls simultaneously. They also retained fluorescence from transiently expressed mTalin-citrine, revealing intact actin microfilaments deep within cleared tobacco leaves[11]. Nevertheless, this protocol used sectioned pea root nodules to aid antibody penetration and could only use short enzyme treatments to avoid structural damage to the tissue. The PEA-CLARITY

[1]School of Environmental and Life Sciences, University of Newcastle, Callaghan, NSW, 2308, Australia. [2]School of Biomedical Sciences and Pharmacy, University of Newcastle, Callaghan, NSW 2308, Australia. [3]CSIRO Agriculture, Black Mountain, ACT, 2601, Australia. [4]ARC Centre of Excellence for Translational Photosynthesis, Australian National University, Acton, ACT, 2601, Australia. Correspondence and requests for materials should be addressed to W.M.P. (email: william.moreau.palmer@gmail.com) or C.P.L.G. (email: chris.grof@newcastle.edu.au)

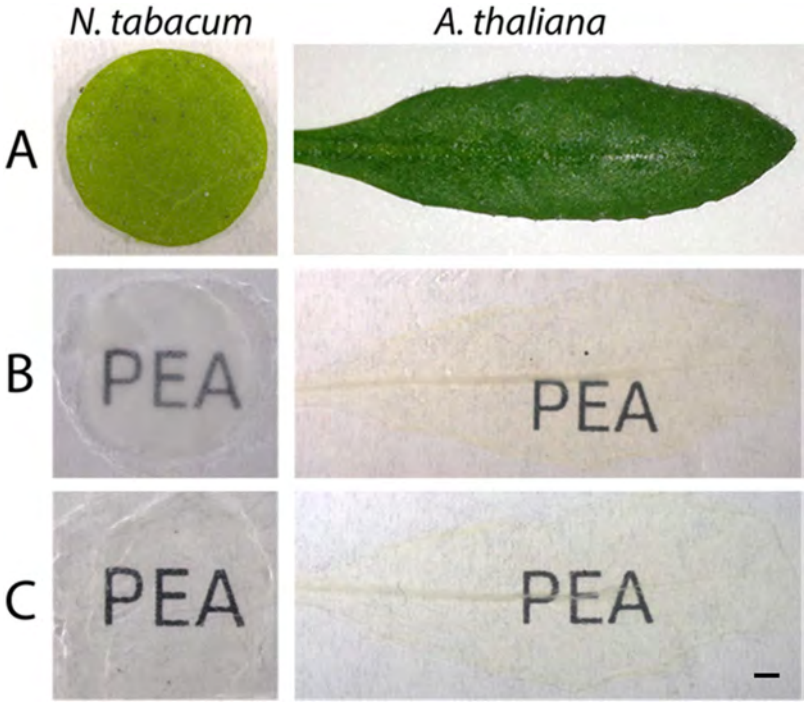


Figure 1. PEA-CLARITY clearing of *N. tabacum* and *A. thaliana* leaves. (**A**) fresh leaf disc from a fully expanded *N. tabacum* leaf (left) and a whole fully expanded *A. thaliana* leaf (right). (**B**) fixed, hydrogel embedded, passively cleared leaves. (**C**) cleared cell wall enzyme treated leaves for immunohistochemistry and imaging. Scale bar: 1 mm

technique described here combines hydrogel fixation, tissue clearing and enzymatic degradation, allowing visualisation of expressed fluorescent proteins, including GFP and CFP, simultaneously with immunofluorescence labelling using common IgG antibodies in whole, intact tissues, without loss of structural integrity.

PEA-CLARITY was applied to leaf tissue from two model species, *A. thaliana* and *N. tabacum* (Fig. 1A), initially to confirm that molecular structures remained intact during fixation, hydrogel polymerization and clearing. During the fixation and clearing process the leaf tissue became transparent due to the removal of lipids, chlorophyll and other pigments (Fig. 1B). Following clearing, cell walls and starch were enzymatically degraded to allow passage of antibodies during immunolabelling and to slightly increase transparency of the tissue (Fig. 1C).

Tobacco leaf tissue fixed in hydrogel and cleared in SDS but not enzymatically degraded was stained with propidium iodide (PI, red-nucleus) and calcofluor white (CW, green—cell wall) to assess the retention of DNA and cellular structure (Fig. 2). Using CLSM the whole, intact tobacco leaf disc was imaged and the z-stacks were assembled into a raw, unprocessed 3D slice reconstruction (Fig. 2B). Optical penetration through the entire leaf thickness (~150 μm) was achieved. Vascular, epidermal, palisade mesophyll and spongy mesophyll cell structure was maintained with all nuclei fixed within the cytosol of the cell. Even long trichome protrusions remained perpendicular to the epidermal surface layer across the entire leaf disc (see supplementary 1 for basal trichome orientation; full trichomes were not imaged). Furthermore the staining of nuclei by PI demonstrated the retention of nuclear DNA. This technique also allowed deep optical penetration into the vascular bundle with nucleus of the sieve element / companion cell (SE/CC) clearly defined without the need for sectioning or epidermal peeling.

Mature leaves from *A. thaliana* lines stably transformed with GFP localised to the endoplasmic reticulum (C16251; ER-GFP) or with CFP localised to peroxisomes (C16259; Px-CFP)[12], were CLARITY treated to assess the retention of endogenous fluorescent proteins. Strong ER-GFP and Px-CFP was observed throughout cleared *A. thaliana* leaves, including the internal leaf phloem at a depth of ~80 μm (Fig. 3).

Mature leaves from the *N. tabacum* stably transformed line, Sv-40, with GFP localised to nuclei[13] was processed using the PEA-CLARITY treatment, incubated with cell wall digestion enzymes and immunolabelled with a tobacco RuBisCO polyclonal antibody[14] to demonstrate the retention of proteins and tissue penetration of IgG antibodies for 3D whole tissue immunolocalisation. Immunolabelling of RuBisCO was observed in intact chloroplasts of both the palisade and spongy mesophyll cells (Fig. 4). Strong, nuclear localised GFP fluorescence was also observed indicating that GFP fluorescence was maintained in the correct subcellular compartment through enzyme treatment. *N. tabacum* negative controls and

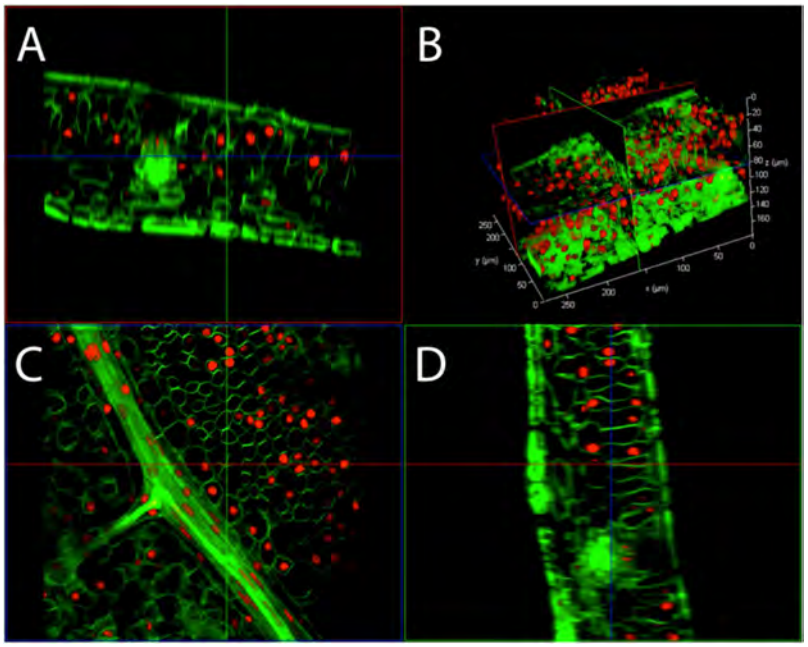


Figure 2. Structurally intact 3D projection of a cleared *N. tabacum* leaf showing retention of nuclear DNA. CLSM 3D projection of a passively cleared (without cell wall enzyme digestion) *N. tabacum* leaf showing nuclei stained with propidium iodide (red), and cell walls stained with calcofluor white (green). The 3D projection with epidermal layer cut away is shown in (**B**) and the x, y, z slices are shown in (**A, D**) and (**C**) respectively. The 3D video file can be viewed in supplementary 1.

further examples of immunolabelling following PEA-CLARITY treatment in the model grass species, *Setaria viridis*, can be found in supplementary Figs 2 & 3 respectively.

Here PEA-CLARITY has been demonstrated as a powerful technique allowing the use of common molecular probes, including stains, endogenous fluorescent proteins and immunohistochemistry, enabling the localisation of cell components in whole unsectioned tissues. Whilst *in situ* hybridization using RNA probes has not been demonstrated here, it is theoretically possible, as demonstrated in animal tissues[4].

As a proof of concept, mature leaves of two model species, *A. thaliana* and *N. tabacum,* were used in this study but it is anticipated that the PEA-CLARITY protocol will be applicable to a wide range of species although it may require tailoring to individual species and tissue types. Preliminary trials were conducted applying this protocol to an array of species and tissues and it was noted that some tissues, usually older or those with high sugar contents, developed a brown coloration during the clearing process (supplementary 4). Other drawbacks include: incomplete clearing of some tissues, fragility of tissue after enzymatic treatment, and fading of GFP fluorescence after extended passive clearing (~6 months or greater). We found that impure enzyme preparations may contain nucleases and proteases that degrade certain cell contents, and note that infusion of antibodies through cell walls remains difficult after incomplete enzyme digestion. Application of intermittent vacuum during enzyme and antibody incubations greatly improved penetration when compared to non-vacuum treated samples. Clearing was improved in many tissues by harvesting at the end of the dark photoperiod which reduced levels of starch.

Barriers to further clearing and penetration of larger fluorophores include tissues containing waxes and phenolics, for example, thick epidermal surfaces and lignified tissues. While no commercially available ligninases and cutinases are currently available, enzymes such as these may be developed in future and chemical treatments are being explored to potentially improve clearing of waxy and phenolic tissues. In addition, optimisation of tissue specific cell wall degrading enzyme cocktails may further enhance antibody penetration and optical clarity. The enzyme degradation step presented here in PEA-CLARITY may also be of benefit to mammalian systems, aiding in clearing and fluorophore penetration of difficult animal tissues. Also in animal tissues, light sheet microscopy has been used to reduce image acquisition times, decrease tissue bleaching, and improve z-plane resolution when generating 3D reconstructions[15].

The adaptation of CLARITY clearing and imaging protocols to plant tissues paves the way for 3D molecular interrogation of intact plant samples. Implementation of PEA-CLARITY will allow a more complete understanding of the relationship between structure and function within plant organs and of spatially regulated molecular processes.

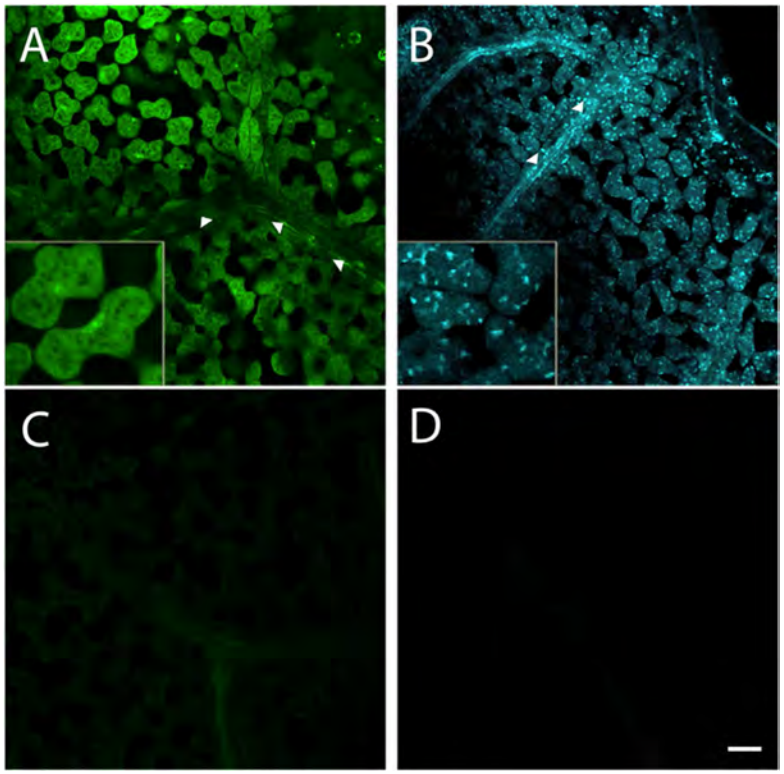


Figure 3. Two-dimensional optical sections of passively cleared, whole *A. thaliana* leaves showing retention of GFP and CFP fluorescence. Whole passively cleared (without cell wall enzyme treatment), whole mount *A. thaliana* leaves containing (**A**) endoplasmic reticulum localised GFP with WT negative control (**C**) and (**B**) peroxisome localised CFP with WT negative control (**D**). All were imaged through the internal phloem (arrows) of the leaf at a depth of ~80μm. Scale bar: 50μm

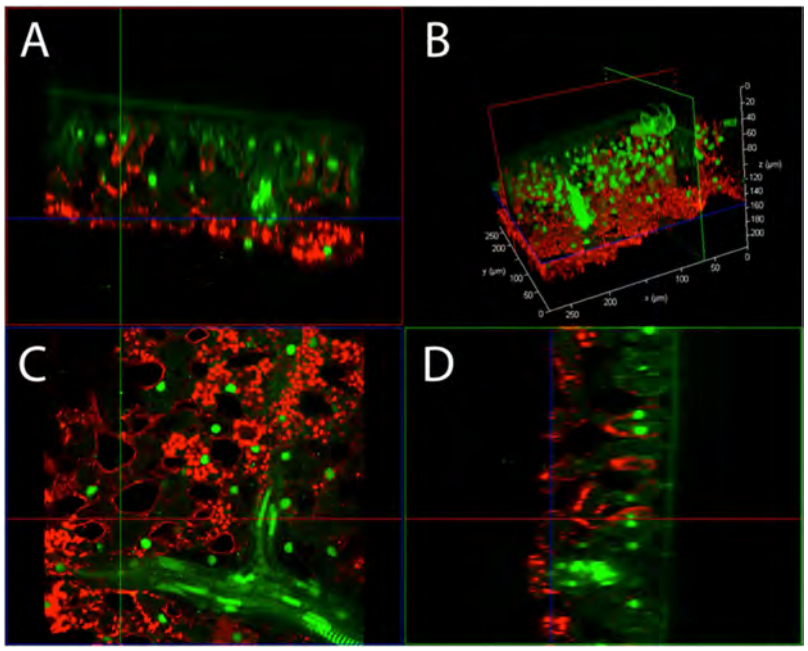


Figure 4. CLSM 3D projection of a PEA-CLARITY treated *N. tabacum* leaf showing immunostaining of RuBisCO and retention of GFP fluorescence. CLSM 3D projection of a passively cleared, cell wall enzyme treated (PEA-CLARITY) Sv-40 (nuclear localised GFP-green) *N. tabacum* leaf, immunostained with tobacco RuBisCO primary and Cy5 secondary antibodies (red). The 3D projection is shown in (**B**) and the x, y, z slices are shown in (**A**, **C**, **D**) respectively.

Experimental Procedures

Plant Growth Conditions. *Nicotiana tabacum* plants were grown in a glasshouse in Newcastle, Australia with supplemented sunlight for 16 hrs light and 8 hrs dark. *Arabidopsis thaliana* GFP and CFP lines C16251 and C16259[12] were grown in a growth cabinet with 16 hr/8 hr, 23/18 °C day night cycles and photon flux of ~150 μmol. m^{-2}. s^{-1} during light hours.

Tissue Fixation and Clearing. *N. tabacum* leaf discs were excised using a 7 mm leaf punch 30 min prior to commencement of the light period thus limiting starch accumulation. Discs were then immediately drop fixed into CLARITY hydrogel solution (see supplementary 6) and kept on ice. Whole leaves of *A. thaliana* were also harvested 30 mins prior to commencement of the light period and drop fixed as above. All tissues were put under vacuum for 1 hr at −100 kPa whilst on ice to facilitate infiltration of the hydrogel. Samples were kept at 4 °C overnight in hydrogel solution. Each disc/leaf was transferred to a separate tube, filled with hydrogel, taking care to remove air bubbles before being sealed and polymerised at 37 °C overnight. Excess hydrogel was removed after polymerisation and samples transferred to 50 ml 4% SDS clearing solution buffered with boric acid (see supplementary 6). Samples were passively cleared at 37 °C with gentle agitation for 4 to 6 weeks (or until clear), changing the clearing solution daily. Once cleared, samples were stained, or those containing endogenous fluorescence were mounted for imaging.

Staining. Cleared tobacco leaf discs were stained with 0.1% aqueous propidium iodide and/or (depending on enzyme cocktail mix) 0.05% aqueous calcofluor white for 20 mins in dark before being washed with 0.001% NaN_3 in PBS pH 7.4.

Enzyme Degradation. Enzymatic degradation of the cell wall was achieved by firstly washing the cleared sample with 0.001% NaN_3 in PBS pH 7.4 with at least 3 changes to remove all SDS (SDS inhibits enzymatic activity). Once all SDS was removed samples were carefully placed into the enzyme cocktail mix containing cellulase, xylanase, arabinofuranosidase, pectate lyase and α-amylase (see supplementary 6) and vacuum infiltrated 3 × 5 mins at −100 kPa. Samples were then incubated at 37 °C with very gentle agitation in darkness. Vacuum infiltration was repeated daily for 5–7 days and samples were transferred into fresh enzyme solution after 3 days. Samples were removed from the enzyme cocktail with care, before washing 3 times with 0.001% NaN_3 in PBS pH 7.4.

Immunofluorescence localisation of RuBisCO. Tissues were placed into solution containing anti-RuBisCO[14] diluted 1:100 in PBST pH 7.4 and incubated for 5 days with intermittent vacuum infiltration at −100 kPa for 3 × 5 mins, 3 × daily. Samples were then washed for 24 hrs in 50 ml PBST pH 7.4 with three solution changes and very gentle agitation. After washing, tissues were transferred into anti-rabbit Cy5 secondary antibody (1:200 in PBST pH 7.4) and incubated for 5 days with intermittent vacuum infiltration as above. Samples were washed in 0.001% NaN_3 in PBS pH 7.4 for 24 hrs and were mounted for imaging.

Imaging. Samples were mounted in PBS pH 7.4 and imaged with a Leica SP8 CLSM using a Leica 20 × na = 0.5 water immersion objective, white light laser and Hybrid Detectors at a resolution of 1024 × 1024 pixels. For full microscope settings (see metadata supplementary). All images displayed in this article are raw images taken from the Leica SP8. Unprocessed 3D reconstructions were performed in the Leica Applications Suite—Fluorescence (LAS-AF) software.

References

1. Ke, M.-T., Fujimoto, S. & Imai, T. SeeDB: a simple and morphology-preserving optical clearing agent for neuronal circuit reconstruction. *Nat Neurosci* **16,** 1154–1161 (2013).
2. Hama, H., Kurokawa, H., Kawano, H., Ando, R., Shimogori, T., Noda, H., Fukami, K., Sakaue-Sawano, A. & Miyawaki, A. Scale: a chemical approach for fluorescence imaging and reconstruction of transparent mouse brain. *Nat Neurosci* **14,** 1481–1488 (2011).
3. Ertürk, A., Becker, K., Jährling, N., Mauch, C. P., Hojer, C. D., Egen, J. G., Hellal, F., Bradke, F., Sheng, M. & Dodt, H.-U. Three-dimensional imaging of solvent-cleared organs using 3DISCO. *Nat Protocols* **7,** 1983–1995 (2012).
4. Chung, K., Wallace, J., Kim, S.-Y., Kalyanasundaram, S., Andalman, A. S., Davidson, T. J., Mirzabekov, J. J., Zalocusky, K. A., Mattis, J. & Denisin, A. K. *et al.* Structural and molecular interrogation of intact biological systems. *Nature* **497,** 332–337 (2013).
5. Yang, B., Treweek Jennifer, B., Kulkarni Rajan, P., Deverman Benjamin, E., Chen, C.-K., Lubeck, E., Shah, S., Cai, L. & Gradinaru, V. Single-Cell Phenotyping within Transparent Intact Tissue through Whole-Body Clearing. *Cell* **158,** 945–958 (2014).
6. Truernit, E. Phloem imaging. *Journal of Experimental Botany* **65,** 1681–1688 (2014).
7. Truernit, E., Bauby, H., Dubreucq, B., Grandjean, O., Runions, J., Barthélémy, J. & Palauqui, J.-C. High-Resolution Whole-Mount Imaging of Three-Dimensional Tissue Organization and Gene Expression Enables the Study of Phloem Development and Structure in Arabidopsis. *The Plant Cell* **20,** 1494–1503 (2008).
8. Nguyen, S. T. & McCurdy, D. High-resolution confocal imaging of wall ingrowth deposition in plant transfer cells: Semi-quantitative analysis of phloem parenchyma transfer cell development in leaf minor veins of Arabidopsis. *BMC Plant Biology* **15,** 1–14 (2015).
9. Read, S. & Bacic, A. Cell Wall Porosity and Its Determination. In *Plant Cell Wall Analysis*. Volume 17. Edited by Linskens H., Jackson J. Springer Berlin Heidelberg, 63–80, 1996: *Modern Methods of Plant Analysis*.
10. Rozier, F., Mirabet, V., Vernoux, T. & Das, P. Analysis of 3D gene expression patterns in plants using whole-mount RNA *in situ* hybridization. *Nat Protocols* **9,** 2464–2475 (2014).

11. Warner, C. A., Biedrzycki, M. L., Jacobs, S. S., Wisser, R. J., Caplan, J. L. & Sherrier, D. J. An Optical Clearing Technique for Plant Tissues Allowing Deep Imaging and Compatible with Fluorescence Microscopy. *Plant Physiology* **166,** 1684–1687 (2014).
12. Nelson, B. K., Cai, X. & Nebenführ, A. A multicolored set of *in vivo* organelle markers for co-localization studies in Arabidopsis and other plants. *The Plant Journal* **51,** 1126–1136 (2007).
13. Hanson, M. R. & Köhler, R. H. GFP imaging: methodology and application to investigate cellular compartmentation in plants. *Journal of Experimental Botany* **52,** 529–539 (2001).
14. Whitney, S. M. & Andrews, T. J. The Gene for the Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (Rubisco) Small Subunits That Assemble into Rubisco. *The Plant Cell* **13,** 193–206 (2001).
15. Tomer, R., Ye, L., Hsueh, B. & Deisseroth, K. Advanced CLARITY for rapid and high-resolution imaging of intact tissues. *Nat Protocols* **9,** 1682–1697 (2014).

Acknowledgments

We would like to thank Mark Talbot for helpful suggestions during CLSM measurements, David McCurdy for providing the Sv-40 *N. tabacum* line, Spencer Whitney for providing the RuBisCO antibody, Vivien Rolland for initial CLSM investigations, and Joe Enright for growing the plants.

Author Contributions

W.P., A.M. and J.F. conceived of the study. W.P. conceived of the enzymatic procedure and conducted initial protocol development. W.P. and A.M. planned and conducted experiments, produced all figures and wrote the draft manuscript. S.R. processed G.F.P. and C.F.P. *A. thaliana* samples and aided in imaging. R.W., R.F. and C.G. provided resources, helpful feedback and reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Palmer, W. M. *et al.* PEA-CLARITY: 3D molecular imaging of whole plant organs. *Sci. Rep.* **5**, 13492; doi: 10.1038/srep13492 (2015).

5.3 Unpublished data: Raman imaging of *Setaria viridis* cell walls

Raman confocal imaging of plant cell walls, *in situ*, has been used since 1985 when the orientation of lignin in black spruce was investigated [223]. Since then it has been used with many species for the investigation of cell wall polymer orientation and composition, including New Zealand Flax [224], rice, polypogon, sweet corn, onion and carrot [225], and poplar wood [226]. Recently, a detailed protocol was published outlining Raman confocal imaging of major cell wall polymers, including lignin [227]. Using this methodology and equipment, imaging of cell wall polymers to a resolution of <0.5 µm was possible [227].

This methodology and equipment, located at the Max Planck Institute of Molecular Plant Physiology, was used to image lignin in a mature region of a developing *S. viridis* internode for the first time. Lignin-rich cell walls were observed in xylem, and schlerenchyma cells and cellulose-rich cell walls were observed in phloem and parenchyma cells (**Figure 5.3.1**). A chemical map of lignin was generated using the area under the raman spectral peak at 1100-1200 cm^{-1} which showed finer detail of lignin distribution in epidermal cell walls and in cell corners of outer parenchyma cells. These results showed that Raman confocal imaging of lignin was applicable to *S. viridis* stem sections.

The methods used for imaging were exactly as outlined in Gierlinger *et. al.*, 2012 [225] and was performed using the same equipment and data analysis pipeline. Since this method was simply applying an established technique to a new species, publication will not be pursued, however, it was included in this dissertation for completeness. We believe it will facilitate an understanding of cell wall biosynthesis through the identification of cell specific cell wall variations in genetically manipulated *S. viridis* plants. It will, therefore, be an important methodology in a toolkit that will ultimately lead to genetic improvement of bioenergy crops.

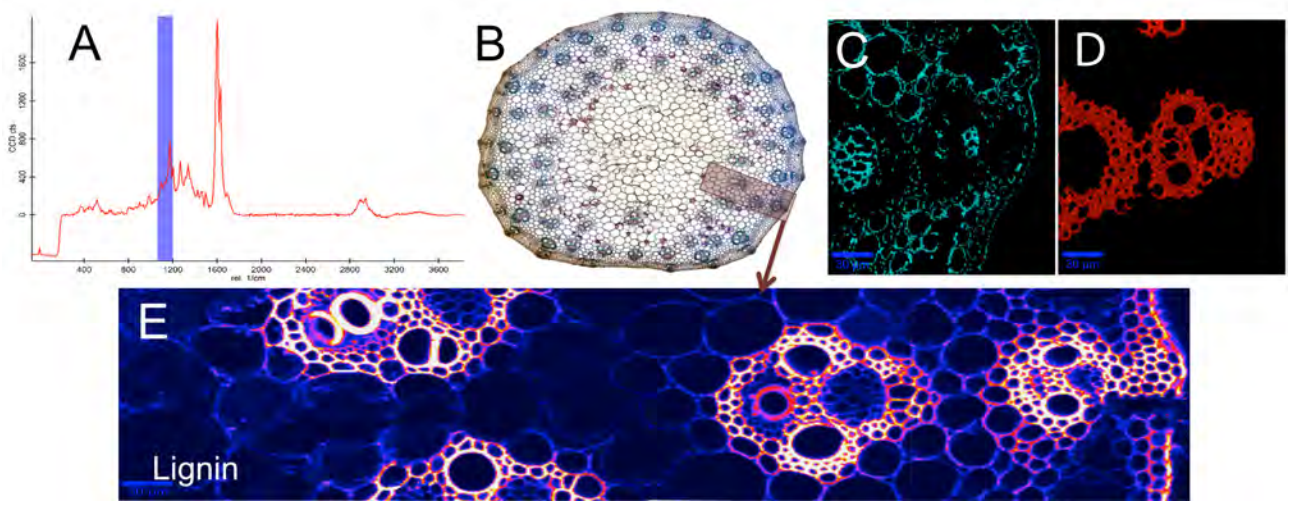


Figure 5.3.1. Raman confocal imaging of lignin in a mature region of a *Setaria viridis* internode. Raman spectra **(A)** were collected on cryosections from the outer region of the mature zone (MZ) of a developing *S. viridis* internode **(B)**. Cellulose-rich cell walls (cyan) and lignin-rich cell walls (red) were spectrally separated using principle component (PC) analysis, and the PC scores were mapped onto the image **(C** and **D** respectively**)**. The integrated area under a raman peak at 1100-1200 cm^{-1} associated with lignin **(A)** [227] was calculated for each pixel on the raman image to produce a chemical map of lignin **(E)**.

CHAPTER SIX

General discussion

The rapid improvement of bioenergy crops will be integral to the survival of the biofuel industry. Here I have presented a toolkit that will enable such rapid improvements and have presented the findings as three major advances:

1) A high throughput screening methodology, utilising FTIR spectroscopy and mathematical predictive modelling, was developed in the bioenergy crop *S. bicolor*. Bioenergy crops can now be screened for variations in biomass, cell wall degradability, and soluble sugar content driving bioenergy specific plant breeding, trait discovery, and genetic linkage maps in a high throughput and efficient manner.
2) *Setaria viridis*, the wild relative of foxtail millet, was developed as an experimental system for generating a fundamental understanding of the genetic and physiological factors that contribute to vegetative biomass production, cell wall composition and soluble sugar content in bioenergy crops. Using transcriptomic and metabolite profiles of a developing internode as a resource for gene discovery, there is now a system for investigating the effects of genetic manipulation of bioenergy crops in a rapid and efficient manner.
3) An imaging technique for 3D visualisation of plant structures and *in situ* protein localisation named PEA-CLARITY was developed. Raman confocal microscopy of cell wall lignin composition was also shown to be an effective method for imaging the spatial distribution of cell wall components in the model C_4 grass *Setaria viridis*. Together, these two imaging technique will move our understanding from tissue level gene discovery (**chapter 4**) to cellular level organisation of proteins in a 3D space, and spatial distribution of cell wall composition.

6.1 High-throughput phenotypic screening to drive bioenergy crop improvement

Phenotyping of plant diversity under different environmental conditions and genetic backgrounds has been identified as the major bottleneck to crop improvement in the genomic era [228]. Sequencing technologies have improved exponentially, however, to interpret this expanse of genetic data, it must be

coupled with high-resolution phenomics data. Even without genomic data, large seed banks containing a huge diversity of species and traits exist around the world [229], but without phenotypic data to accompany them, mining this diversity for breeding is not feasible.

Phenotyping platforms have seen a rapid improvement over the last 10 years with most utilising electromagnetic radiation, from visible light up to infrared, coupled with computational interpretation, for non-destructive 'imaging' of plant phenotypes. Automated platforms for imaging and interpreting plant morphology in 3D, thermal imaging for crop stress, chlorophyll fluorescence imaging for crop photosynthetic performance, and hyperspectral imaging for crop composition are all under constant development [228]. They are, however generally expensive to setup and maintain, and are either applicable to small model species on a lab scale, or by bulk imaging of crops from above. None of these techniques are, currently, easily deployable on a small scale, or provide organ level resolution for 'real-world' field phenotyping and have not been tailored for examining bioenergy traits other than simple measures of biomass and growth rates.

In 2013, we published a holistic high-throughput screening method that utilised Fourier Transform Mid-infrared (FTIR) spectroscopy, coupled with partial least square algorithms for screening cell wall and soluble sugar composition of *S. bicolor* stem tissue [230]. We also included a mathematical approximation of stalk biomass, allowing a holistic high-throughput assessment of yield and quality of *S. bicolor* stalks. This methodology was designed to be simple and cheap to setup and deployable from small to large scale. Since then, this approach has been improved upon with near infrared modelling of more detailed cell wall components including hemicellulose, cellulose, lignin and stalk sugar yields [231, 232]. These advances, coupled with the simple, cheap, holistic approach presented here can now be used to screen diverse populations in high-throughput and high-resolution for chemical phenotyping of bioenergy traits without expensive wet chemistry or bioreactor testing.

Numerous mutagenesis, recombinant inbred, and natural diversity populations exist for most potential bioenergy crops. Included in this dissertation is a *S. bicolor* EMS mutagenesis population (**3.3 publication 3**) generated and propagated, now to the M6 generation, in Gatton, Queensland. Additionally, during the course of this project, a *S. bicolor* natural diversity panel and a recombinant inbred population was propagated and sampled for high-throughput screening. In the literature there are also numerous reports of *S. bicolor* populations for which genomic data is available [233, 234]. High-throughput screening for bioenergy traits could be applied to these populations, utilising the genomic data for quantitative trait loci (QTL) and single nucleotide polymorphism (SNP) discovery considering genomic data is already available. This rationale applies to switchgrass, *miscanthus*, maize, rice, and sugarcane (although confounded by its complex genome) populations that currently exist. The holistic screening methodology is applicable to any crop species with similar morphology and composition to sweet sorghums, although mathematical modelling would need to be recalibrated.

Additionally, the application of this high-throughput bioenergy phenotyping platform to a C_4 grass model species will rapidly improve our understanding of the relationship between genotype and phenotype in these Panicoid grasses. *Setaria viridis* has recently been adopted by the C_4 photosynthesis research community for the study of C_4 photosynthesis in a C_4 grass model species [235]. Diversity panels, mutagenesis populations, and recombinant inbred populations have all, recently, been produced [236] and we are currently using this screening approach to process thousands of *S. viridis* mutagenesis lines for cell wall variations (unpublished data). The rapid life cycle and small stature of *S. viridis* [235] makes it ideal for high-throughput phenotyping such as this, and its small genome of ~515 Mb [211] simplifies genetic mapping.

We hope that combining this high-throughput morphological and chemical phenotyping approach with recent advances in genome sequencing technologies, will rapidly improve yields of bioenergy crops such as sorghum, and using *S. viridis*

as a model species, will rapidly improve our understanding of the relationship between genotype and phenotype in C_4 monocot grasses.

6.2 A C_4 grass model system for the study of bioenergy traits

The lack of a model species with genetic and morphological similarities to bioenergy crops within the Panicoideae clade has hindered rapid progress in the understanding of bioenergy traits. *Setaria viridis* has recently been adopted by the C_4 photosynthesis research community as their C_4 grass model species [235]. It has also been shown to have similar lignocellulosic feedstock composition and cell wall saccharification dynamics to bioenergy crops sorghum, maize and switchgrass [237]. While the recent C_3 monocot grass model *Brachypodium distachyon* [238-240], and dicot *Arabidopsis* will complement our understanding of plant biology, *S. viridis* will allow a much more direct comparison with panicoid grass crop species and genetic engineering of bioenergy traits will be more transferable.

Within this dissertation we identified and characterised a developing *S. viridis* internode showing that it contained an active meristem, a zone of cell expansion that was synthesising primary cell walls, a transitional zone that was actively synthesising secondary cell walls, and a mature zone that was accumulating soluble sugars (**4.2 publication 4**). This experimental system, therefore, contained the biology that dictates:

a) Stem sink size in bioenergy crops via transcriptomic and metabolic control of an active intercalary meristem and cell expansion.
b) Cell wall composition via control of primary cell wall biosynthesis during cell expansion, and secondary cell wall biosynthesis (the major source of cell wall sugars in bioenergy crops) in the transitional zone of the internode
c) Soluble sugar content via control of the accumulation of soluble sugars in mature cells of the internode.

The focus of this dissertation was the utilisation of cell walls as a biofuel feedstock, therefore this discussion will mostly be limited to discussion of (b).

Transcriptomics of these tissues provided a platform for gene discovery, specifically with respect to these three processes. The resource provided here will, therefore, be an invaluable tool for selecting candidate genes and utilising *S. viridis* transformation [241, 242] for the genetic manipulation of stem sink size, cell wall composition, and soluble sugar content.

Towards this goal, a number of cell wall biosynthesis genes were identified from the transcriptomic dataset that was presented in **4.2**. Genes that were expressed highly in stem tissue and showed a transcriptional profile that was consistent with their role in either primary or secondary cell wall biosynthesis were selected and listed for future genetic manipulation, or *in situ* investigation of protein localisation.

Table 1. Candidate genes for genetic manipulation of primary cell walls in stem tissue of *Setaria viridis*. Genes affecting cellulose, hemicellulose and lignin are clustered. Gene IDs, common name, mean expression in each stem zone (MsZ-meristem, CEZ-cell expansion zone, TZ-transitional zone, MZ-mature zone) and a description of the gene is displayed.

Polymer	Gene ID	Name	MsZ	CEZ	TZ	MZ	Description
Cellulose	Si005721	CSLD	20	24	7	3	Cell plate/early cell wall cellulose
Hemicellulose	Si016912	CSLA	266	213	32	17	Mannan backbone synthesis
	Si016839	CSLA	39	21	4	2	
	Si029676	CSLA	42	46	10	10	
	Si021376	CSLC	62	94	42	14	Glucan backbone of hemicellulose
	Si009413	CSLH	13	20	1	0	Mixed linkage glucan
	Si006727	GT31	18	29	18	13	Primary cell wall
	Si006587	GT31	19	15	7	4	Early cell division cell wall?
	Si035653	GT34	40	66	16	11	Primary cell wall
	Si019396	GT37	26	6	5	4	Cell plate/early division cell wall?
Lignin	Si026523	COMT	79	4	0	0	Cell plate/early division cell wall? Vascular tissue in meristem?
	Si026536	COMT	41	2	0	0	

Cellulose synthesis is performed primarily by CesA proteins which form a rosette consisting of 6 CesAs, likely comprised of three different CesA proteins [36]. Interestingly, in primary cell walls, there was little *CesA* expression, however, there was high expression in secondary cell walls (**4.2 publication 4**). *CesA*s were highly expressed in the transitional and mature tissues of an *S. viridis* internode and three candidates were selected with the largest up-regulation in the transitional zone (**table 2**) where secondary cell wall synthesis was occurring at its highest rate (**4.2 publication 4**). Each showed similar expression profiles raising the possibility that

these three genes are co-expressed, forming a CesA rosette responsible for secondary cell wall synthesis of cellulose. COBRA proteins, whilst not directly responsible for cellulose synthesis, have been shown to form part of the cellulose synthesis complex as a "polysaccharide chaperone" which binds the emerging β1,4-glucan chain to facilitate cellulose crystallisation [40]. One *COBRA* gene was highly expressed in *S.* viridis stems and it showed a large spike in expression during the transitional phase of internode development, co-incident with the expression of *CesA* genes. Together, genetic manipulation of these genes has the potential to alter cellulose deposition in secondary cell walls and may lead to increased proportions of cellulose for improved bioethanol production. Whilst low, but significant, levels of CesA expression were observed in primary cell walls, some CesA-like gene families (annotated from the glycosyltransferase 2 (GT2) family) have also been shown to be capable of synthesising cellulose and may play a role in this process [36]. CSLD families have been shown to synthesise cellulose in dividing/expanding cells of maize coleoptiles, likely playing a role in formation of the cell plate and cell expansion [243]. In *S. viridis* stems there was only one *CSLD* gene that was highly expressed, and it was predominantly found in meristematic and cell expansion tissues (**table 1**). *Si005721* (*CSLD*) is therefore a strong candidate for manipulation of cellulose synthesis during cell division (i.e. cell plate formation) and cell expansion in early internode development.

Hemicellulose synthesis is more complex than cellulose synthesis and a number of large gene families, predominantly within the GT family, are thought to be responsible for their synthesis. The CSLA gene family has been shown to synthesise mannan backbones of hemicellulose polymers in dicots [36]. In *S. viridis* stems, only three *CSLA* genes were expressed highly in stems and these showed expression in meristematic and cell expansion tissues (**table 1**). Specifically, *Si016912* (*CSLA*) was highly expressed and is a candidate for genetic manipulation of the mannan component of primary cell walls. Little is known about the role of mannan backbones in cell walls of dividing and expanding cells in grasses, therefore, manipulation of this *CSLA* gene may provide novel insights into cell wall formation during early internode development. The CSLC gene family has been shown to

synthesise glucan backbones of xyloglucan hemicellulose polymers, but also may be capable of synthesising cellulose [36, 244]. In *S. viridis* stems, only one *CSLC* gene was highly expressed in stems, and this was in meristematic and cell expansion tissues (**table 1**). Si021376 (*CSLC*) may therefore be a candidate for xyloglucan backbone synthesis in grass primary cell walls. Whilst xyloglucan is a major component of type I cell walls, it is only present in the primary cell walls of grasses and little is known of its role during cell division and expansion, especially during internode development. Alternatively, CSLC may play a role in cellulose synthesis in primary cell walls of grasses. Genetic manipulation of this gene will provide insight into the role of CSLC in xyloglucan or cellulose synthesis in primary cell walls of grasses. CSLH has been shown to synthesise small amounts of mixed linkage glucans (MLG), however, playing a relatively minor role in comparison to *CSLF* genes [64]. In grasses, MLGs are found in high amounts (10-30 % cell wall dry weight) [58] in primary cell walls, and are absent in secondary cell walls, however, recently rice has been shown to contain MLG in secondary cell walls of its internode tissue [67]. In *S. viridis*, no *CSLF* genes were expressed in the primary cell wall tissue, however, *CSLH* was (**table 1**). This suggests that *Si009413* (*CSLH*) plays a major role in synthesis of primary cell wall MLGs in *S. viridis* internodes and is an interesting target for genetic manipulation of these MLGs, which have implications for cell expansion [57] and are also a useable product for bioethanol production, since MLG is a polymer of glucose.

These CSL gene families are annotated from the glycosyltransferase (GT) family 2, however, there are 97 GT families [245] of which many (including GT8, GT31, GT34, GT37 and GT47) have a suggested role in hemicellulose synthesis [246]. The GT families are generally large gene families that are not well characterised in terms of their role in cell wall synthesis, especially in grass species. From the large GT31, GT34 and GT37 gene families, surprisingly only a small number were highly expressed in stem tissue of *S. viridis*. Two *GT31* genes were identified that were highly expressed in meristematic and cell expansion tissues that contain primary cell walls, whilst one GT31 (*Si017391*) was identified that was expressed highly in the transitional and mature tissue (**table 1** and **2**) which are undergoing secondary

cell wall synthesis (**4.2 publication 4**). Genetic manipulation of these genes will provide insight into the role of the GT31 family in both primary and secondary cell wall synthesis in *S. viridis* internodes. From the GT34 and GT37 gene families, only one highly expressed gene was identified in each and both were expressed predominantly in meristematic and cell expansion tissues (**table 1**). Interestingly, *Si019396* (GT37) showed much higher expression in meristematic tissue than all other internode zones suggesting that it may have a more specific role in cell plate formation during cell division. Conversely, three *GT47* genes, which have been shown to be involved in xylan backbone synthesis, showed high expression in stems, specifically in the transitional and mature tissues where secondary cell wall synthesis was occurring (**table 2**). Since glucuronoarabinoxylans (GAX) are a major component of secondary cell walls of grasses (40-50 % cell wall dry weight [28]), these genes may play a major role in their synthesis. If these *GT47* genes do play a major role in GAX synthesis, genetic manipulation should result in a severe phenotype and will be of interest to the biofuel research community. Reductions in the amount of xylan in secondary cell walls of bioenergy crops has the potential to increase yields of bioethanol through indirect increases in the cellulose content which may also have implications for accessibility of cellulase enzymes to cellulose microfibrils embedded in the cell wall matrix of hemicelluloses and lignin.

Table 2. Candidate genes for genetic manipulation of secondary cell walls in stem tissue of *Setaria viridis*. Genes affecting cellulose, hemicellulose and lignin are clustered. Gene IDs, common name, mean expression in each stem zone (MsZ-meristem, CEZ-cell expansion zone, TZ-transitional zone, MZ-mature zone) and a description of the gene is displayed.

Polymer	Gene ID	Name	MsZ	CEZ	TZ	MZ	Description
Cellulose	Si000179	CesA	51	21	864	816	May interact to form a CesA rosette
	Si028761	CesA	27	12	613	579	
	Si034020	CesA	27	13	554	463	
	Si029839	COBRA	17	11	127	16	Cellulose synthesis - indirectly
Hemicellulose	Si013204	CSLF	36	110	265	270	Mixed linkage glucan synthesis
	Si001623	GT47	71	46	615	398	Xylan backbone synthesis
	Si022172	GT47	25	27	392	215	
	Si022174	GT47	26	31	189	122	
	Si017391	GT31	62	71	222	243	Galactomannan synthesis
Lignin	Si013348	PAL	68	122	1643	1371	Directs carbon to lignin/ flavanoid synthesis
	Si016504	PAL	60	98	1295	1019	
	Si016467	PAL	85	80	705	557	
	Si022114	C4H	31	57	273	197	
	Si006172	4CL	22	27	192	95	
	Si009939	HCT	23	44	236	367	Lignin biosynthesis
	Si021789	C3H	61	76	304	188	
	Si013961	CCR	41	51	248	330	
	Si014344	CCoAOMT	185	229	2280	864	
	Si014900	COMT	298	252	3861	2108	
	Si035174	F5H	11	7	90	12	
	Si017253	CAD	66 ±	42	414	364	

Lignin biosynthesis is well understood and is generally highly conserved between both type I and type II cell walls. Phenylalanine lyase (PAL) uses the amino acid, phenylalanine as a substrate which undergoes a series of enzyme catalysed steps to produce monolignols which are excreted into the cell wall space where polymerisation of the irregular polymer, lignin, occurs [94]. This process is a defining feature of secondary cell walls and is responsible for the cessation of cell expansion and the rigidity of mature tissues. All enzymes involved in the biosynthesis of lignin (i.e. *PAL*, *C4H*, *4CL*, *HCT*, *C3H*, *CCR*, *CCoAOMT*, *COMT*, *F5H* and *CAD*) showed high expression of at least one gene in transitional and mature tissues of the *S. viridis* internode (**table 2**). Three *PAL* genes, and one *C4H*, *4CL*, *HCT*, *C3H*, *CCR*, *CCoAOMT*, *COMT*, *F5H* and *CAD* were highly expressed in stem tissue, and especially in the transitional region of the *S. viridis* internode (**table 2**) where the

highest rates of lignin deposition occur (**4.2 publication 4**). Interestingly, two *COMT* genes were relatively highly expressed in the meristematic region of the internode (**table 1**) and are therefore candidates for lignification of developing protoxylem within this tissue. Disruption of *C4H*, *HCT*, *C3H*, *CCoAOMT*, *F5H*, *COMT* and *CAD* genes has been observed in alfalfa plants by RNAi knockdown [247-249], and *COMT* and *CAD* in mutagenesis populations of maize, perl millet, and sorghum producing Brown midrib (Bmr) mutants [201]. All of these gene disruptions caused major reductions in lignin content/composition and lignocellulosic hydrolysis rates [249], and therefore, genetic manipulation of these stem specific gene candidates identified here in *S. viridis* will likely have the potential to alter lignin content in stems of *S. viridis* without affecting other organs. It would also be of great interest to investigate the cell specificity of such manipulations using Raman confocal imaging of cell walls (**5.3 unpublished data**). Using this *S. viridis* model system, the optimal cell and organ specific gene disruptions can be rapidly determined and then applied to crop species for bioenergy crop improvement.

Table 3. Candidate genes for genetic manipulation of cell walls precursors, proteins, modifications and transcriptional regulation in stem tissue of *Setaria viridis*. Gene IDs, common name, mean expression in each stem zone (MsZ-meristem, CEZ-cell expansion zone, TZ-transitional zone, MZ-mature zone) and a description of the gene is displayed.

Process	Gene ID	Name	MsZ	CEZ	TZ	MZ	Description
CW precursors	Si036432	UXS	800	2246	3239	518	UDP-xylose cell wall precursor synthesis
	Si035448	UGD	474	1406	2254	481	Diverts UDPG to cell wall precursors
	Si013662	UGD	148	414	1002	81	
	Si010650	UGUT	80	114	735	305	UDPG uridyltransferase
CW protein	Si002509	AGP	135	152	1630	632	Arabinogalactan protein
	Si022977	AGP	250	865	1960	1861	
	Si037230	POE I	1	11	501	0	Type I allergen extensin
	Si036300	RGP	714	2020	833	319	1,4-glucan protein synthase
Cellulase	Si016976	CEL	23	9	0	0	Division CW cellulase (GH 9B1)
	Si034713	CEL	206	265	1043	1122	SCW cellulase (GH 9A1)
TFs	Si001565	MYB61	19	18	124	109	Resource allocation, lignin biosynthesis
	Si030750	MYB42	7	2	183	221	Lignin biosynthesis activation

While biosynthesis of the major cell wall components is very important for dictating cell wall composition, there are many other genes that also play a major

role. These roles include co-ordinating gene expression, producing cell wall precursors, structural proteins, and modification/re-organisation of already synthesised cell wall polymers. Some important, and highly expressed genes from each of these categories were also identified as candidates for genetic manipulation.

UDP-glucose is diverted toward the production of cell wall precursors, and ultimately cell walls, by UDP-glucose dehydrogenase (UGD) (**2.2 publication 1**). Two UGD genes were identified with very high expression in stems (**table 3**), specifically in the cell expansion and transitional regions where the majority of primary cell wall and secondary cell wall synthesis occur respectively (**4.2 publication 4**). These two genes, *Si013662* and *Si035448*, are therefore interesting candidates for the manipulation of carbon partitioning between cell walls, metabolism, and storage. A UPD-xylose synthase (UXS - precursor for xylan, the major 5 carbon sugar of grass cell walls [28]) and a UDP-glucose uridyltransferase (UGUT-which regulates UDP-glucose utilisation) with similar profiles were also identified as candidate genes. Genetic manipulation of these genes, especially the UXS has the potential to alter the 5 carbon/6 carbon composition of cell walls of *S. viridis* which has implications for bioethanol production since 6 carbon sugars (glucose) are more fermentable than 5 carbon sugars (xylose).

Structural proteins are also a very important component of cell walls and can affect its physical properties in many ways. Arabinogalactan glycoproteins (AGPs) have been shown to covalently link hemicellulosic and pectic polymers within the *Arabidopsis* type I cell wall [250], and are therefore thought to act as a structural component of the cell wall matrix. Two highly expressed AGP glycoproteins were identified in the transitional region of the *S. viridis* internode suggesting that they are a structural component of the secondary cell wall. A 1,4-glucan protein synthase (responsible for linking 1,4-glucans to protein), and a type I allergen extension protein (also called expansin B proteins that are thought to be involved in cell expansion in grasses [251]) showed high expression in the cell expansion and transitional regions of the *S. viridis* internode respectively. This is somewhat in

contrast to their predicted roles. Genetic manipulation will provide insights into the contribution of these highly expressed genes in cell expansion, an important process dictating (along with cell division) the sink size of bioenergy crops.

Deposition of cell walls also requires constant modification, especially in complex processes such as the formation of the cell plate during cell division, or during cell expansion. Modification and degradation of cellulose requires three enzymatic activities, cellobiohydrolase, endoglucanase, and β-glucosidase, [252]. The *in situ* role of these enzymes in plant development are not well understood, however they have been implicated in cell expansion and cellulose deposition [160]. Two cellulase genes were identified with high expression in *S. viridis* stems. One (*Si016976*) was a homologue of the endoglucanase AtCEL1 and was expressed uniquely in meristematic tissue. The other (*Si034713*) was a homologue of the endoglucanase AtKORRIGAN1, which has been shown to be required for normal cellulose synthesis [253], and was expressed in mature tissues during secondary cell wall deposition (**table 3**). *Si016976* (*SiCEL1*) is therefore a candidate for its involvement in the formation of cell plates during cell division, whilst *Si034713* (*SiKOR1*) is a candidate for re-organisation of secondary cell walls and production of crystalline cellulose. This has implications for biofuel production as utilisation of lignocellulosic material requires efficient enzymatic hydrolysis of cellulose. These candidate cellulase enzymes will likely be specialised for cellulose hydrolysis in *S. viridis*. Particularly *SiKOR1* would likely be specialised for the efficient hydrolysis of secondary cell walls within grass stems. Cloning and overproduction of these cellulase enzymes has the potential to improve the efficiency of enzymatic saccharification of lignocellulosic material from panicoid grasses for bioethanol production.

The co-ordination of transcription, especially during the transitional phase from meristematic/expanding tissue with primary cell walls, to mature tissues with lignified secondary cell walls, is the most important point of control determining how an internode develops. Whilst an extensive investigation of transcription factors required for cell wall synthesis was beyond the scope of this body of work,

two of the most highly expressed MYB transcription factors were identified as gene candidates for genetic manipulation. These two most highly expressed MYBs were, in fact, homologous to *MYB42* and *MYB61*, which have been shown in *Arabidopsis* to form part of the transcriptional cascade that is required for activation of secondary cell wall synthesis [146]. *MYB61* has been shown to affect lignin biosynthesis and resource allocation, and is expressed in xylem cells of sink tissues [254]. *MYB42* is expressed in fiber and xylem cells of stems and has been shown to repress the phenylpropanoid pathway in maize [255], but overexpression of its homologues cause ectopic deposition of lignin in *Arabidopsis* [256]. Both genes were highly expressed in transitional and mature tissues of the *S. viridis* internode where secondary cell wall synthesis occurs (**4.2 publication 4**).

All of the genes listed above are now targets for the manipulation of cell wall composition in *S. viridis* stems via gene knockdown or overexpression, utilising transformation of *S. viridis* (which is currently being performed in our laboratory). These genetically modified plants will be screened at the organ level for cell wall variations using an FTIR screening approach (**3.3 publication 2 & 3**), and on a cellular level using *in situ* chemometric Raman imaging of cell wall composition (**5.3 unpublished data**). It will also be of great interest to examine the protein localisation of these genes in 3D within the *S. viridis* experimental system using PEA-CLARITY (**5.2 publication 5**) and correlate that with cell specific compositional data obtained using Raman imaging of cell walls.

6.3 Concluding statement

Within this body of work we have developed a framework for the rapid improvement of bioenergy crops. A holistic high-throughput screening strategy for *S. bicolor*, that can be applied to other panicoid grass crops, will drive rapid phenotype-to-genotype discoveries in bioenergy crops. A powerful experimental system in the new C_4 model species *Setaria viridis* that shows metabolic and transcriptional changes through development of an internode will drive rapid gene-to-phenotype discoveries with respect to bioenergy traits. Coupled with efficient transformation [241, 242] and crossing [257] techniques (not explored in

this dissertation), two new imaging techniques, PEA-CLARITY and Raman confocal imaging of cell walls, will facilitate a detailed understanding of the effects of gene modification at a cellular, and sub-cellular level, within the *S. viridis* internode experimental system. This understanding will unravel further opportunities for bioenergy crop improvement.

Appendices

All additional files are available electronically on a USB supplied with this thesis and are labelled as they appear in text. All publications and their corresponding supplementary material, and all conference papers, posters and talks are supplied.

Bibliography

1. Byrt CS, Grof CPL, Furbank RT: **C4 Plants as Biofuel Feedstocks: Optimising Biomass Production and Feedstock Quality from a Lignocellulosic PerspectiveFree Access**. *Journal of Integrative Plant Biology* 2011, **53**(2):120-135.
2. Doust AN, Kellogg EA, Devos KM, Bennetzen JL: **Foxtail Millet: A Sequence-Driven Grass Model System**. *Plant Physiology* 2009, **149**(1):137-141.
3. Wang X, Gowik U, Tang H, Bowers J, Westhoff P, Paterson A: **Comparative genomic analysis of C4 photosynthetic pathway evolution in grasses**. *Genome Biology* 2009, **10**(6):R68.
4. Becker A, Winter K-U, Meyer B, Saedler H, Theißen G: **MADS-Box Gene Diversity in Seed Plants 300 Million Years Ago**. *Molecular Biology and Evolution* 2000, **17**(10):1425-1434.
5. Brutnell TP, LinWang, Swartwood K, Goldschmidt A, Jackson D, Zhu X-G, Kellogg E, Eck JV: **Setaria viridis: A Model for C4 Photosynthesis**. *The Plant Cell* 2010, **22**:2537-2544.
6. Friedl A, Padouvas E, Rotter H, Varmuza K: **Prediction of heating values of biomass fuel from elemental composition**. *Analytica Chimica Acta* 2005, **544**(1–2):191-198.

7. DL C: **GenChem Textbook**. In: *Molecules in living systems.* Edited by Moore JW, Holmes JL: ChemEd DL; 2013.
8. [http://www.docstoc.com/]
9. Sage RF: **The evolution of C4 photosynthesis**. *New Phytologist* 2004, **161**(2):341-370.
10. Taiz L, Zeiger E: **Plant Physiology**, 4th edn. Sunderland, Massachusetts: Sinauer Associates, Inc.; 2006.
11. Furbank RT, Taylor WC: **Regulation of Photosynthesis in C3 and C4 Plants: A Molecular Approach**. *The Plant Cell* 1995, **7**:797-807.
12. Gunning BES, Steer MW: **Plant Cell Biology: Structure and Function**. World Headquaters: Jones and Bartlett; 1996.
13. Cummings B: **Biology 8th Edition**, 8th edn: Cambell & Reece; 2008.
14. Clarke JMB, Tymoczko JL, Stryer L: **Biochemistry**, 5th edn. New York: W.H. Freeman; 2002.
15. Sage RF, Christin P-A, Edwards EJ: **The C4 plant lineages of planet Earth**. *Journal of Experimental Botany* 2011, **62**(9):3155-3169.
16. Brutnell TP, Wang L, Swartwood K, Goldschmidt A, Jackson D, Zhu X-G, Kellogg E, Van Eck J: **Setaria viridis: A Model for C4 Photosynthesis**. *The Plant Cell Online* 2010, **22**(8):2537-2544.
17. **Angiosperm Phylogeny Website** [http://www.mobot.org/MOBOT/research/APweb/]
18. Brown WH, Poon T: **Introduction to organic chemistry**, 3rd edn. New York: Wiley; 2005.
19. Hettinga WG, Junginger HM, Dekker SC, Hoogwijk M, McAloon AJ, Hicks KB: **Understanding the reductions in US corn ethanol production costs: An experience curve approach**. *Energy Policy* 2009, **37**(1):190-203.
20. Harrison RW: **The Food versus Fuel Debate: Implications for Consumers**. *Journal of Agricultural and Applied Economics* 2009, **41**(2):493-500.
21. Aoki N, Hirose T, Furbank R: **Sucrose Transport in Higher Plants: From Source to Sink**. In: *Photosynthesis.* Edited by Eaton-Rye JJ, Tripathy BC, Sharkey TD, vol. 34: Springer Netherlands; 2012: 703-729.
22. Kühn C, Grof CPL: **Sucrose transporters of higher plants**. *Current Opinion in Plant Biology* 2010, **13**(3):287-297.
23. Braun DM: **SWEET! The Pathway Is Complete**. *Science* 2012, **335**(6065):173-174.
24. Frank Baker R, Leach KA, Braun DM: **SWEET as Sugar: New Sucrose Effluxers in Plants**. *Molecular Plant* 2012, **5**(4):766-768.
25. Ayre BG: **Membrane-Transport Systems for Sucrose in Relation to Whole-Plant Carbon Partitioning**. *Molecular Plant* 2011, **4**(3):377-394.
26. Sticklen MB: **Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol**. *Nature Reviews Genetics* 2008, **9**:433-443.
27. Bar-Peled M, O'Neill MA: **Plant Nucleotide Sugar Formation, Interconversion, and Salvage by Sugar Recycling***. *Annual Review of Plant Biology* 2011, **62**(1):127-155.
28. Vogel J: **Unique aspects of the grass cell wall**. *Current Opinion in Plant Biology* 2008, **11**:301-307.

29. Taylor NG: **Cellulose biosynthesis and deposition in higher plants**. *New Phytologist* 2008, **178**:239-252.
30. Richmond T: **Higher plant cellulose synthases**. *Genome Biology* 2000, **1**(4):3001.3001-3001.3006.
31. Kimura S, Laosinchai W, Itoh T, Cui X, Linder R, Jr. RMB: **Immunogold Labeling of Rosette Terminal Cellulose-Synthesizing Complexes in the Vascular Plant Vigna angularis**. *The Plant Cell* 1999, **11**:2075-2085.
32. Emons AMC, Mulder BM: **How the deposition of cellulose microfibrils builds cell wall architecture**. *Trends in Plant Science* 2000, **5**(1):35-40.
33. Cosgrove DJ: **Growth of the Plant Cell Wall**. *Nature Reviews Molecular Cell Biology* 2005, **6**:850-861.
34. Hill JL, Hammudi MB, Tien M: **The Arabidopsis Cellulose Synthase Complex: A Proposed Hexamer of CESA Trimers in an Equimolar Stoichiometry**. *The Plant Cell* 2014, **26**(12):4834-4842.
35. Mutwil M, Debolt S, Persson S: **Cellulose synthesis: a complex complex**. *Current Opinion in Plant Biology* 2008, **11**(3):252-257.
36. Doblin MS, Pettolino F, Bacic A: **Plant cell walls: the skeleton of the plant world**. *Functional Plant Biology* 2010, **37**(5):357-381.
37. Tanaka K, Murata K, Yamazaki M, Onosato K, Miyao A, Hirochika H: **Three Distinct Rice Cellulose Synthase Catalytic Subunit Genes Required for Cellulose Synthesis in the Secondary Wall**. *Plant Physiology* 2003, **133**(1):73-83.
38. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A *et al*: **The Sorghum bicolor genome and the diversification of grasses**. *Nature* 2009, **457**(7229):551-556.
39. Appenzeller L, Doblin M, Barreiro R, Wang H, Niu X, Kollipara K, Carrigan L, Tomes D, Chapman M, Dhugga K: **Cellulose synthesis in maize: isolation and expression analysis of the cellulose synthase (CesA) gene family**. *Cellulose* 2004, **11**(3-4):287-299.
40. Sorek N, Sorek H, Kijac A, Szemenyei HJ, Bauer S, Hématy K, Wemmer DE, Somerville CR: **The Arabidopsis COBRA Protein Facilitates Cellulose Crystallization at the Plasma Membrane**. *Journal of Biological Chemistry* 2014, **289**(50):34911-34920.
41. Ben-Tov D, Abraham Y, Stav S, Thompson K, Loraine A, Elbaum R, De Souza A, Pauly M, Kieber JJ, Harpaz-Saad S: **COBRA-LIKE 2, a member of the GPI-anchored COBRA-LIKE family, plays a role in cellulose deposition in Arabidopsis seed coat mucilage secretory cells**. *Plant Physiology* 2015.
42. Somerville C: **Cellulose Synthesis in Higher Plants**. *Annual Review of Cell and Developmental Biology* 2006, **22**(1):53-78.
43. McFarlane HE, Döring A, Persson S: **The Cell Biology of Cellulose Synthesis**. *Annual Review of Plant Biology* 2014, **65**(1):69-94.
44. Arioli T, Peng L, Betzner AS, Burn J, Wittke W, Herth W, Camilleri C, Höfte H, Plazinski J, Birch R *et al*: **Molecular Analysis of Cellulose Biosynthesis in Arabidopsis**. *Science* 1998, **279**(5351):717-720.
45. Aohara T, Kotake T, Kaneko Y, Takatsuji H, Tsumuraya Y, Kawasaki S: **Rice BRITTLE CULM 5 (BRITTLE NODE) is Involved in Secondary Cell**

Wall Formation in the Sclerenchyma Tissue of Nodes. *Plant and Cell Physiology* 2009, **50**(11):1886-1897.
46. Sindhu A, Langewisch T, Olek A, Multani DS, McCann MC, Vermerris W, Carpita NC, Johal G: **Maize Brittle stalk2 Encodes a COBRA-Like Protein Expressed in Early Organ Development But Required for Tissue Flexibility at Maturity**. *Plant Physiology* 2007, **145**(4):1444-1459.
47. Kokubo A, Sakurai N, Kuraishi S, Takeda K: **Culm Brittleness of Barley (Hordeum vulgare L.) Mutants Is Caused by Smaller Number of Cellulose Molecules in Cell Wall**. *Plant Physiology* 1991, **97**(2):509-514.
48. Faik A: **Xylan Biosynthesis: News from the Grass**. *Plant Physiology* 2010, **153**(2):396-402.
49. York WS, O'Neill MA: **Biochemical control of xylan biosynthesis — which end is up?** *Current Opinion in Plant Biology* 2008, **11**(3):258-265.
50. Carpita NC, Gibeaut DM: **Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth**. *The Plant Journal* 1993, **3**(1):1-30.
51. Carpita NC, McCann MC: **Maize and sorghum: genetic resources for bioenergy grasses**. *Trends in Plant Science* 2008, **13**(8):415-420.
52. Carpita NC: **Hemicelluloses of the cell walls of Zea coleoptiles**. *Plant Physiology* 1983, **72**:515-521.
53. Mortimer JC, Miles GP, Brown DM, Zhang Z, Segura MP, Weimar T, Yu X, Seffen KA, Stephens E, Turner SR *et al*: **Absence of branches from xylan in Arabidopsis gux mutants reveals potential for simplification of lignocellulosic biomass**. *Proceedings of the National Academy of Sciences* 2010, **107**(40):17409-17414.
54. Doering A, Lathe R, Persson S: **An Update on Xylan Synthesis**. *Molecular Plant* 2012.
55. Pauly M, Gille S, Liu L, Mansoori N, de Souza A, Schultink A, Xiong G: **Hemicellulose biosynthesis**. *Planta* 2013, **238**(4):627-642.
56. Hahn-Hägerdal B, Karhumaa K, Fonseca C, Spencer-Martins I, Gorwa-Grauslund M: **Towards industrial pentose-fermenting yeast strains**. *Appl Microbiol Biotechnol* 2007, **74**(5):937-953.
57. Kim J-B, Olek AT, Carpita NC: **Cell Wall and Membrane-Associated Exo-b-D-Glucanases from Developing Maize Seedlings**. *Plant Physiology* 2000, **123**:471-485.
58. Kozlova LV, Snegireva AV, Gorshkova TA: **Distribution and structure of mixed linkage glucan at different stages of elongation of maize root cells**. *Russ J Plant Physiol* 2012, **59**(3):339-347.
59. Ermawar RA, Collins HM, Byrt CS, Betts NS, Henderson M, Shirley NJ, Schwerdt J, Lahnstein J, Fincher GB, Burton RA: **Distribution, structure and biosynthetic gene families of (1,3;1,4)-β-glucan in Sorghum bicolor**. *Journal of Integrative Plant Biology* 2015, **57**(4):429-445.
60. Carpita NC, Defernez M, Findlay K, Wells B, Shoue DA, Catchpole G, Wilson RH, McCann MC: **Cell Wall Architecture of the Elongating Maize Coleoptile**. *Plant Physiology* 2001, **127**:551-565.
61. Wilson SM, Ho YY, Lampugnani ER, Van de Meene AML, Bain MP, Bacic A, Doblin MS: **Determining the Subcellular Location of Synthesis and**

Assembly of the Cell Wall Polysaccharide (1,3; 1,4)-β-d-Glucan in Grasses. *The Plant Cell* 2015, **27**(3):754-771.
62. Lockhart J: **Uncovering the Unexpected Site of Biosynthesis of a Major Cell Wall Component in Grasses**. *The Plant Cell* 2015, **27**(3):483.
63. Burton RA, Wilson SM, Hrmova M, Harvey AJ, Shirley NJ, Medhurst A, Stone BA, Newbigin EJ, Bacic A, Fincher GB: **Cellulose Synthase-Like CslF Genes Mediate the Synthesis of Cell Wall (1,3;1,4)-ß-d-Glucans**. *Science* 2006, **311**(5769):1940-1942.
64. Doblin MS, Pettolino FA, Wilson SM, Campbell R, Burton RA, Fincher GB, Newbigin E, Bacic A: **A barley cellulose synthase-like CSLH gene mediates (1,3;1,4)-β-d-glucan synthesis in transgenic Arabidopsis**. *Proceedings of the National Academy of Sciences* 2009, **106**(14):5996-6001.
65. Kim S-J, Zemelis S, Keegstra K, Brandizzi F: **The cytoplasmic localization of the catalytic site of CSLF6 supports a channeling model for the biosynthesis of mixed-linkage glucan**. *The Plant Journal* 2015, **81**(4):537-547.
66. Vega-Sánchez ME, Verhertbruggen Y, Scheller HV, Ronald PC: **Abundance of mixed linkage glucan in mature tissues and secondary cell walls of grasses**. *Plant Signaling & Behavior* 2013, **8**(2):e23143.
67. Vega-Sánchez ME, Verhertbruggen Y, Christensen U, Chen X, Sharma V, Varanasi P, Jobling SA, Talbot M, White RG, Joo M *et al*: **Loss of Cellulose Synthase-Like F6 Function Affects Mixed-Linkage Glucan Deposition, Cell Wall Mechanical Properties, and Defense Responses in Vegetative Tissues of Rice**. *Plant Physiology* 2012, **159**(1):56-69.
68. Whetten R, Sederoff R: **Lignin Biosynthesis**. *The Plant Cell* 1995, **7**:1001-1013.
69. Hisano H, Nandakumar R, Wang Z-Y: **Genetic modification of lignin biosynthesis for improved biofuel production**. *In Vitro Cell Developmental Biology* 2009, **45**:306-313.
70. Li X, Weng J-K, Chapple C: **Improvement of biomass through lignin modification**. *The Plant Journal* 2008, **54**:569-581.
71. Koch G, Kleist G: **Application of Scanning UV Microspectrophotometry to Localise Lignins and Phenolic Extractives in Plant Cell Walls**. *Holzforschung* 2001, **55**(6):563-567.
72. Vanholme R, Morreel K, Ralph J, Boerjan W: **Lignin engineering**. *Current Opinion in Plant Biology* 2008, **11**:278-285.
73. Abreu HS, Latorraca JVF, Pereira RPW, Monteiro MBO, Abreu FA, Amparado KF: **A supramolecular proposal of lignin structure and its relation with the wood properties**. *Annals of the Brazilian Academy of Sciences* 2009, **81**(1):137-142.
74. Holton TA, Cornish EC: **Genetics and Biochemistry of Anthocyanin Biosynthesis**. *The Plant Cell* 1995, **7**:1071-1083.
75. Dixon RA, Paiva NL: **Stress-lnduced Phenylpropanoid Metabolism**. *The Plant Cell* 1995, **7**:1085-1097.
76. Ware D, Jaiswal P, Pan JX, Chang K, Clark K, Teytelman L, Schmidt S, Zhao W, Cartinhour S, McCouch S *et al*: **Gramene: a Resource for Comparative Grass Genomics**. *Nucleic Acids Research* 2002, **30**(1):103-105.

77. Rasmussen S, Dixon RA: **Transgene-Mediated and Elicitor-Induced Perturbation of Metabolic Channeling at the Entry Point into the Phenylpropanoid Pathway**. *The Plant Cell* 1999, **11**:1537-1551.
78. Liu C-J, Dixon RA: **Elicitor induced association of isoflavone O-methyltransferase with endomembranes prevents the formation and 7-O-methylation of daizein during isoflavonoid phytoalexin biosynthesis**. *The Plant Cell* 2001, **13**:2643-2658.
79. Winkel BSJ: **Metabolic Channeling In Plants**. *Annual Review of Plant Biology* 2004, **55**:85-107.
80. Hrazdina G, Wagner GJ: **Metabolic pathways as enzyme complexes: Evidence for the synthesis of phenylpropanoids and flavonoids on membrane-associatede enzyme complexes**. *Archive Of Biocemistry and Biophysics* 1985, **237**:88-100.
81. Hrazdina G, Jensen RA: **Spatial Organization Of Enzymes In Plant Metabolic Pathways**. *Annual Review of Plant Physiology and Plant Molecular Biology* 1992, **43**:241-267.
82. Sewalt VJ, Ni W, Blount JW, Jung HC, Masoud SA, Howles PA, Lamb C, Dixon RA: **Reduced Lignin Content and Altered Lignin Composition in Transgenic Tobacco Down-Regulated in Expression of L-Phenylalanine Ammonia-Lyase or Cinnamate 4-Hydroxylase**. *Plant Physiology* 1997, **115**:41-50.
83. Whetten RW, MacKay JJ, Sederoff RR: **Recent Advances In Understanding Lignin Biosynthesis**. *Annual Review of Plant Physiology and Molecular Biology* 1998, **49**:585-609.
84. Chiang VL: **Monolignol biosynthesis and genetic engineering of lignin in trees, a review**. *Environmental Chemistry Letters* 2006, **4**:143-146.
85. Ye Z-H, Kneusel RE, Matern U, Varner JE: **An Alternative Methylation Pathway in Lignin Biosynthesis in Zinnia**. *The Plant Cell* 1994, **6**:1427-1439.
86. Vogt T: **Phenylpropanoid Biosynthesis**. *Molecular Plant* 2010, **3**(1):2-20.
87. Grabber JH: **How Do Lignin Composition, Structure, and Cross-Linking Affect Degradability? A Review of Cell Wall Model Studies**. *Crop Science* 2005, **45**:820-831.
88. Samaj J, Hawkins S, Lauvergeat V, Grima-Pettenati J, Boudet A: **Immunolocalization of cinnamyl alcohol dehydrogenase 2 (CAD 2) indicates a good correlation with cell-speci®c activity of CAD 2 promoter in transgenic poplar shoots**. *Planta* 1998, **204**:437-443.
89. Smith CG, Rodgers MW, Zimmerlin A, Ferdinando D, Bolwell GP: **Tissue and subcellular immunolocalisation of enzymes of lignin synthesis in differentiating and wounded hypocotyl tissue of French bean (Phaseolus vulgaris L.)**. *Planta* 1994, **192**:155-164.
90. Kaneda M, Rensing KH, Wong JC, Banno B, Mansfield SD, Samuels AL: **Tracking Monolignols during Wood Development in Lodgepole Pine**. *Plant Physiology* 2008, **147**:1750-1760.
91. Ehlting J, Mattheus N, Aeschliman DS, Li E, Hamberger B, Cullis IF, Zhuang J, Kaneda M, Mansfield SD, Samuels L *et al*: **Global transcript profiling of primary stems from Arabidopsis thaliana identifies candidate genes**

for missing links in lignin biosynthesis and transcriptional regulators of fiber differentiation. *Plant Journal* 2005, **42**:618-640.
92. Wang Y, Chantreau M, Sibout R, Hawkins S: **Plant cell wall lignification and monolignol metabolism**. *Frontiers in Plant Science* 2013, **4**.
93. Ralph J, Brunow G, Harris PJ, Dixon RA, Schatz PF, Boerjan W: **Lignification: are Lignins Biosynthesized via simple Combinatorial Chemistry or via Proteinaceous Control and Template Replication?** In: *Recent Advances in Polyphenol Research.* Edited by Daayf F, Lattanzio V, vol. 1. Oxford: Wiley-Blackwell; 2008.
94. Boerjan W, Ralph J, Baucher M: **Lignin Biosynthesis**. *Annual Review of Plant Biology* 2003, **54**:519-546.
95. Karkonen A, Koutaniemi S: **Lignin Biosynthesis Studies in Plant Tissue Cultures**. *Journal of Integrative Plant Biology* 2010, **52**(2):176-185.
96. Shah K, Penel C, Gagnon J, Dunand C: **Purification and identification of a Ca2+-pectate binding peroxidase from Arabidopsis leaves**. *Pytochemistry* 2004, **65**:307-312.
97. Carpin S, Crevecoeur M, Greppin H, Penel C: **Molecular Cloning and Tissue-Specific Expression of an Anionic Peroxidase in Zucchini1**. *Plant Physiology* 1999, **120**:799-810.
98. Carpin S, Crèvecoeur M, Meyer Md, Simon P, Greppin H, Penel C: **Identification of a Ca2+ -Pectate Binding Site on an Apoplastic Peroxidase**. *The Plant Cell* 2001, **13**:511-520.
99. Hatfield R, Vermerris W: **Lignin Formation in Plants. The Dilemma of Linkage Specificity**. *Plant Physiology* 2001, **126**:1351-1357.
100. Liu C-J: **Deciphering the Enigma of Lignification: Precursor Transport, Oxidation, and the Topochemistry of Lignin Assembly**. *Molecular Plant* 2012, **5**(2):304-317.
101. Cesarino I, Araújo P, Sampaio Mayer JL, Vicentini R, Berthet S, Demedts B, Vanholme B, Boerjan W, Mazzafera P: **Expression of SofLAC, a new laccase in sugarcane, restores lignin content but not S:G ratio of Arabidopsis lac17 mutant**. *Journal of Experimental Botany* 2013, **64**(6):1769-1781.
102. Zhao Q, Nakashima J, Chen F, Yin Y, Fu C, Yun J, Shao H, Wang X, Wang Z-Y, Dixon RA: **LACCASE Is Necessary and Nonredundant with PEROXIDASE for Lignin Polymerization during Vascular Development in Arabidopsis**. *The Plant Cell* 2013, **25**(10):3976-3987.
103. Schuetz M, Benske A, Smith RA, Watanabe Y, Tobimatsu Y, Ralph J, Demura T, Ellis B, Samuels AL: **Laccases Direct Lignification in the Discrete Secondary Cell Wall Domains of Protoxylem**. *Plant Physiology* 2014, **166**(2):798-807.
104. Lee Y, Rubio Maria C, Alassimone J, Geldner N: **A Mechanism for Localized Lignin Deposition in the Endodermis**. *Cell* 2013, **153**(2):402-412.
105. Alejandro S, Lee Y, Tohge T, Sudre D, Osorio S, Park J, Bovet L, Lee Y, Geldner N, Fernie Alisdair R *et al*: **AtABCG29 Is a Monolignol Transporter Involved in Lignin Biosynthesis**. *Current Biology* 2012, **22**(13):1207-1212.

106. Hatfield RD, Ralph J, Grabber JH: **Cell wall cross-linking by ferulates and diferulates in grasses**. *Journal of the Science of Food and Agriculture* 1999, **79**:403-407.
107. Yu P, McKinnon JJ, Christensen DA: **Hydroxycinnamic acids and ferulic acid esterase in relation to biodegradation of complex plant cell walls**. *Canadian Journal of Animal Science* 2005, **85**(3):255-267.
108. Mohnen D: **Pectin structure and biosynthesis**. *Current Opinion in Plant Biology* 2008, **11**:266-277.
109. Zhu J, Chen S, Alvarez S, Asirvatham VS, Schachtman DP, Wu Y, Sharp RE: **Cell Wall Proteome in the Maize Primary Root Elongation Zone. I. Extraction and Identification of Water-Soluble and Lightly Ionically Bound Proteins**. *Plant Physiology* 2006, **140**(1):311-325.
110. Hunt JW, Dean AP, Webster RE, Johnson GN, Ennos AR: **A Novel Mechanism by which Silica Defends Grasses Against Herbivory**. *Annals of Botany* 2008, **102**:653-656.
111. Sparks J, Chandra S, Derry L, Parthasarathy M, Daugherty C, Griffin R: **Subcellular localization of silicon and germanium in grass root and leaf tissues by SIMS: evidence for differential and active transport**. *Biogeochemistry* 2011, **104**(1-3):237-249.
112. Zhang C, Wang L, Zhang W, Zhang F: **Do lignification and silicification of the cell wall precede silicon deposition in the silica cell of the rice (Oryza sativa L.) leaf epidermis?** *Plant Soil* 2013:1-13.
113. Chauhan DK, Tripathi DK, Rai NK, Rai AK: **Detection of Biogenic Silica in Leaf Blade, Leaf Sheath, and Stem of Bermuda Grass (Cynodon dactylon) Using LIBS and Phytolith Analysis**. *Food Biophysics* 2011, **6**(3):416-423.
114. Inanaga S, Okasaka A: **Calcium and silicon binding compounds in cell walls of rice shoots**. *Soil Science and Plant Nutrition* 1995, **41**(1):103-110.
115. Inanaga S, Okasaka A, Tanaka S: **Does silicon exist in association with organic compounds in rice plant?** *Soil Science and Plant Nutrition* 1995, **41**(1):111-117.
116. Hossain MT, Mori R, Soga K, Wakabayashi K, Kamisaka S, Fujii S, Yamamoto R, Hoson T: **Growth promotion and an increase in cell wall extensibility by silicon in rice and some other Poaceae seedlings**. *J Plant Res* 2002, **115**(1):0023-0027.
117. Hossain MT, Soga K, Wakabayashi K, Kamisaka S, Fujii S, Yamamoto R, Hoson T: **Modification of chemical properties of cell walls by silicon and its role in regulation of the cell wall extensibility in oat leaves**. *Journal of Plant Physiology* 2007, **164**(4):385-393.
118. Hattori T, Inanaga S, Tanimoto E, Lux A, Luxová M, Sugimoto Y: **Silicon-Induced Changes in Viscoelastic Properties of Sorghum Root Cell Walls**. *Plant and Cell Physiology* 2003, **44**(7):743-749.
119. Ma JF, Miyake Y, Takahashi E: **Chapter 2 Silicon as a beneficial element for crop plants**. In: *Studies in Plant Science.* Edited by L.E. Datnoff GHS, Korndörfer GH, vol. Volume 8: Elsevier; 2001: 17-39.
120. Currie HA, Perry CC: **Silica in Plants: Biological, Biochemical and Chemical Studies**. *Annals of Botany* 2007, **100**(7):1383-1389.

121. Epstein E: **Chapter 1 Silicon in plants: Facts vs. concepts**. In: *Studies in Plant Science.* Edited by L.E. Datnoff GHS, Korndörfer GH, vol. Volume 8: Elsevier; 2001: 1-15.
122. Ma JF, Yamaji N: **Silicon uptake and accumulation in higher plants**. *Trends in Plant Science* 2006, **11**(8):392-397.
123. Van Soest PJ, Jones LHP: **Effect of Silica in Forages upon Digestibility**. *Journal of Dairy Science* 1968, **51**(10):1644-1648.
124. Mitani N, Yamaji N, Ma JF: **Identification of Maize Silicon Influx Transporters**. *Plant and Cell Physiology* 2009, **50**(1):5-12.
125. Yamaji N, Chiba Y, Mitani-Ueno N, Feng Ma J: **Functional Characterization of a Silicon Transporter Gene Implicated in Silicon Distribution in Barley**. *Plant Physiology* 2012, **160**(3):1491-1497.
126. Montpetit J, Vivancos J, Mitani-Ueno N, Yamaji N, Rémus-Borel W, Belzile F, Ma J, Bélanger R: **Cloning, functional characterization and heterologous expression of TaLsi1, a wheat silicon transporter gene**. *Plant Molecular Biology* 2012, **79**(1-2):35-46.
127. Mitani N, Yamaji N, Ago Y, Iwasaki K, Ma JF: **Isolation and functional characterization of an influx silicon transporter in two pumpkin cultivars contrasting in silicon accumulation**. *The Plant Journal* 2011, **66**(2):231-240.
128. Ma JF, Tamai K, Yamaji N, Mitani N, Konishi S, Katsuhara M, Ishiguro M, Murata Y, Yano M: **A silicon transporter in rice**. *Nature* 2006, **440**(7084):688-691.
129. Ma JF, Yamaji N, Mitani N, Tamai K, Konishi S, Fujiwara T, Katsuhara M, Yano M: **An efflux transporter of silicon in rice**. *Nature* 2007, **448**(7150):209-212.
130. Yamaji N, Mitatni N, Ma JF: **A Transporter Regulating Silicon Distribution in Rice Shoots**. *The Plant Cell Online* 2008, **20**(5):1381-1389.
131. Mitani N, Chiba Y, Yamaji N, Ma JF: **Identification and Characterization of Maize and Barley Lsi2-Like Silicon Efflux Transporters Reveals a Distinct Silicon Uptake System from That in Rice**. *The Plant Cell Online* 2009, **21**(7):2133-2142.
132. Ubeda-Tomás S, Beemster GTS, Bennett MJ: **Hormonal regulation of root growth: integrating local activities into global behaviour**. *Trends in Plant Science* 2012, **17**(6):326-331.
133. Xie L, Yang C, Wang X: **Brassinosteroids can regulate cellulose biosynthesis by controlling the expression of CESA genes in Arabidopsis**. *Journal of Experimental Botany* 2011, **62**(13):4495-4506.
134. Todaka D, Nakashima K, Maruyama K, Kidokoro S, Osakabe Y, Ito Y, Matsukura S, Fujita Y, Yoshiwara K, Ohme-Takagi M *et al*: **Rice phytochrome-interacting factor-like protein OsPIL1 functions as a key regulator of internode elongation and induces a morphological response to drought stress**. *Proceedings of the National Academy of Sciences* 2012, **109**(39):15947-15952.
135. Benatti MR, Penning BW, Carpita NC, McCann MC: **We are good to grow: dynamic integration of cell wall architecture with the machinery of growth**. *Frontiers in Plant Science* 2012, **3**(187).

136. Muller B, Bourdais G, Reidy B, Bencivenni C, Massonneau As, Condamine P, Rolland Gl, Cone´je´ro Gv, Rogowsky P, Tardieu F: **Association of Specific Expansins with Growth in Maize Leaves Is Maintained under Environmental, Genetic, and Developmental Sources of Variation**. *Plant Physiology* 2007, **143**:278-290.
137. Cosgrove DJ, Bedinger P, Durachko DM: **Group I allergens of grass pollen as cell wall-loosening agents**. *PNAS* 1997, **94**:6559-6564.
138. Tabuchi A, Li L-C, Cosgrove DJ: **Matrix solubilization and cell wall weakening by β-expansin (group-1 allergen) from maize pollen**. *The Plant Journal* 2011, **68**(3):546-559.
139. Li L-C, Cosgrove DJ: **Grass group I pollen allergens (b-expansins) lack proteinase activity and do not cause wall loosening via proteolysis**. *European Journal of Biochemistry* 2001, **268**:4217-4226.
140. McCann MC, Carpita NC: **Designing the deconstruction of plant cell walls**. *Current Opinion in Plant Biology* 2008, **11**:314-320.
141. Boudet A-M: **Towards an understanding of the supramolecular organization of the lignified wall**. In: *The Plant Cell Wall.* Edited by Rose JKC, vol. 8. Oxford: Blackwell; 2003.
142. Zhong R, Lee C, Ye Z-H: **Evolutionary conservation of the transcriptional network regulating secondary cell wall biosynthesis**. *Trends in Plant Science* 2010, **15**(11):625-632.
143. Handakumbura PP, Hazen SP: **Transcriptional Regulation of Grass Secondary Cell Wall Biosynthesis: Playing Catch-Up with Arabidopsis thaliana**. *Frontiers in Plant Science* 2012, **74**(3).
144. Fornalé S, Shi X, Chai C, Encina A, Irar S, Capellades M, Fuguet E, Torres J-L, Rovira P, Puigdomènech P *et al*: **ZmMYB31 directly represses maize lignin genes and redirects the phenylpropanoid metabolic flux**. *The Plant Journal* 2010, **64**(4):633-644.
145. Shen H, He X, Poovaiah CR, Wuddineh WA, Ma J, Mann DGJ, Wang H, Jackson L, Tang Y, Neal Stewart C *et al*: **Functional characterization of the switchgrass (Panicum virgatum) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks**. *New Phytologist* 2012, **193**(1):121-136.
146. Zhong R, Lee C, McCarthy RL, Reeves CK, Jones EG, Ye Z-H: **Transcriptional Activation of Secondary Wall Biosynthesis by Rice and Maize NAC and MYB Transcription Factors**. *Plant and Cell Physiology* 2011, **52**(10):1856-1871.
147. Fornalé S, Sonbol F-M, Maes T, Capellades M, Puigdomènech P, Rigau J, Caparrós-Ruiz D: **Down-regulation of the maize and Arabidopsis thaliana caffeic acid O-methyl-transferase genes by two new maize R2R3-MYB transcription factors**. *Plant Molecular Biology* 2006, **62**(6):809-823.
148. Terashima N, Kitano K, Kojima M, Yoshida M, Yamamoto H, Westermark U: **Nanostructural assembly of cellulose, hemicellulose, and lignin in the middle layer of secondary wall of ginkgo tracheid**. *Journal of Wood Science* 2009, **55**:409-416.
149. Barrett KE, Boitano S, Brooks HL: **Review of medical physiology**: McGraw-Hill Lange; 2012.

150. Tomotani EJ, Vitolo M: **Production of high-fructose syrup using immobilized invertase in a membrane reactor**. *Journal of Food Engineering* 2007, **80**(2):662-667.
151. Willey JM, Sherwood LM, Woolverton CJ: **Prescott, Harley, and Klein's microbiology**, 7th edn. New York: McGraw-Hill; 2008.
152. van der Maarel MJEC, van der Veen B, Uitdehaag JCM, Leemhuis H, Dijkhuizen L: **Properties and applications of starch-converting enzymes of the α-amylase family**. *Journal of Biotechnology* 2002, **94**(2):137-155.
153. Gilbert HJ, Stålbrand H, Brumer H: **How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation**. *Current Opinion in Plant Biology* 2008, **11**(3):338-348.
154. McDougall EI: **Studies on ruminant saliva. 1. The composition and output of sheep's saliva**. *Biochemistry journal* 1948, **43**(1):99-109.
155. Mosier N, Wyman C, Dale B, Elander R, Lee YY, Holtzapple M, Ladisch M: **Features of promising technologies for pretreatment of lignocellulosic biomass**. *Bioresource Technology* 2005, **96**(6):673-686.
156. Hendriks ATWM, Zeeman G: **Pretreatments to enhance the digestibility of lignocellulosic biomass**. *Bioresource Technology* 2009, **100**(1):10-18.
157. Vidal B, Jr., Dien B, Ting KC, Singh V: **Influence of Feedstock Particle Size on Lignocellulose Conversion—A Review**. *Appl Biochem Biotechnol* 2011, **164**(8):1405-1421.
158. Hofmann RR: **Evolutionary Steps of Ecophysiological Adaptation and Diversification of Ruminants: A Comparative View of Their Digestive System**. *Oecologia* 1989, **78**(4):443-457.
159. Jordan DB, Bowman MJ, Braker JD, Dien BS, Hector RE, Lee CC, Mertens JA, Wagschal K: **Plant cell walls to ethanol**. *Biochemistry journal* 2012, **442**:241-252.
160. Lopez-Casado G, Urbanowicz BR, Damasceno CMB, Rose JKC: **Plant glycosyl hydrolases and biofuels: a natural marriage**. *Current Opinion in Plant Biology* 2008, **11**(3):329-337.
161. Bugg TDH, Ahmad M, Hardiman EM, Rahmanpour R: **Pathways for degradation of lignin in bacteria and fungi**. *Natural Product Reports* 2011, **28**(12):1883-1896.
162. Hammel KE, Cullen D: **Role of fungal peroxidases in biological ligninolysis**. *Current Opinion in Plant Biology* 2008, **11**(3):349-355.
163. Zumdahl SS: **Introductory chemistry: A foundation**, 5th edn. Boston New York: Houghton Miffin Company; 2002.
164. Dien SB, Bothast RJ, Nichols NN, Cotta MA: **The U.S. corn ethanol industry: an overview of current technology and future prospects**. *International Sugar Journal* 2002, **104**(1241):204-208.
165. **Ethanol technologies** [http://www.ethtec.com.au/]
166. Stavrinides AJ, Phipps DA, Al-Shamma'a A: **Review; Current And Developing Ligno-Cellulosic Pretreatment Methods For Bioethanol Production** In: *Liverpool Conference on the Built Environment and Natural Environment: 2010; Liverpool*. Liverpool BEAN: 233-242.
167. Guo B, Zhang Y, Yu G, Lee W-H, Jin Y-S, Morgenroth E: **Two-Stage Acidic–Alkaline Hydrothermal Pretreatment of Lignocellulose for the High**

Recovery of Cellulose and Hemicellulose Sugars. *Appl Biochem Biotechnol* 2013, **169**(4):1069-1087.
168. Grange D, Haan R, Zyl W: **Engineering cellulolytic ability into bioprocessing organisms**. *Appl Microbiol Biotechnol* 2010, **87**(4):1195-1208.
169. Wilson DB: **Cellulases and biofuels**. *Current Opinion in Biotechnology* 2009, **20**(3):295-299.
170. Beckham GT, Dai Z, Matthews JF, Momany M, Payne CM, Adney WS, Baker SE, Himmel ME: **Harnessing glycosylation to improve cellulase activity**. *Current Opinion in Biotechnology* 2012, **23**(3):338-345.
171. Georgianna DR, Mayfield SP: **Exploiting diversity and synthetic biology for the production of algal biofuels**. *Nature* 2012, **488**(7411):329-335.
172. Rosenberg JN, Mathias A, Korth K, Betenbaugh MJ, Oyler GA: **Microalgal biomass production and carbon dioxide sequestration from an integrated ethanol biorefinery in Iowa: A technical appraisal and economic feasibility evaluation**. *Biomass and Bioenergy* 2011, **35**(9):3865-3876.
173. Atabani AE, Silitonga AS, Badruddin IA, Mahlia TMI, Masjuki HH, Mekhilef S: **A comprehensive review on biodiesel as an alternative energy resource and its characteristics**. *Renewable and Sustainable Energy Reviews* 2012, **16**(4):2070-2093.
174. Lam MK, Lee KT: **Microalgae biofuels: A critical review of issues, problems and the way forward**. *Biotechnology Advances* 2012, **30**(3):673-690.
175. Halim R, Danquah MK, Webley PA: **Extraction of oil from microalgae for biodiesel production: A review**. *Biotechnology Advances* 2012, **30**(3):709-732.
176. **Algenol DIRECT TO ETHANOL** [http://www.algenolbiofuels.com/]
177. Peteavich R: **Method of producing hydrocarbon biofuels using genetically modified seaweed**. In. Edited by office Up, vol. US 2010/0120111 A1. US; 2009.
178. Pauly M, Keegstra K: **Physiology and metabolism 'Tear down this wall'**. *Current Opinion in Plant Biology* 2008, **11**(3):233-235.
179. Hoffmann-Thoma G, Hinkel K, Nicolay P, Willenbrink J: **Sucrose accumulation in sweet sorghum stem internodes in relation to growth**. *Physiologia Plantarum* 1996, **97**(2):277-284.
180. Rae AL, Grof CPL, Casu RE, Bonnett GD: **Sucrose accumulation in the sugarcane stem: pathways and control points for transport and compartmentation**. *Field Crops Research* 2005, **92**(2–3):159-168.
181. Vandenbrink JP, Hammonds RE, Hilten RN, Das KC, Henson JM, Paterson AH, Feltus FA: **Tissue specific analysis of bioconversion traits in the bioenergy grass Sorghum bicolor**. *Industrial Crops and Products* 2013, **50**(0):118-130.
182. Kimber CT: **Sorghum: Origin, history, technology, and production** John Wiley and Sons; 2000.
183. Vandenbrink JP, Delgado MP, Frederick JR, Feltus FA: **A sorghum diversity panel biofuel feedstock screen for genotypes with high hydrolysis yield potential**. *Industrial Crops and Products* 2010, **31**(3):444-448.

184. Sluiter A, Hames B, Ruiz R, Scarlata C, Sluiter J, Templeton D, Crocker D: **Determination of Structural Carbohydrates and Lignin in Biomass**. *NREL technical report* 2011.
185. Stefaniak TR, Dahlberg JA, Bean BW, Dighe N, Wolfrum EJ, Rooney WL: **Variation in Biomass Composition Components among Forage, Biomass, Sorghum-Sudangrass, and Sweet Sorghum Types**. *Crop Sci* 2012, **52**(4):1949-1954.
186. Mace ES, Jordan DR: **Integrating sorghum whole genome sequence information with a compendium of sorghum QTL studies reveals uneven distribution of QTL and of gene-rich regions with significant implications for crop improvement**. *Theoretical and Applied Genetics* 2011, **123**(1):169-191.
187. Wang Y-H, Bible P, Loganantharaj R, Upadhyaya H: **Identification of SSR markers associated with height using pool-based genome-wide association mapping in sorghum**. *Molecular Breeding* 2012, **30**(1):281-292.
188. Guan Y-a, Wang H-l, Qin L, Zhang H-w, Yang Y-b, Gao F-j, Li R-y, Wang H-g: **QTL mapping of bio-energy related traits in Sorghum**. *Euphytica* 2011, **182**(3):431-440.
189. Zou G, Zhai G, Feng Q, Yan S, Wang A, Zhao Q, Shao J, Zhang Z, Zou J, Han B *et al*: **Identification of QTLs for eight agronomically important traits using an ultra-high-density map based on SNPs generated from high-throughput sequencing in sorghum under contrasting photoperiods**. *Journal of Experimental Botany* 2012, **63**(15):5451-5462.
190. Shiringani A, Friedt W: **QTL for fibre-related traits in grain × sweet sorghum as a tool for the enhancement of sorghum as a biomass crop**. *Theoretical and Applied Genetics* 2011, **123**(6):999-1011.
191. Vandenbrink J, Goff V, Jin H, Kong W, Paterson A, Alex Feltus F: **Identification of bioconversion quantitative trait loci in the interspecific cross Sorghum bicolor × Sorghum propinquum**. *Theoretical and Applied Genetics* 2013:1-14.
192. Kong W, Jin H, Franks CD, Kim C, Bandopadhyay R, Rana MK, Auckland SA, Goff VH, Rainville LK, Burow GB *et al*: **Genetic Analysis of Recombinant Inbred Lines for Sorghum bicolor × Sorghum propinquum**. *G3: Genes|Genomes|Genetics* 2013, **3**(1):101-108.
193. Salas Fernandez MG, Becraft PW, Yin Y, Lübberstedt T: **From dwarves to giants? Plant height manipulation for biomass yield**. *Trends in Plant Science* 2009, **14**(8):454-461.
194. Pereira MG, Lee M: **Identification of genomic regions affecting plant height in sorghum and maize**. *Theoretical and Applied Genetics* 1995, **90**(3-4):380-388.
195. Multani DS, Briggs SP, Chamberlin MA, Blakeslee JJ, Murphy AS, Johal GS: **Loss of an MDR Transporter in Compact Stalks of Maize br2 and Sorghum dw3 Mutants**. *Science* 2003, **302**(5642):81-84.
196. Parry MAJ, Madgwick PJ, Bayon C, Tearall K, Hernandez-Lopez A, Baudo M, Rakszegi M, Hamada W, Al-Yassin A, Ouabbou H *et al*: **Mutation discovery for crop improvement**. *Journal of Experimental Botany* 2009, **60**(10):2817-2825.

197. Ahloowalia BS, Maluszynski M: **Induced mutations – A new paradigm in plant breeding**. *Euphytica* 2001, **118**(2):167-173.
198. Porter KS, Axtell JD, Lechtenberg VL, Colenbrander VF: **Phenotype, Fiber Composition, and in vitro Dry Matter Disappearance of Chemically Induced Brown Midrib (bmr) Mutants of Sorghum**. *Crop Science* 1977, **18**(2):205-208.
199. Pedersen JF, Vogel KP, Funnell DL: **Impact of Reduced Lignin on Plant Fitness**. *Crop Science* 2005, **45**:812-819.
200. Dien B, Sarath G, Pedersen J, Sattler S, Chen H, Funnell-Harris D, Nichols N, Cotta M: **Improved sugar conversion and ethanol yield for forage sorghum (Sorghum bicolor L. Moench) lines with reduced lignin contents**. *BioEnergy Res* 2009, **2**:153 - 164.
201. Sattler SE, Funnell-Harris DL, Pedersen JF: **Brown midrib mutations and their importance to the utilization of maize, sorghum, and pearl millet lignocellulosic tissues**. *Plant Science* 2010, **178**:229-238.
202. Xin Z, Li Wang M, Barkley N, Burow G, Franks C, Pederson G, Burke J: **Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population**. *BMC Plant Biology* 2008, **8**(1):103.
203. Xin Z, Wang M-L, Chopra S, Wang P: **Gene Mutagenesis Systems and Resources for the Saccharinae**. In: *Genomics of the Saccharinae.* Edited by Paterson AH, vol. 11: Springer New York; 2013: 169-185.
204. Xin Z, Wang M, Burow G, Burke J: **An Induced Sorghum Mutant Population Suitable for Bioenergy Research**. *BioEnergy Research* 2009, **2**(1-2):10-16.
205. Jones T: **Maize Tissue Culture and Transformation: The First 20 Years**. In: *Molecular Genetic Approaches to Maize Improvement.* Edited by Kriz A, Larkins B, vol. 63: Springer Berlin Heidelberg; 2009: 7-27.
206. Li R, Qu R: **High throughput Agrobacterium-mediated switchgrass transformation**. *Biomass and Bioenergy* 2011, **35**(3):1046-1054.
207. Arruda P: **Genetically modified sugarcane for bioenergy generation**. *Current Opinion in Biotechnology* 2012, **23**(3):315-322.
208. Liu G, Godwin I: **Highly efficient sorghum transformation**. *Plant Cell Rep* 2012, **31**(6):999-1007.
209. Alves SC, Worland B, Thole V, Snape JW, Bevan MW, Vain P: **A protocol for Agrobacterium-mediated transformation of Brachypodium distachyon community standard line Bd21**. *Nat Protocols* 2009, **4**(5):638-649.
210. Brkljacic J, Grotewold E, Scholl R, Mockler T, Garvin DF, Vain P, Brutnell T, Sibout R, Bevan M, Budak H *et al*: **Brachypodium as a Model for the Grasses: Today and the Future**. *Plant Physiology* 2011, **157**(1):3-13.
211. Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, Estep M, Feng L, Vaughn JN, Grimwood J *et al*: **Reference genome sequence of the model plant Setaria**. *Nat Biotech* 2012, **30**(6):555-561.
212. Li P, Brutnell TP: **Setaria viridis and Setaria italica, model genetic systems for the Panicoid grasses**. *Journal of Experimental Botany* 2011, **62**(9):3031-3037.

213. Ambavaram MMR, Krishnan A, Trijatmiko KR, Pereira A: **Coordinated Activation of Cellulose and Repression of Lignin Biosynthesis Pathways in Rice**. *Plant Physiology* 2011, **155**(2):916-931.
214. Coleman HD, Yan J, Mansfield SD: **Sucrose synthase affects carbon partitioning to increase cellulose production and altered cell wall ultrastructure**. *Proceedings of the National Academy of Sciences* 2009, **106**(31):13118-13123.
215. Morikawa H, Senda M: **Infrared analysis of oat coleoptile cell walls and oriented structure of matrix polysaccharides in the walls**. *Plant and Cell Physiology* 1978, **19**(2):327-336.
216. Brown DM, Zeef LAH, Ellis J, Goodacre R, Turner SR: **Identification of Novel Genes in Arabidopsis Involved in Secondary Cell Wall Formation Using Expression Profiling and Reverse Genetics**. *The Plant Cell* 2005, **17**:2281-2295.
217. Mouille G, Robin S, Lecomte M, Pagant S, Hofte H: **Classification and Identification of *Arabidopsis* cell wall mutations using Fourier-Transform InfraRed (FT-IR) microscopy**. *The Plant Journal* 2003, **35**:393-404.
218. McCann MC, Defernez M, Urbanowicz BR, Tewari JC, Langewisch T, Olek A, Wells B, Wilson RH, Carpita NC: **Neural Network Analyses of Infrared Spectra for Classifying Cell Wall Architectures1**. *Plant Physiology* 2007, **143**:1314-1326.
219. Sills DL, Gossett JM: **Using FTIR to Predict Saccharification From Enzymatic Hydrolysis of Alkali-Pretreated Biomasses**. *Biotechnology and Bioengineering* 2012, **109**(2).
220. Duarte IF, Barros A, Delgadillo I, Almeida C, Gil AM: **Application of FTIR spectroscopy for the quantification of sugars in mango juice as a function of ripening**. *Journal of agricultural and food chemistry* 2002, **50**(11):3104-3111.
221. Irudayaraj J, Tewari J: **Simultaneous Monitoring of Organic Acids and Sugars in Fresh and Processed Apple Juice by Fourier Transform Infrared?Attenuated Total Reflection Spectroscopy**. *Appl Spectrosc* 2003, **57**(12):1599-1604.
222. Wang J, Kliks MM, Jun S, Jackson M, Li QX: **Rapid Analysis of Glucose, Fructose, Sucrose, and Maltose in Honeys from Different Geographic Regions using Fourier Transform Infrared Spectroscopy and Multivariate Analysis**. *Journal of Food Science* 2010, **75**(2):C208-C214.
223. ATALLA RH, AGARWAL UP: **Raman Microprobe Evidence for Lignin Orientation in the Cell Walls of Native Woody Tissue**. *Science* 1985, **227**(4687):636-638.
224. Richter S, Müssig J, Gierlinger N: **Functional plant cell wall design revealed by the Raman imaging approach**. *Planta* 2011, **233**(4):763-772.
225. Sene C, McCann MC, Wilson RH, Grinter R: **Fourier-Transform Raman and Fourier-Transform Infrared Spectroscopy (An Investigation of Five Higher Plant Cell Walls and Their Components)**. *Plant Physiology* 1994, **106**(4):1623-1631.

226. Gierlinger N, Schwanninger M: **Chemical Imaging of Poplar Wood Cell Walls by Confocal Raman Microscopy**. *Plant Physiology* 2006, **140**(4):1246-1254.
227. Gierlinger N, Keplinger T, Harrington M: **Imaging of plant cell walls by confocal Raman microscopy**. *Nat Protocols* 2012, **7**(9):1694-1708.
228. Furbank RT, Tester M: **Phenomics – technologies to relieve the phenotyping bottleneck**. *Trends in Plant Science* 2011, **16**(12):635-644.
229. McCouch S, Baute GJ, Bradeen J, Bramel P, Bretting PK, Buckler E, Burke JM, Charest D, Cloutier S, Cole G *et al*: **Agriculture: Feeding the future**. *Nature* 2013, **499**(7456):23-24.
230. Martin AP, Palmer WM, Byrt CS, Furbank RT, Grof CP: **A holistic high-throughput screening framework for biofuel feedstock assessment that characterises variations in soluble sugars and cell wall composition in Sorghum bicolor**. *Biotechnology for Biofuels* 2013, **6**(1):1-13.
231. Wu L, Li M, Huang J, Zhang H, Zou W, Hu S, Li Y, Fan C, Zhang R, Jing H *et al*: **A near infrared spectroscopic assay for stalk soluble sugars, bagasse enzymatic saccharification and wall polymers in sweet sorghum**. *Bioresource Technology* 2015, **177**(0):118-124.
232. Payne C, Wolfrum E: **Rapid analysis of composition and reactivity in cellulosic biomass feedstocks with near-infrared spectroscopy**. *Biotechnology for Biofuels* 2015, **8**(1):43.
233. Bekele WA, Wieckhorst S, Friedt W, Snowdon RJ: **High-throughput genomics in sorghum: from whole-genome resequencing to a SNP screening array**. *Plant Biotechnology Journal* 2013, **11**(9):1112-1125.
234. Higgins RH, Thurber CS, Assaranurak I, Brown PJ: **Multiparental Mapping of Plant Height and Flowering Time QTL in Partially Isogenic Sorghum Families**. *G3: Genes|Genomes|Genetics* 2014, **4**(9):1593-1602.
235. Brutnell TP, Wang L, Swartwood K, Goldschmidt A, Jackson D, Zhu X-G, Kellogg E, Van Eck J: **Setaria viridis: A Model for C4 Photosynthesis**. *The Plant Cell* 2010, **22**(8):2537-2544.
236. Brutnell TP, Bennetzen JL, Vogel JP: **Brachypodium distachyon and Setaria viridis: Model Genetic Systems for the Grasses**. *Annual Review of Plant Biology* 2015, **66**(1):465-485.
237. Petti C, Shearer A, Tateno M, Ruwaya M, Nokes S, Brutnell T, DeBolt S: **Comparative feedstock analysis in Setaria viridis L. as a model for C4 bioenergy grasses and Panicoid crop species**. *Frontiers in Plant Science* 2013, **4**:181.
238. Vogel JP, Garvin DF, Mockler TC, Schmutz J, Rokhsar D, Bevan MW: **Genome sequencing and analysis of the model grass Brachypodium distachyon**. *Nature* 2010, **463**(7282):763-768.
239. Vogel J, Bragg J: **Brachypodium distachyon, a New Model for the Triticeae**. *Plant Genetics and Genomics: Crops and Models* 2009, **7**(2):427-449.
240. Garvin DF, Gu Y-Q, Hasterok R, Hazen SP, Jenkins G, Mockler TC, Mur LAJ, Vogel JP: **Development of Genetic and Genomic Research Resources for Brachypodium distachyon, a New Model System for Grass Crop Research**. *The Plant Genome* 2008, **1**:69-84.

241. Martins PK, Nakayama TJ, Ribeiro AP, Cunha BADBd, Nepomuceno AL, Harmon FG, Kobayashi AK, Molinari HBC: **Setaria viridis floral-dip: A simple and rapid Agrobacterium-mediated transformation method**. *Biotechnology Reports* 2015, **6**(0):61-63.
242. Van Eck J, Swartwood K: **Setaria viridis**. In: *Agrobacterium Protocols.* Edited by Wang K, vol. 1223: Springer New York; 2015: 57-67.
243. Hunter CT, Kirienko DH, Sylvester AW, Peter GF, McCarty DR, Koch KE: **Cellulose Synthase-Like D1 Is Integral to Normal Cell Division, Expansion, and Leaf Development in Maize**. *Plant Physiology* 2012, **158**(2):708-724.
244. Dwivany FM, Yulia D, Burton RA, Shirley NJ, Wilson SM, Fincher GB, Bacic A, Newbigin E, Doblin MS: **The CELLULOSE-SYNTHASE LIKE C (CSLC) Family of Barley Includes Members that Are Integral Membrane Proteins Targeted to the Plasma Membrane**. *Molecular Plant* 2009, **2**(5):1025-1039.
245. **CAZY database** [http://www.cazy.org/]
246. **Cell wall genomics** [https://cellwall.genomics.purdue.edu/]
247. Shadle G, Chen F, Reddy MSS, Jackson L, Nakashima J, Dixon RA: **Down-regulation of hydroxycinnamoyl CoA: Shikimate hydroxycinnamoyl transferase in transgenic alfalfa affects lignification, development and forage quality**. *Phytochemistry* 2007, **68**:1521-1529.
248. Guo D, Chen F, Wheeler J, Winder J, Selman S, Peterson M, Dixon RA: **Improvement of in-rumen digestibility of alfalfa forage by genetic manipulation of lignin O-methyltransferases** *Transgenic Research* 2001, **10**(5):457-464.
249. Chen F, Dixon RA: **Lignin modification improves fermentable sugar yields for biofuel production**. *Nature Biotechnology* 2007, **25**(7):759-761.
250. Tan L, Eberhard S, Pattathil S, Warder C, Glushka J, Yuan C, Hao Z, Zhu X, Avci U, Miller JS *et al*: **An Arabidopsis Cell Wall Proteoglycan Consists of Pectin and Arabinoxylan Covalently Linked to an Arabinogalactan Protein**. *The Plant Cell* 2013, **25**(1):270-287.
251. Sampedro J, Guttman M, Li L-C, Cosgrove DJ: **Evolutionary divergence of β–expansin structure and function in grasses parallels emergence of distinctive primary cell wall traits**. *The Plant Journal* 2015, **81**(1):108-120.
252. Van Dyk JS, Pletschke BI: **A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes—Factors affecting enzymes, conversion and synergy**. *Biotechnology Advances* 2012, **30**(6):1458-1480.
253. Ueda T: **Cellulase in Cellulose Synthase: A Cat among the Pigeons?** *Plant Physiology* 2014, **165**(4):1397-1398.
254. Zhao K, Bartley LE: **Comparative genomic analysis of the R2R3 MYB secondary cell wall regulators of Arabidopsis, poplar, rice, maize, and switchgrass**. *BMC Plant Biology* 2014, **14**(1):135.
255. Sonbol F-M, Fornalé S, Capellades M, Encina A, Touriño S, Torres J-L, Rovira P, Ruel K, Puigdomènech P, Rigau J *et al*: **The maize ZmMYB42 represses the phenylpropanoid pathway and affects the cell wall**

structure, composition and degradability in Arabidopsis thaliana. *Plant Molecular Biology* 2009, **70**(3):283-296.
256. Zhong R, Lee C, Zhou J, McCarthy RL, Ye Z-H: **A Battery of Transcription Factors Involved in the Regulation of Secondary Cell Wall Biosynthesis in Arabidopsis**. *The Plant Cell* 2008, **20**:2763-2782.
257. Jiang H, Barbier H, Brutnell T: **Methods for Performing Crosses in Setaria viridis, a New Model System for the Grasses**. *Journal of visualized experiments : JoVE* 2013(80):50527.