Title
Reducing Xylan and Improving Lignocellulosic Biomass through Antimorphic and Heterologous Enzyme Expression

Author
Brandon, Andrew G

Publication Date
2019

Peer reviewed|Thesis/dissertation

Reducing Xylan and Improving Lignocellulosic Biomass through Antimorphic and Heterologous Enzyme Expression

By
Andrew Brandon

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Plant Biology
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:
Professor Henrik V. Scheller, Chair
Professor Sheng Luan
Dr. Peggy Lemaux
Professor Michael Marletta

Fall 2019

Abstract

Reducing Xylan and Improving Lignocellulosic Biomass through Antimorphic and Heterologous Enzyme Expression

By

Andrew Brandon
Doctor of Philosophy in Plant Biology
University of California, Berkeley
Professor Henrik V. Scheller, Chair

Large-scale, sustainable production of lignocellulosic bioenergy from plant biomass will likely depend on a variety of dedicated bioenergy crops. Compared to their genetic and genomic diversity, polysaccharide composition varies little from one species to the next. Genetic engineering can be used to dramatically improve plant biomass for biofuel production. Engineering strategies that act independently of genetic context are particularly important, since they can improve the polysaccharide composition in a variety of species. Cellulose is the most abundant polysaccharide in all plants and the source of the six-carbon sugar glucose that is, ultimately, converted by genetically engineered microbes into liquid fuels and a variety of useful commodity chemicals. The ideal bioenergy crop would be as cellulose-dense as possible, and overexpression of plant and fungal enzymes can dramatically increase cellulose production and make simpler to break down. The second most abundant polysaccharide in bioenergy crops is xylan, composed of the five-carbon sugar xylose. Xylan associates tightly with cellulose microfibrils and, via its side chains, forms covalent linkages with other components of the cell wall, making xylan a significant contributor to the difficulty of biomass deconstruction. Xylan degrades and releases toxic compounds during the pretreatment of lignocellulosic biomass. These compounds and xylose itself are significantly detrimental to the microbial fermentation of glucose. Thus, a reduction in the amount of xylan and its complexity are additional characteristics that should be engineered into the ideal bioenergy crop. From a plant genetic engineering perspective, reducing the biosynthesis of a polysaccharide is significantly more challenging than increasing it, as a loss-of-function mutation in a specific gene would typically be required. Bioenergy crops often have multiple, functionally redundant, copies of a gene in their genome. Even with the advancement of genome editing technologies like CRISPR, homologous enzymes may have diverged in DNA sequence, making the task of knocking out each copy very laborious and time-consuming. I developed a protein-level genetic engineering approach to significantly reduce xylan content by overexpressing a catalytically dead or diminished (i.e. antimorphic) version of the xylan biosynthetic enzyme Irregular xylem10 (IRX10) in the model plant *Arabidopsis thaliana.* This work additionally provides experimental evidence supporting the hypothesis that xylan biosynthesis requires a complex of proteins in the Golgi, a complex in which IRX10 is the agent of catalysis. The journal publication of this work has been adapted in Chapter 1. High-level expression of the antimorph “out-competed” the native enzyme for its place in the complex, thereby generating xylan synthase complexes with little to no catalytic activity. A major advantage to this strategy is that sufficiently high expression would likely have the same effect on any redundant IRX10 homologs the plant may express. Additionally, the amino acid residues mutated to create the antimorphic protein are fully conserved in all land plants, so the technology could be translated to a variety of bioenergy crops with relative ease and the US Patent Application regarding it is the subject of Chapter 2. Continuing my study of dominant genetic engineering strategies for improving the qualities of biomass crops, I conducted an intensive study of carbohydrate active enzymes and proteins that have been heterologously expressed in plants to modify polysaccharides in muro

This invited review, for future publication in Frontiers in Plant Science, is covered in Chapter 3.

Table of Contents

This work is dedicated, first, to the memory of my mother, Carol. I've always thought my aptitude for science came from her and I know she would be proud of this accomplishment. I dedicate it to my father Douglas who, despite these past few years of turmoil and change in his own life, continues to support me every day. I thank my former undergraduate advisor, Dr. Barry Bruce, who encouraged me to apply to UC Berkeley in the first place. I also owe a great deal of knowledge to the innumerable post docs and research assistants that have passed through JBEI during my time here. Finally, of course, I could not have reached this point without the intuition and guidance of my advisor Dr. Henrik Scheller. It may have taken a little longer than expected, but I think the work was worth it.

The following chapter is adapted from "A dominant negative approach to reduce xylan in plants." *Plant Biotechnology Journal*. 25 June 2019 https://doi.org/10.1111/pbi.13198. More detail has been added to the introduction, experimental methods and discussion.

Chapter 1: A dominant negative approach to reduce xylan in bioenergy crop

Andrew G. Brandon*[1,2], Devon Birdseye[2] and Henrik V. Scheller[1,2,*1]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA

[2]Feedstocks Division, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

Summary

This study involves the identification of potentially catalytic residues in the xylan biosynthetic enzyme IRX10. In an effort to dominantly suppress plant xylan biosynthesis, mutations to these residues were made individually and each mutated IRX10 was overexpressed in wild-type *Arabidopsis thaliana*. Two of the mutant alleles were capable of significantly suppressing xylan biosynthesis: G238A and E293Q (mutation of conserved glycine residue 238 to alanine and glutamine residue 293 to glutamate, respectively). This was evidenced by a >55% decrease in stem cell wall content of xylose residues, and a reduced in vitro xylan biosynthetic activity using microsomes prepared from the mutant plants. The reduction in xylan is presumably a result of out-competing the native IRX10 for its place in a xylan synthase complex. Few tools are available to reduce the activity of a biosynthetic pathway in a targeted manner. The approach of using dominant negative protein variants as a biotechnology tool could be employed in many other systems where it is desired to inhibit an enzyme activity only in precisely defined cell types or under specific environmental or developmental conditions.

Introduction

Lignocellulosic biomass has a huge potential as a source for biofuels and bioproducts as an alternative to fossil fuels. Concerns about climate change and energy security have made the development of new, renewable sources of commodity chemicals and fuels a necessity. Metabolically engineered microbes can produce a near limitless variety of compounds from carbohydrates, including molecules identical to those comprising modern gasoline, diesel and jet fuels. The most abundant and renewable source of carbohydrates is found in the lignocellulosic biomass of plants, contained in the cell walls surrounding plant cells. However, the plant cell wall has evolved to be rigid and nearly impenetrable, making the conversion of plant biomass to simple sugars a laborious and expensive endeavor. This fact has created significant research interest in reducing the recalcitrance of plant biomass and improving its conversion to useable sugars.

The plant secondary cell wall is a complex structure made primarily of polysaccharides embedded in a matrix of the phenolic polymer lignin. After cellulose, a class of polysaccharides collectively known as hemicelluloses make up the majority of the carbohydrate content of plant biomass. The most abundant of the hemicelluloses in secondary cell walls of angiosperms is the β-(1,4)-linked xylose homopolymer xylan (Scheller & Ulvskov 2010). The xylan backbone is decorated with acetyl, glucuronic acid (GlcA) and 4-*O*-methyl glucuronic acid (Me-GlcA) residues. In grasses, xylan is also substituted with arabinofuranose, which can be further modified with xylose, ferulic acid or *p*-coumaric acid (Rennie 2014). Many genes with unique, if not well understood, roles in xylan biosynthesis have been identified. Members of two

glycosyltransferase (GT) families are required for synthesis of the xylan backbone: IRREGULAR XYLEM 9 (IRX9) and IRX14 from GT43 and IRX10 from GT47. In *A. thaliana*, each protein also has a close paralog (IRX9L, IRX14L and IRX10L) that is at least partially functionally redundant, although recent work suggests a role for these in primary cell wall biosynthesis (Mortimer et al. 2015). Mutants of each gene and its homolog have reduced xylan content, shorter xylan chains, and lower endogenous xylan synthase activity (Brown et al. 2009; Lee et al. 2010; Peña et al. 2007; Wu et al. 2010). The reduction in xylan results in much thinner and weaker secondary cell walls, leading to the irregular collapsed xylem vessels for which the genes are named. Of the proteins involved in xylan backbone elongation, only IRX10 and IRX10L have been demonstrated unambiguously to have β-(1,4)-xylosyltransferase activity *in vitro* (Urbanowicz et al. 2014; Jensen et al. 2014). It has also been shown that mutation of several conserved residues in IRX9 and IRX14 do not impair the protein's ability to complement their respective mutant phenotypes, suggesting non-catalytic roles for these proteins (Zeng et al. 2016; Ren et al. 2014). Thus, a xylan synthase complex has been proposed wherein IRX9(L) and/or IRX14(L) play primarily structural roles and IRX10(L) is responsible for elongation of the xylan backbone. The IRX10 homologs generally do not have predicted transmembrane regions and hence it is conceivable that IRX9(L) and IRX14(L) have a role in retaining IRX10(L) in the Golgi lumen. Nevertheless, biochemical evidence for the existence of a xylan synthase complex has been difficult to obtain, but has been reported for wheat and asparagus (Zeng et al. 2016, 2010).

While xylan plays a critical role in the structural integrity of the cell wall by interacting with cellulose microfibrils, this also makes xylan a significant contributor to the recalcitrance of plant biomass to enzymatic degradation. An additional, downstream obstacle presented by xylan is the large amount of pentoses and acetyl groups released during its enzymatic breakdown. Pentoses cannot be efficiently metabolized by and actually inhibit the fermentation of hexoses like glucose in commonly used microbial platforms, e.g. *Saccharomyces*, and acetate is generally detrimental to their growth (Laureano-Perez et al. 2005; Young et al. 2010). Thus, an important goal in the development of dedicated biofuel crops is reducing the amount of xylan in the secondary cell wall. Ideally, the approach would be able to reduce the amount of xylan in a wild-type plant since few knockout mutant lines of relevant biofuel crops exist. Generating mutations for example in IRX10 homologs would be possible using genome editing, but often many genes would have to be targeted, e.g. while *A. thaliana* has two IRX10 homologs, sorghum has four, and rice has seven. Furthermore, complete loss of xylan is detrimental to plant growth. We have shown that this can be addressed by complementing xylan deficiency specifically in vessels of xylan synthase mutants (Petersen et al. 2012). However, this is a complex approach and transforming unmodified crop plants with a dominant negative gene variant would allow much easier targeting to any tissue of choice than transformation of a xylan biosynthesis mutant that would have to be generated first. Here we report the dominant suppression of xylan biosynthesis in wild-type *A. thaliana* via the overexpression of two mutated forms of IRX10. This effect suggests that the mutated IRX10 is able to compete with the native form for its place in a protein complex, thereby significantly reducing xylan synthase activity.

Experimental procedures

Plant lines and growth conditions

A. thaliana Heynh. (L) accession Columbia-0 (Col-0) was obtained from the Arabidopsis Biological Resource Center (ABRC, Ohio State University). Plants were grown on soil in a

growth chamber under long-day conditions (16-h light/8-h dark) at 22°C and 60% relative humidity after being stratified at 4°C in the dark for 3 days. Arabidopsis plants were transformed using *A. tumefaciens* PGV3101 via the floral dip method (Clough & Bent, 1998). Seeds were germinated on soil under the conditions described above and 10-day old seedlings were sprayed once a day for a total of six days with glufosinate-ammonium solution (120 µg/mL). Resistant plants were then transplanted to new pots and grown as described above.

Bioinformatics and sequence alignments

The homologous GT47 sequences most similar to that of AtIRX10 (At1g27440) were chosen from 7 species: *A. thaliana* (IRX10-L, AT5G61840), *P. trichocarpa* (XP_002297880.1, POPTR_0001s12940g), *O. sativa* (Os01g0926700), *B. distachyon* (BRADI_2g59410), *P. ovata* (IRX10_6, APW77260.1), *P. patens* (XP_001753186), and *H. sapiens* (EXT1, AAB62718.1, 35% id.)

Site-directed mutagenesis

The wild-type coding sequence of IRX10 was amplified by PCR from *A. thaliana* cDNA and cloned into pDONR223 as described previously (Lao et al., 2014). Mutated versions of IRX10 were generated via the Gibson cloning method using the primers listed in Table 1 and using the pDONR223 clone as a template. Mutations were confirmed with sequencing in the pDONR223 entry clone. Each of the IRX10 point mutant sequences and the unmutated IRX10 sequence were inserted via LR clonase (Life Technologies) reaction into pEarleygate101 destination vector (Earley et al. 2006) modified to replace the C-terminal YFP-HA tag with a FLAG epitope tag. This vector contains the cauliflower mosaic virus 35S (pCaMV35S) promoter to drive constitutive, high-level expression in plants and glufosinate ammonium (BASTA) resistance marker. After BASTA selection, resistant plants were genotyped using a p35S forward primer and an OCS terminator reverse primer to confirm the presence of the transgene. The PCR product was then sequenced using an IRX10-specific sequencing primer to confirm the presence or absence of the point mutation.

Cell wall monosaccharide composition analysis

Alcohol-insoluble residue (AIR) was prepared from 6 cm of basal inflorescence stem of fully-senesced and dry plants as previously described (Harholt et al. 2006). AIR samples (1-3 mg) were then hydrolyzed with 2 M trifluoroacetic acid for 1 h at 120 °C. The samples were then lyophilized, resuspended in water and centrifuged at 20,000 g for 10 min. The supernatant was filtered through a 3 kDa MWCO spin filter and 4- and 25-fold dilutions in water were prepared for analysis by high-performance anion exchange chromatography (HPAEC) on an ICS 5000 instrument (ThermoFisher) using a CarboPac^{TM} PA20 (3 x 150 mm, ThermoFisher) anion exchange column (Øbro et al. 2004; Yin et al. 2011).

Staining and xylan immunolocalization microscopy

The basal regions of inflorescence stems (3 cm above the rosette) were collected from 10-week-old plants, and fixed in 4% paraformaldehyde in PEM buffer (50 mM PIPES, 5 mM EGTA, 5 mM $MgSO_4$, pH 6.9) overnight at 4 °C. The stems were embedded in 7% agarose and cut into 100 µm thick sections using a Leica VT1000S vibratome. Sections were washed three times with phosphate-buffered saline (PBS) pH 7 and incubated overnight with a 10-fold dilution of LM10 rat primary antibody in PBS (PlantProbes, Leeds, UK), which recognizes the xylan

backbone (McCartney et al. 2005). After three washes with PBS, the sections were then incubated fluorescein isothiocyanate (FITC)-conjugated goat anti-rat secondary antibodies for 1 hour. After three additional washes with PBS, the sections were imaged with a confocal microscope. Immunolabelled stem cuts were mounted on slides in a glycerol anti-fade solution (CitiFluor AF1, Agar Scientific) and pictures were taken using a Zeiss LSM710 confocal microscope and monitored and imaged by the ZEN2010 software (Zeiss).

Labeling with 8-aminonaphthalene-1,3,6-trisulfonic acid (ANTS) and analysis of labeled xylan substrates

Linear β-1,4-xylan oligomer (Xyl_6) was obtained from Megazyme (Bray, Ireland). Xylan substrates, 200 µg oligosaccharide dissolved in H_2O, was reductively aminated with ANTS (Invitrogen) as follows: oligosaccharides were dried on a speedvac and resuspended in 5 µL of 0.2 M ANTS solution (dissolved in 17:3, water:acetic acid), and 5 µl of 0.2 M 2-picoline borane [resuspended in dimethylsulfoxide (DMSO)] was added to each tube, as described in Mortimer et al. (2015). Samples were dried and dissolved in 100 µL water. Excess fluorophore was removed using GlykoClean S Cartridges (Prozyme). Labeled oligosaccharides were dried and resuspended to reach 1 µg/µL and stored at -80 °C until use.

Polysaccharide Analysis by Carbohydrate gel Electrophoresis (PACE)

All reactions were performed in a total of 25 µL containing $MnCl_2$ (10 mM), Triton X-100 (1% v/v) in buffer (50 mM MES, pH 6.5). The xylosyltransferase assays included 2 µg acceptor substrate, UDP-Xyl (200 µM) and microsomal membranes (50 µg total protein). Reactions were incubated for 2 h at 30 °C with agitation (800 rpm). Reactions were terminated by heating (100 °C, 3 min), and precipitated protein and lipids were sedimented by centrifugation (10,000 g for 10 min). Supernatants (15 µL) were mixed with 15 µL urea (3 M), and the samples (5 µL) were analyzed by separation on large format Tris-borate acrylamide gel prepared as described elsewhere (Goubet et al., 2002), and electrophoresed at 200 V for 30 min followed by 1000 V for 1.5 h. The PACE gels were visualized with a G-box (Syngene, www.syngene.com) equipped with a UV detection filter and long-wave UV tubes (365 nm emission).

Results and Discussion

Bioinformatic analysis of homologous sequences to AtIRX10

As of this publication, no x-ray crystal structure of a GT47 protein has been resolved and no conserved catalytic motifs have been identified. Therefore, the amino acid sequence of AtIRX10 was aligned with those of the closest homologs in an evolutionarily divergent range of species, and the degree of conservation of amino acid residues was used to infer potential roles of specific residues in catalysis. The amino acid sequences of *A. thaliana* IRX10 (AT1G27440) and a range of land plant species (*Oryza sativa, Populus trichocarpa, Brachypodium distachyon, Plantago ovata, Physcomitrella patens*), as well as the closest *Homo sapiens* homolog *Exostosin1* (EXT1), were used in an amino acid alignment (Fig. 1). The GT47 domain of HsEXT1, which also contains a GT64 domain, is an α-*N*-acetylglucosaminyltransferase required for the synthesis of heparan-sulfate (Duncan et al. 2001). Several mutant alleles of *HsEXT1* are linked to a bone development disorder called hereditary multiple exostoses. Only one of these disease-causing mutations is found in the GT47 portion of the protein: a missense mutation changing Glycine-339 into Asparagine (Phillipe et al. 1994). Glycine-339 in EXT1 corresponds

to Gly-283 in the AtIRX10 sequence. This glycine residue is conserved across all the land plant amino acid sequences analyzed, indicating it may be critical for the function of the protein and making G283 a promising candidate for mutation. G283 is clustered with three other residues that are fully conserved across land plants: C277, F278, and E293. The proximity of these residues to G283 and their large and/or charged functional groups made these residues attractive additional targets for mutation. C277 and F278 were mutated to alanine. E293 was mutated to glutamine to maintain the size but remove the charge of the functional group. A fifth fully conserved residue in a different portion of the protein, H146, was mutated to aspartate.

Transformation and phenotypic analysis

The native and mutated IRX10 coding sequences were LR cloned (Life Technologies) from their pDONR223 entry clone into a pEarleygate101 destination vector modified to replace the C-terminal YFP-HA tag with a FLAG epitope tag. This vector contains the cauliflower mosaic virus 35S (pCaMV35S) promoter to drive constitutive, high-level expression in plants. Each of these constructs was used to transform wild-type Columbia-0 *A. thaliana* via *Agrobacterium tumefaciens*-mediated floral dip (Clough & Bent 1998).
Several independent T1 lines (n=4-7) transformed with each T-DNA construct were chosen after selection for resistance to BASTA. Xylan biosynthesis mutants are characteristically dwarfed in stature. Transgenic plants carrying the F277A, G283D, and E293Q forms of IRX10 displayed significantly reduced height, an early indication that xylan biosynthesis may be suppressed (Fig. 2). Once the plants became fully senesced and dry, the monosaccharide composition of the stem (6 cm section, 2 cm from the rosette) was analyzed for each of the overexpressor lines. Each of the mutant IRX10 overexpressors, with the exception of F277A, demonstrated a reduction in mean xylose content, while the xylose content of plants overexpressing the unmutated form of IRX10 was comparable to empty vector control plants. Interestingly, the missense mutation associated with disease in the human homolog of IRX10, G283D, displayed the greatest decrease in xylose. Expression of the transgene was confirmed via immunoblot and lines with high-level expression of the IRX10, H146D, C278A, G283D and E293Q transgenes were chosen for further analysis in the T2 generation (Fig. 3). The suppression of xylan biosynthesis was even more pronounced in the T2 generation for constructs G283D and E293Q, but not for H146D or C278A (Fig. 4).

Microscopic analysis and comparison of xylan biosynthesis suppressors

The effects of the drastically reduced xylose content in G283D and E293Q were further investigated with xylan immunolabelling and toluidine blue-O staining of transverse sections of the basal stem (Fig. 5). The xylem vessels of wild-type plants and the native IRX10 overexpressors are characteristically large and round in shape with thick secondary cell walls. The xylem vessels of E293Q overexpressors, in contrast, are much smaller and more irregular in shape with significantly thinner secondary cell walls. Curiously, the G283D overexpressors exhibit a regular vessel phenotype much like wild type and have similarly thick secondary cell walls, despite having reduced stature and an even greater reduction in xylose than E293Q overexpressors.

In order to observe differences in cell wall xylan, stem sections were labeled with the LM10 monoclonal antibody, which recognizes unsubstituted or lowly substituted β-(1→4)-xylans (McCartney et al. 2005). The LM10 signal is specifically intense in cell types with secondary cell walls like the xylem vessels and interfascicular fibers in wild-type stem sections.

The intensity of LM10 labeling was comparable between wild type and transformants overexpressing unmutated IRX10, while it was reduced in the G283D and E293Q overexpressors consistent with the reduction in cell wall xylose.

The suppression of xylan biosynthesis was confirmed using polyacrylamide carbohydrate electrophoresis (PACE). Reactions using membrane prepared from the stems of plants were mixed with UDP-Xyl and fluorescently labeled Xyl_6 oligomer as the acceptor. Activity was measured by the intensity of bands corresponding to Xyl_7 and Xyl_8 on the gel (Fig. 6). $IRX10^{G283D}$ and $IRX10^{E293Q}$ had 43% and 3% of empty vector control (EVC), respectively.

Conclusion

This work reveals that there are at least two amino acid residues, G283 and E293, which are critically important in the xylan biosynthesis activity of IRX10. The corresponding amino acid to G283 in the human Exostosin 1 protein, G299, has been identified previously as a naturally occurring mutation associated with a disease caused by the faulty function of EXT1. The fact that overexpressing a mutated form of IRX10 is capable of disrupting the activity of the native IRX10 is consistent with the existence of a xylan synthase complex. The more abundant IRX10 mutant protein is apparently able to displace the functional IRX10 in the complex, thereby abrogating xylan biosynthetic activity. This additionally confirms previous work demonstrating that IRX10 is not alone sufficient for normal xylan backbone elongation *in vivo.* The proteins that IRX10 and IRX10L interact directly with in a biosynthetic complex have not been unambiguously identified, but there is significant evidence that IRX9(L) and IRX14(L) are part of the complex. It is conceivable that additional proteins are part of the complex as well (Jiang et al. 2016). The fact that IRX9 and IRX14 are required for xylan biosynthesis despite their apparent lack of catalytic activity can be explained by the need for these proteins in organizing and anchoring the catalytic IRX10(L) in the Golgi membrane. The existence of protein complexes required for biosynthesis of cell wall polysaccharides has often been suggested but hard evidence is lacking except for a few cases. A well-established example with some similarity to what is now understood for xylan synthesis is the homogalacturonan-synthesizing complex with GAUT1 as the catalytic subunit that, like IRX10, lacks a membrane anchor, and is anchored in the Golgi by interaction with the non-catalytic GAUT7 (Atmodjo et al. 2011).

From a practical perspective, the findings reported here readily suggest a method to improve biomass composition. Downregulation of xylan biosynthesis in bioenergy crops is desirable because an increased hexose/pentose ratio is advantageous for the downstream processing of biomass. Obviously, in the practical implementation downregulation must be done in a way that does not cause a yield reduction. However, prior work has demonstrated that when xylan is specifically reduced in interfascicular fiber cell while maintained in vessels the resulting plants have a large increase in hexose/pentose ratio while not showing any reduction in biomass yield (Petersen et al. 2012; Aznar et al. 2018). Indeed, *A. thaliana* engineered in this way even show improved drought tolerance (Yan et al. 2018). Hence, expression of the mutated IRX10 forms described here under strong fiber-specific promoters would be an obvious way to engineer plants with the desired properties. Alternatively, expression of the mutated form in vessels could be eliminated in various ways. The approach described here would be easily adapted to a wide variety of industrially relevant crop species.

The target we describe in this paper is xylan biosynthesis. However, the same kind of dominant negative approach can be used in most cases that involve a protein complex. The approach of using dominant negative protein variants as a biotechnology tool could be employed in many other systems where it is desired to inhibit an enzyme activity only in precisely defined cell types.

Bibliography

Atmodjo MA, Sakuragi Y, Zhu X, Burrell AJ, Mohanty SS, Atwood JA, Orlando R, Scheller HV & Mohnen D (2011) Galacturonosyltransferase (GAUT)1 and GAUT7 are the core of a plant cell wall pectin biosynthetic homogalacturonan:galacturonosyltransferase complex. *Proc. Natl. Acad. Sci. USA* 108: 20225–20230.

Aznar A, Chalvin C, Shih PM, Maimann M, Ebert B, Birdseye DS, Loqué D & Scheller HV (2018) Gene stacking of multiple traits for high yield of fermentable sugars in plant biomass. *Biotechnol Biofuels* 11: 2.

Brown DM, Zhang Z, Stephens E, Dupree P & Turner SR (2009) Characterization of IRX10 and IRX10-like reveals an essential role in glucuronoxylan biosynthesis in Arabidopsis. *Plant J.* 57: 732–746.

Clough SJ & Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal* 16: 735–743.

Duncan G, McCormick C & Tufaro F (2001) The link between heparan sulfate and hereditary bone disease: finding a function for the EXT family of putative tumor suppressor proteins. *J. Clin. Invest.* 108: 511–516.

Harholt J, Jensen JK, Sørensen SO, Orfila C, Pauly M & Scheller HV (2006) ARABINAN DEFICIENT 1 is a putative arabinosyltransferase involved in biosynthesis of pectic arabinan in Arabidopsis. *Plant Physiol.* 140: 49–58.

Jensen JK, Johnson NR & Wilkerson CG (2014) Arabidopsis thaliana IRX10 and two related proteins from psyllium and Physcomitrella patens are xylan xylosyltransferases. *Plant J.* 80: 207–215.

Jiang N, Wiemels RE, Soya A, Whitley R, Held M & Faik A (2016) Composition, assembly, and trafficking of a wheat xylan synthase complex. *Plant Physiol.* 170: 1999–2023.

Laureano-Perez L, Teymouri F, Alizadeh H & Dale BE (2005) Understanding Factors that Limit Enzymatic Hydrolysis of Biomass. *Twenty-Sixth Symposium on Biotechnology for Fuels and Chemicals*, (Davison, BH, Evans, BR, Finkelstein, M, & McMillan, JD, eds), pp. 1081–1099. Humana Press, Totowa, NJ.

Lee C, Teng Q, Huang W, Zhong R & Ye Z-H (2010) The Arabidopsis family GT43 glycosyltransferases form two functionally nonredundant groups essential for the elongation of glucuronoxylan backbone. *Plant Physiol.* 153: 526–541.

McCartney L, Marcus SE & Knox JP (2005) Monoclonal antibodies to plant cell wall xylans and arabinoxylans. *J. Histochem. Cytochem.* 53: 543–546.

Mortimer JC et al. (2015) An unusual xylan in Arabidopsis primary cell walls is synthesised by GUX3, IRX9L, IRX10L and IRX14. *Plant J.* 83: 413–426.

Øbro J, Harholt J, Scheller HV & Orfila C (2004) Rhamnogalacturonan I in Solanum tuberosum tubers contains complex arabinogalactan structures. *Phytochemistry* 65: 1429–1438.

Peña MJ, Zhong R, Zhou G-K, Richardson EA, O'Neill MA, Darvill AG, York WS & Ye Z-H (2007) Arabidopsis irregular xylem8 and irregular xylem9: implications for the complexity of glucuronoxylan biosynthesis. *Plant Cell* 19: 549–563.

Petersen PD, Lau J, Ebert B, Yang F, Verhertbruggen Y, Kim JS, Varanasi P, Suttangkakul A, Auer M, Loqué D, Scheller HV. (2012) Engineering of plants with improved properties as biofuels feedstocks by vessel-specific complementation of xylan biosynthesis mutants. *Biotechnol Biofuels* 5: 84.

Phillipe C, Porter DE, Emerton ME, Wells DE, Simpson AHRW & Monaco AP (1994) Mutation Screening of the EXT1 and EXT2 Genes in Patients with Hereditary Multiple Exostoses . *Am. J. Hum. Genet.*

Ren Y, Hansen SF, Ebert B, Lau J & Scheller HV (2014) Site-directed mutagenesis of IRX9, IRX9L and IRX14 proteins involved in xylan biosynthesis: glycosyltransferase activity is not required for IRX9 function in Arabidopsis. *PLoS One* 9: e105014.

Rennie E (2014) Xylan Biosynthesis. *Curr. Opin. Biotechnol.*

Scheller HV & Ulvskov P (2010) Hemicelluloses. *Annu. Rev. Plant Biol.* 61: 263–289.

Urbanowicz BR, Peña MJ, Moniz HA, Moremen KW & York WS (2014) Two Arabidopsis proteins synthesize acetylated xylan in vitro. *Plant J.* 80: 197–206.

Wu A-M, Hörnblad E, Voxeur A, Gerber L, Rihouey C, Lerouge P & Marchant A (2010) Analysis of the Arabidopsis IRX9/IRX9-L and IRX14/IRX14-L pairs of

glycosyltransferase genes reveals critical contributions to biosynthesis of the hemicellulose glucuronoxylan. *Plant Physiol.* 153: 542–554.

Yan J et al. (2018) Increased drought tolerance in plants engineered for low lignin and low xylan content. *Biotechnol Biofuels* 11: 195.

Yin L et al. (2011) The cooperative activities of CSLD2, CSLD3, and CSLD5 are required for normal Arabidopsis development. *Mol. Plant* 4: 1024–1037.

Young E, Lee S-M & Alper H (2010) Optimizing pentose utilization in yeast: the need for novel tools and approaches. *Biotechnol Biofuels* 3: 24.

Zeng W et al. (2016) Asparagus IRX9, IRX10, and IRX14A Are Components of an Active Xylan Backbone Synthase Complex that Forms in the Golgi Apparatus. *Plant Physiol.* 171: 93–109.

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

Table 1

Mutation	
H146D	F: 5'-TgACTTCTTTGTTGTGCCCCATGACTTTGG-3'
	R: 5'-AGTcATCAGCTCCTTCAGTCCGATTCC-3'
F277A	F: 5'-5'ATCgcCTGTCTATGTCCTCTAGGATGG3'-3'
	R: 5'-ACAGgcGATGGCTCTCTGCATGTC-3'
C278A	F: 5'-TCgcTCTATGTCCTCTAGGATGGGC-3'
	R: 5'-AGAgcGAAGATGGCTCTCTGCATGTC-3'
G283D	F: 5'-TAGatTGGGCTCCATGGAGCCCTAG-3'
	R: 5'-AatCTAGAGGACATAGACAGAAGATGGCTCTC-3'
E293Q	F: 5'-TTcAAGCAGTTGTGTTTGGGTGCATTCCAG-3'
	R: 5'-TTgAACTAACCTAGGGCTCCATGGAGC-3'
	lower case indicates mutated bases

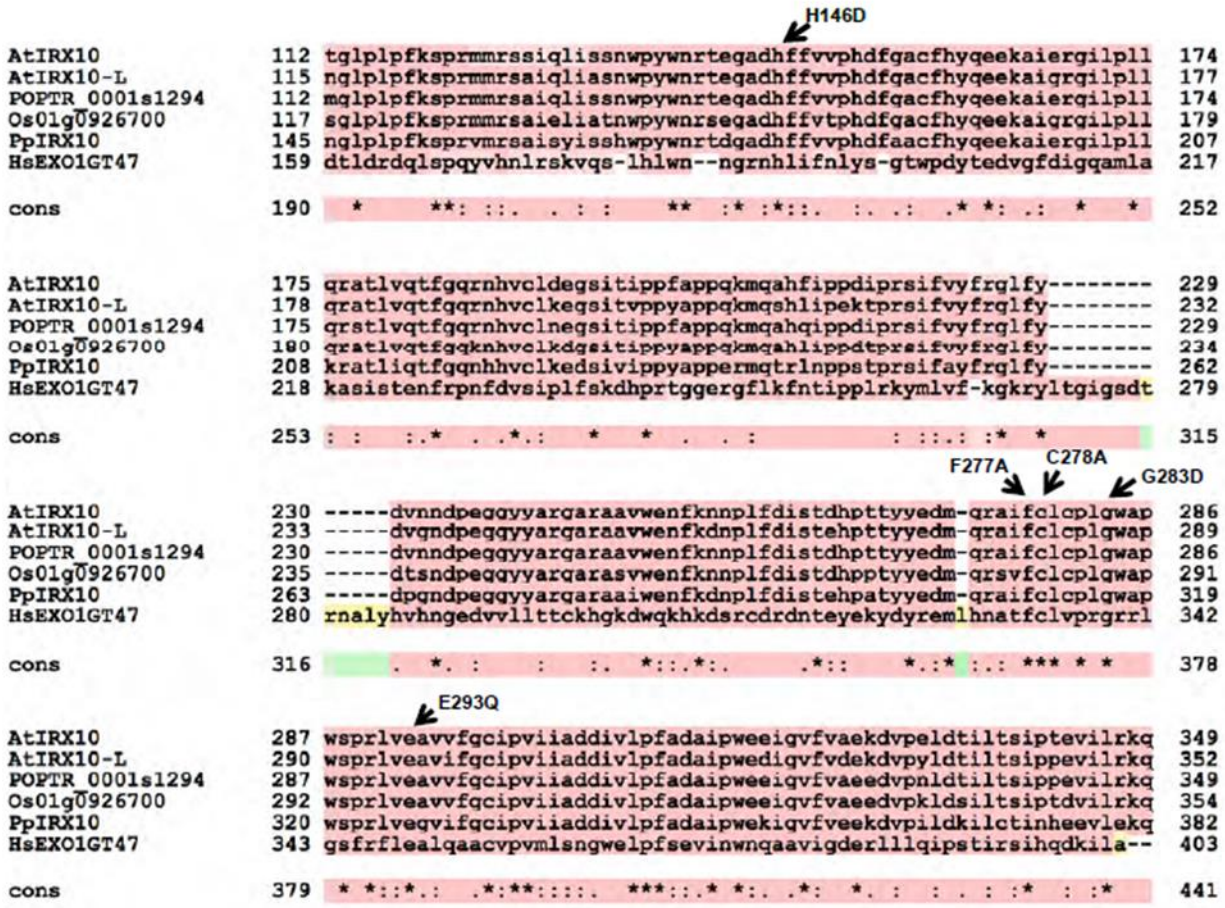

Fig. 1 Amino acid sequence alignment of Arabidopsis IRX10 and IRX10-L, with homologs from *Populus tremulus* (POPTR_0001s1294), *Oryza sativa* (Os01g0926700), *Physcomitrella patens* (PpIRX10), and the GT47-region of *Homo sapiens* Exostosin 1 (HsEXO1GT47). The five amino acid residues targeted in this study are indicated.

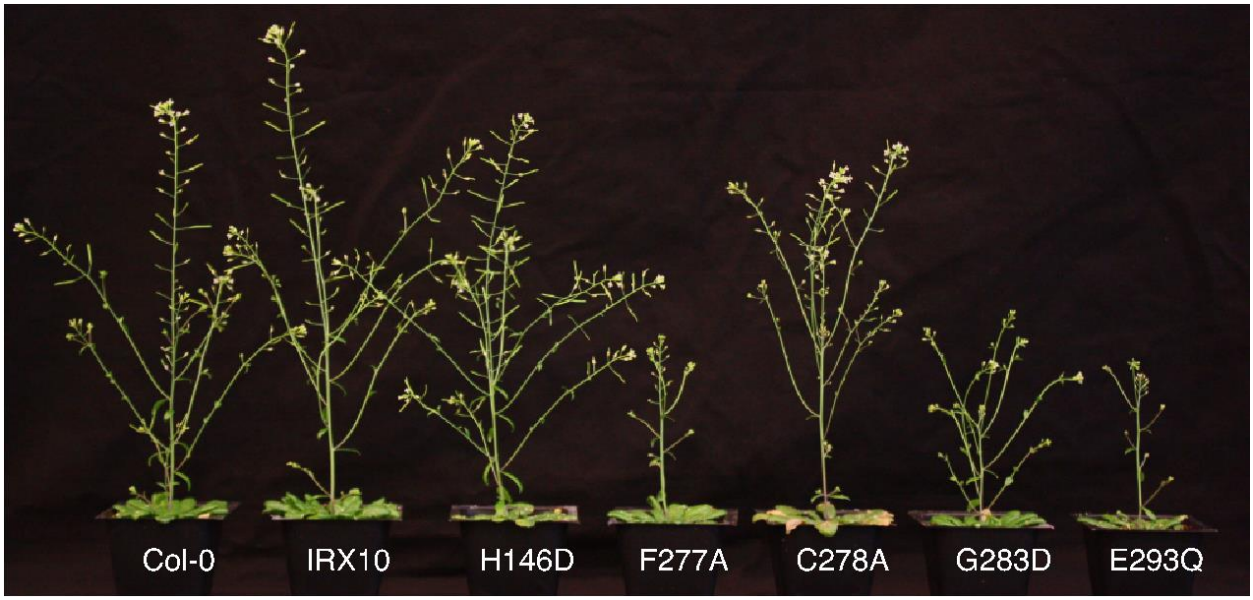

Fig 2. Representative plants from the T1 generation. Col-0 is a non-transformed control, and 'IRX10' is a wild-type plant transformed with the unmodified IRX10.

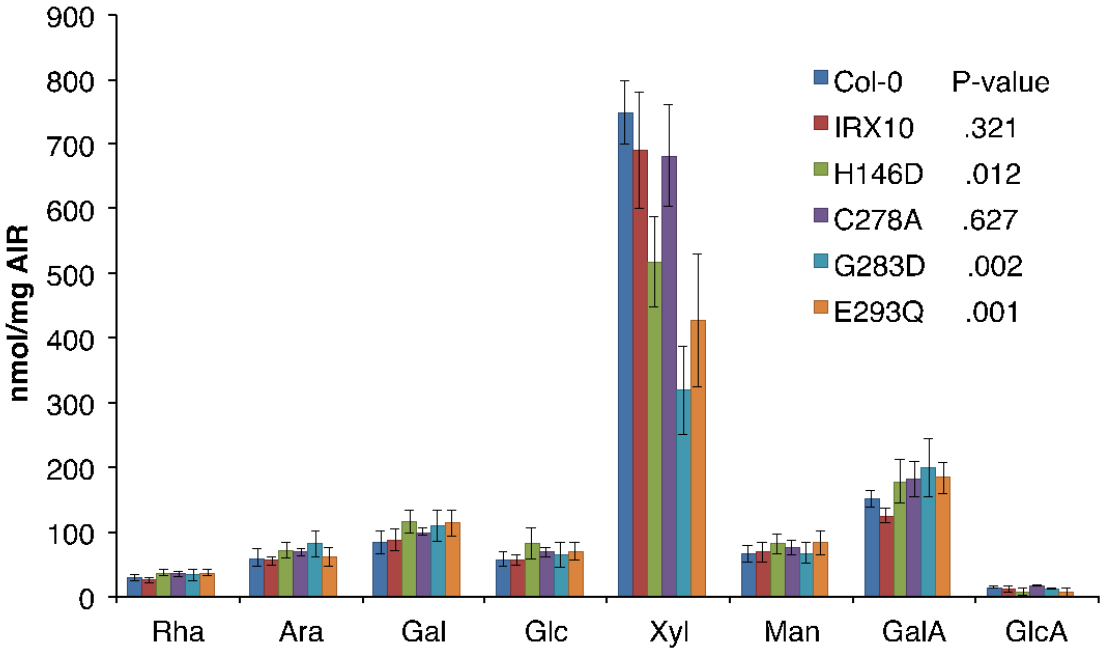

Fig 4. Monosaccharide composition of the basal stem. Data shows average ± SD (n = 4 biological replicates). P-values indicates significance of difference in xylose content compared to Col-0 (t-test).

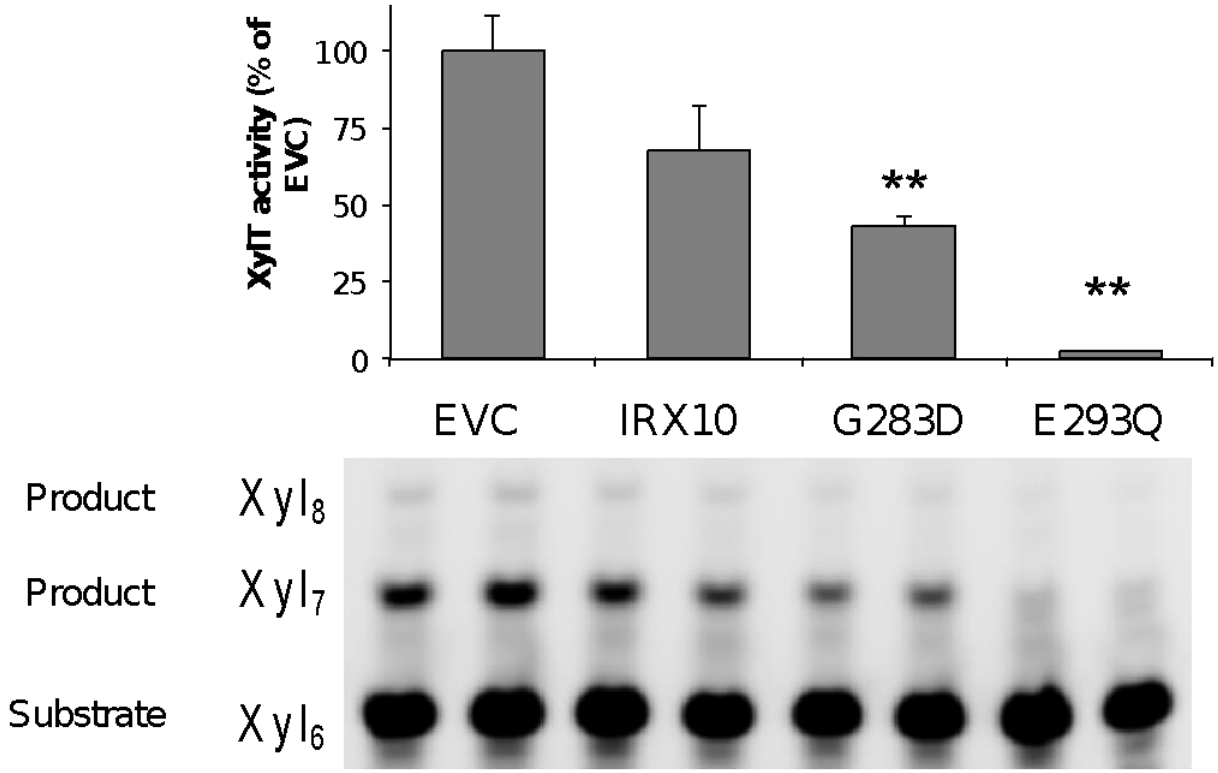

Fig 6. Xylan synthase activity in transgenic plants. Microsomes were prepared from EVC, IRX10, G283D, and E293Q Arabidopsis stems and used in assays using UDP-xylose and ANTS-labeled xylohexaose as acceptor. Reaction products were analyzed by PACE.

Chapter 2: MUTANT XYLAN BIOSYNTHETIC ENZYMES CAPABLE OF DOMINANT SUPPRESSION OF XYLAN BIOSYNTHESIS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Patent Application Ser. No. 62/435,687, filed on Dec. 16, 2016, which is hereby incorporated by reference.

STATEMENT OF GOVERNMENTAL SUPPORT

[0002] The invention was made with government support under Contract No. DE-AC02-05CH11231 awarded by the U.S. Department of Energy. The government has certain rights in the invention.

FIELD OF THE INVENTION

[0003] The present invention is in the field of xylan biosynthesis in plants.

BACKGROUND OF THE INVENTION

[0004] Xylan is the most abundant non-cellulosic poly-saccharide in plant biomass and one of the most abundant biopolymers on earth. The xylan backbone is a homopolymer of β-(1,4)-linked xylose, decorated at regular intervals with GlcA, 4-O-MeG!cA and acetyl groups. As a hemicellulose, xylan is thought to coat and crosslink cellulose microfibrils, promoting their crystallinity. Indeed, xylan is critical for the overall health and mechanical strength of the plant. Xylan biosynthesis mutants are severely dwarfed due to cell wall collapse in the water-conducting xylem vessels. While important, the relatively high amount of xylan in plant biomass creates several problems for the development of advanced biofuels. Xylose, a 5-carbon sugar, is poorly utilized by microorganisms and strongly inhibits the fermentation of 6-carbon sugars like glucose. Additionally, the acetate released from the xylan backbone creates a toxic environment for microbial growth. Any way to reduce the amount of xylan in plant biomass could have a significant effect on the conversion efficiency to biofuel. Since few, if any, mutants of biotechnologically relevant crops exist, the ideal approach would act as a dominant suppressor of xylan biosynthesis.

SUMMARY OF THE INVENTION

[0005] The present invention provides for a polypeptide capable of dominant suppression of a first naturally occur-ring IRXl0, wherein the polypeptide comprises an amino acid sequence having at least 70% identity as compared to a second naturally occurring IRXl0 wherein the

polypeptide comprises one or more of the conserved amino acid indi-cated in FIG. **2** substituted with a different amino acid residue. The conserved amino acid residues are the ones which are identical for IRXl0, IRXl0-L, OsIRXl0, PpIRXl0, as indicated in FIG. **2.** In some embodiments, the conserved amino acid residue is indicated by an asterisk in FIG. **2.** In some embodiments, the first naturally occurring IRXl0 and the second naturally occurring IRXl0 are the same IRXl0.

[0006] In some embodiments, the conserved amino acid residue corresponds to the histidine at position 146, the phenylalanine at position 277, the cysteine at position 278, the glycine at position 283, or glutamate at position 293 of *Arabidopsis* IRXl0. In some embodiments, the polypeptide has one or more of the following substitutions: H146D, F277A, C278A, G283D, or E293Q.

[0007] In some embodiments, the naturally occurring IRXl0, for the first or second naturally occurring IRXl0, or both, is *Arabidopsis* IRXl0, IRXl0-L, OsIRXl0, PpIRXl0, or HsEXOl.

[0008] This invention provides for a means to identify potential catalytic residues in the xylan biosynthetic enzyme IRXl0 and mutating them. Overexpression of the mutated IRXl0 outcompetes the native form of the enzyme, sup- pressing the biosynthesis of the polymer.

[0009] The present invention provides for a genetically modified eukaryotic host cell comprising (a) a gene encoding a polypeptide of the present invention operably linked to a promoter, wherein the gene and/or the promoter is heter- ologous to the cell. The host cell has a native IRXl0. In some embodiment, the native IRXl0 is disrupted. The modified cell is altered in producing xylan and produces modified cellulose and/or cell wall that comprises less xylan. In some embodiments, the host cell is a plant cell. In some embodiments, the host cell is part of a plant. In some embodiments, the host cell is a plant cell wherein all of the cells of the plant are similarly modified.

[0010] The present invention provides for a plant comprising the cell of the present invention, or a progeny thereof.

[0011] The present invention provides for a seed from the plant of the present invention.

[0012] The present invention provides for a biomass comprising plant tissue from the plant of the present invention.

[0013] The present invention provides for a method of obtaining a polypeptide of the present invention, comprising: (a) providing a nucleic acid encoding a naturally occurring IRXl0, (b) introducing or generating a mutation into an open reading frame (ORF) encoding the naturally occurring IRXl0 which results in an amino acid substitution of a conserved amino acid residue, as indicated in FIG. **2,** in the naturally occurring IRXl 0, (c) optionally introducing the nucleic acid into a eukaryotic host cell, and (d) optionally culturing or growing the eukaryotic host cell.

[0014] In some embodiments, the method results in a decrease of the amount of xylan in a plant comprising the plant cell.

[0015] In some embodiments, host cell is part of a plant and the method further comprises: collecting plant material from the plant, and optionally incubating the plant material from the plant in a saccharification reaction.

[0016] The present invention provides for a method of improving the amount of soluble sugar obtained from a plant biomass material, comprising: (a) providing plant biomass material from a plant which expresses the polypeptide of the present invention, (b) performing a saccharification reaction on the plant biomass material, and (c) obtaining soluble sugar.

[0017] The present invention provides for a saccharification reaction comprising grass plant biomass material from a plant which expresses the polypeptide of the present invention.

[0018] The present invention provides for a method of engineering a plant to increase the content of a sugar in a desired tissue, comprising: (a) introducing an expression cassette into the plant, wherein the expression cassette comprises a polynucleotide encoding the polypeptide of the present invention operably linked to a heterologous promoter, and (b) culturing the plant under conditions in which the polypeptide is expressed in the desired tissue; wherein the heterologous promoter specifically expresses in the desired tissue. In some embodiments, the desired tissue is plant vessel tissue.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] The foregoing aspects and others will be readily appreciated by the skilled artisan from the following description of illustrative embodiments when read in con- junction with the accompanying drawings.

[0020] FIG. **1** shows the function ofIRXl0 in the synthesis of cellulose.

[0021] FIG. **2** shows a comparison the amino acid sequences between *Arabidopsis* IRXl0, IRXl0-L, OsIRXl0, PpIRXl0, and HsEXOl. IRXl0 is well con- served within diverse homologs, including human EXOl. Conserved amino acid residues are indicated.

[0022] FIG. **3** shows that mutant plants overexpressing the mutant IRXl0 have a phenotype consistent with reduced xylan.

[0023] FIG. **4** shows that suppressors may alter the substitution pattern of the xylan backbone.

[0024] FIG. **5** shows the monosaccharide analysis of T2 stems.

[0025] FIG. **6** shows a construct for restricting the expression of the suppressor in the plant vessels.

DETAILED DESCRIPTION OF THE INVENTION

[0026] Before the present invention is described, it is to be understood that this invention is not limited to particular embodiments described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims.

[0027] Where a range of values is provided, it is under- stood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limits of that range is also specifically disclosed. Each smaller range between any stated value or intervening value in a stated range and any other stated or intervening value in that stated range is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included or excluded in the range, and each range where either, neither or both limits are included in the smaller ranges is also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention.

[0028] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are now described. All publications mentioned herein are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited.

[0029] The terms "polynucleotide" and "nucleic acid" are used interchangeably and refer to a single or double- stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. A nucleic acid of the present invention will generally contain phosphodiester bonds, although in some cases, nucleic acid analogs may be used that may have alternate backbones, comprising, e.g., phosphoramidate, phosphorothioate, phosphorodithioate, or O-methylphophoroamidite linkages (see Eckstein, Oligonucleotides and Analogues: A Practical Approach, Oxford University Press); positive backbones; non-ionic backbones, and non-ribose backbones. Thus, nucleic acids or polynucleotides may also include modified nucleotides that permit correct read-through by a polymerase. "Polynucleotide sequence" or "nucleic acid sequence" includes both the sense and antisense strands of a nucleic acid as either individual single strands or in a duplex. As will be appreciated by those in the art, the depiction of a single strand also defines the sequence of the complementary strand; thus the sequences described herein also provide the complement of the sequence. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses variants thereof (e.g., degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. The nucleic acid may be DNA, both genomic and cDNA, RNA or a hybrid, where the nucleic acid may contain combinations of deoxyriboand ribonucleotides, and combinations of bases,

including uracil, adenine, thymine, cytosine, guanine, inosine, xanthine hypoxanthine, isocytosine, isoguanine, etc.

[0030] Two nucleic acid sequences or polypeptide sequences are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below. The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence over a comparison window, as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. When percentage of sequence identity is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions, where amino acids residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non- conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated according to, e.g., the algorithm of Meyers & Miller, *Computer Applic Biol. Sci.* 4:11-17 (1988) e.g., as implemented in the pro- gram PC/GENE (Intelligenetics, Mountain View, Calif., USA).

[0031] For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

[0032] A "comparison window," as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, e.g. by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mal. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'!. Acad. Sci. USA* 85:2444 (1988),

by computerized implementations of these algorithms (GAP, BESTFIT, PASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by manual alignment and visual inspection.

[0033] Algorithms that are suitable for determining per-cent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al. (1990) *J. Mal. Biol.* 215: 403-410 and Altschul et al. (1977) *Nucleic Acids Res.* 25: 3389-3402, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (NCBI) web site. The algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive- valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al, supra). These initial neighborhood word hits acts as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >O) and N (penalty score for mismatching residues; always <O). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a word size (W) of 28, an expectation (E) of 10, M=l, N=-2, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a word size (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)).

[0034] The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, *Proc. Nat'!. Acad. Sci. USA* 90:5873- 5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.01, more preferably less than about 10^{-5} , and most preferably less than about 10^{-20}.

[0035] Nucleic acid or protein sequences that are substantially identical to a reference sequence include "conservatively modified variants." With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at

every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein, which encodes a polypeptide, also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid, which encodes a polypeptide, is implicit in each described sequence.

[0036] As to amino acid sequences, one of skill will recognize that individual substitutions, in a nucleic acid, peptide, polypeptide, or protein sequence which alters a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art.

[0037] The following six groups each contain amino acids that are illustrative conservative substitutions for one another: (1) Alanine (A), Serine (S), Threonine (T); (2) Aspartic acid (D), Glutamic acid (E); (3) Asparagine (N), Glutamine (Q); (4) Arginine (R), Lysine (K); (5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and (6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W). (see, e.g., Creighton, *Proteins* (1984)).

[0038] Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other, or a third nucleic acid, under stringent conditions. Stringent conditions are sequence dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5° C. lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, stringent conditions will be those in which the salt concentration is about 0.02 molar at pH 7 and the temperature is at least about 60° C. For example, stringent conditions for hybridization, such as RNA-DNA hybridizations in a blotting technique are those which include at least one wash in 0.2xSSC at 55° C. for 20 minutes, or equivalent conditions.

[0039] The term "promoter," as used herein, refers to a polynucleotide sequence capable of driving transcription of a DNA sequence in a cell. Thus, promoters used in the polynucleotide constructs of the invention include cis- and trans-acting transcriptional control elements and regulatory sequences that are involved in regulating or modulating the timing and/or rate of transcription of a gene. For example, a promoter can be a cis-acting transcriptional control element, including an enhancer, a promoter, a transcription terminator, an origin of replication, a chromosomal integration sequence, 5' and 3' untranslated regions, or an intronic sequence, which are involved in transcriptional regulation. These cis-acting sequences typically interact with proteins or other biomolecules to carry out (tum on/off, regulate, modulate, etc.) gene transcription. Promoters are located 5' to the transcribed gene, and as used herein, include the sequence 5' from the translation start codon (i.e., including the 5' untranslated region of the

mRNA, typically comprising 100-200 bp). Most often the core promoter sequences lie within 1-2 kb of the translation start site, more often within 1 kbp and often within 500 bp of the translation start site. By convention, the promoter sequence is usually provided as the sequence on the coding strand of the gene it controls. In the context of this application, a promoter is typically referred to by the name of the gene for which it naturally regulates expression. A promoter used in an expression construct of the invention is referred to by the name of the gene. Reference to a promoter by name includes a wild type, native promoter as well as variants of the promoter that retain the ability to induce expression. Reference to a promoter by name is not restricted to a particular species, but also encompasses a promoter from a corresponding gene in other species.

[0040] A "constitutive promoter" in the context of this invention refers to a promoter that is capable of initiating transcription in nearly all cell types, whereas a "cell type- specific promoter" or "tissue-specific promoter" initiates transcription only in one or a few particular cell types or groups of cells forming a tissue (for a multi-cellular organ- ism). In some embodiments, a plant promoter is tissue- specific if the transcription levels initiated by the promoter in the cell wall are at least 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 50-fold, 100-fold, 500-fold, 1000-fold higher or more as compared to the transcription levels initiated by the promoter in non-cell wall tissues

[0041] A polynucleotide is "heterologous" to an organism or a second polynucleotide sequence if it originates from a foreign species, or, if from the same species, is modified from its original form. For example, when a polynucleotide encoding a polypeptide sequence is said to be operably linked to a heterologous promoter, it means that the poly- nucleotide coding sequence encoding the polypeptide is derived from one species whereas the promoter sequence is derived from another, different species; or, if both are derived from the same species, the coding sequence is not naturally associated with the promoter (e.g., is a genetically engineered coding sequence, e.g., from a different gene in the same species, or an allele from a different ecotype or variety).

[0042] The term "operably linked" refers to a functional relationship between two or more polynucleotide (e.g., DNA) segments. Typically, it refers to the functional relationship of a transcriptional regulatory sequence to a transcribed sequence. For example, a promoter or enhancer sequence is operably linked to a DNA or RNA sequence if it stimulates or modulates the transcription of the DNA or RNA sequence in an appropriate host cell or other expression system. Generally, promoter transcriptional regulatory sequences that are operably linked to a transcribed sequence are physically contiguous to the transcribed sequence, i.e., they are cis-acting. However, some transcriptional regulatory sequences, such as enhancers, need not be physically contiguous or located in close proximity to the coding sequences whose transcription they enhance.

[0043] The term "expression cassette" or "DNA construct" or "expression construct" refers to a nucleic acid construct that, when introduced into a host cell, results in transcription and/or translation of an RNA or polypeptide, respectively. Antisense or sense constructs that are not or cannot be translated are expressly included by this definition. In the case of both expression of transgenes and suppression of endogenous genes (e.g., by antisense, RNAi, or sense

suppression) one of skill will recognize that the inserted polynucleotide sequence need not be identical, but may be only substantially identical to a sequence of the gene from which it was derived. As explained herein, these substantially identical variants are specifically covered by reference to a specific nucleic acid sequence. One example of an expression cassette is a polynucleotide construct that comprises a polynucleotide sequence encoding a protein operably linked to a heterologous promoter. In some embodiments, an expression cassette comprises a polynucleotide sequence encoding a protein that is targeted to a position in a plant genome such that expression of the polynucleotide sequence is driven by a promoter that is present in the plant

[0044] The term "plant" as used herein can refer to a whole plant or part of a plant, e.g., seeds, and includes plants of a variety of ploidy levels, including aneuploid, polyploid, diploid and haploid. The term "plant part," as used herein, refers to shoot vegetative organs and/or structures (e.g., leaves, stems and tubers), branches, roots, flowers and floral organs (e.g., bracts, sepals, petals, stamens, carpels, anthers), ovules (including egg and central cells), seed (including zygote, embryo, endosperm, and seed coat), fruit (e.g., the mature ovary), seedlings, and plant tissue (e.g., vascular tissue, ground tissue, and the like), as well as individual plant cells, groups of plant cells (e.g., cultured plant cells), protoplasts, plant extracts, and seeds. The class of plants that can be used in the methods of the invention is generally as broad as the class of higher and lower plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), gymnosperms, ferns, bryophytes, and multicellular algae.

[0045] The term "biomass," as used herein, refers to plant material that is processed to provide a product, e.g., a biofuel such as ethanol, or livestock feed, or cellulose for paper and pulp industry products. Such plant material can include whole plants, or parts of plants, e.g., stems, leaves, branches, shoots, roots, tubers, and the like.

[0046] The term "saccharification reaction" refers to a process of converting biomass, usually cellulosic or lignocellulosic biomass, into monomeric sugars, such as glucose and xylose.

[0047] The term "soluble sugar" refers to monomeric, dimeric, or trimeric sugar that is produced from the saccharification of biomass.

[0048] The term "increased amount," when referring to an amount of sugar or soluble sugar obtained from an engineered plant of the present invention, refers to an increase in the amount or yield of sugar that is obtained from saccharification of biomass per amount of starting material, in comparison to corresponding biomass from a wild-type (i.e., naturally occurring) plant. In the context of the present invention, "corresponding biomass from a wild-type plant" refers to plant material that is from the same part of the plant as the biomass from a plant engineered to have modified sugar levels. As understood in the art, increased amount or increased yield is based upon comparisons of the same amount of corresponding plant material.

[0049] The term "conversion reaction," as used herein, refers to a reaction that converts biomass into a form of bioenergy. Examples of conversion reactions include, but are not limited to,

combustion (burning), gasification, pyrolysis, and polysaccharide hydrolysis (enzymatic or chemical).

[0050] The term "increased production," when referring to an amount of bioenergy production obtained from an engineered plant of the present invention, refers to an increased amount of bioenergy that is produced from subjecting bio- mass from an engineered plant to a conversion reaction (e.g., combustion, gasification, pyrolysis, or polysaccharide hydrolysis) as compared to the amount of bioenergy that is produced from corresponding biomass from a wild-type (i.e., naturally occurring) plant.

[0051] The terms "optional" or "optionally" as used herein mean that the subsequently described feature or structure may or may not be present, or that the subsequently described event or circumstance may or may not occur, and that the description includes instances where a particular feature or structure is present and instances where the feature or structure is absent, or instances where the event or circumstance occurs and instances where it does not.

[0052] These and other objects, advantages, and features of the invention will become apparent to those persons skilled in the art upon reading the details of the invention as more fully described below.

[0053] This invention is useful for engineering bioenergy plants with a cell wall composition that makes their sugars easier accessible. This will be interesting for a variety of industries, such as biofuel production or sugar producing industries. Likewise, the invention could be useful for developing plants for other purposes, such as for feed and forage.

[0054] In one aspect, the invention provides a method of engineering plants to decrease xylan content. Eukaryotic cells can be engineered to overexpress one or more poly- peptide in a cell by genetically modifying the cell to over- express one or more polypeptide genes as described herein. In some embodiments, plants can be engineered to overexpress express one or more polypeptide in the plant by genetically modifying the plant to overexpress one or more polypeptide genes as described herein. Typically, overexpression is targeted to cell wall using a tissue-specific promoter. An example of a method for fine-tuning gene expression to increase expression in the cell wall is taught in PCT/US2012/023182, which is incorporated by reference.

[0055] The invention employs various routine recombinant nucleic acid techniques. Generally, the nomenclature and the laboratory procedures in recombinant DNA technology described below are those well known and commonly employed in the art. Many manuals that provide direction for performing recombinant DNA manipulations are available, e.g., Sambrook & Russell, Molecular Cloning, A Laboratory Manual (3rd Ed, 2001); and Current Protocols in Molecular Biology (Ausubel, et al., John Wiley and Sons, New York, 2009).

[0056] In some embodiments, the IRXl0 is an IRXl0 of *Arabidopsis,* poplar, eucalyptus, rice, com, cotton, switch- grass, sorghum, millet, miscanthus, sugarcane, pine, alfalfa, wheat, soy, barley, turfgrass, tobacco, hemp, bamboo, rape, sunflower, willow, or *Brachypodium.*

[0057] The amino acid sequence of *Arabidopsis thaliana*IRXl0 is as follows:
(SEQ ID NO, 1)
MKIHSCLSAILLFLFFSASSAKQNVRTERISGSAGDVLEDDPVGKLKVYVYELPSKYNK
KLLQKDPRCLTHMFAAEIFMHRFLLSSPVRTRNPDEADWFYTPIYPTCDLTPTGLPLPFK
SPRMMRSSIQLISSNWPYWNRTEGADHFFVVPHDFGACFHYQEEKAIERGILPLLQRAT
LVQTFGQRNHVCLDEGSITIPPFAPPQKMQAHFIPPDIPRSIFVYFRGLFYDVNNDPEGGY
YARGARAAVWENFKNNPLFDISTDHPTTYYEDMQRAIFCLCPLGWAPWSPRLVEAVVF
GCIPVIIADDIVLPFADAIPWEEIGVFVAEKDVPELDTILTSIPTEVILRKQRLLANPSMKR
AMLFPQPAQPGDAFHQILNGLARKLPHDKSIYLKTGEKALNWTAGPVADLKPW

[0058] The amino acid sequence of *Arabidopsis thaliana* IRXl0L is as follows:
(SEQ ID NO, 2)
MKLSSCVLIFLLCNTFSSISAFRLSRSQPTERISGSAGDVLEDDPVGRLKVFVYELPSKYN
KKILQKDPRCLNHMFAAEIYMQRFLLSSPVRTLNPEEADWFYVPVYTTCDLTPNGLPLP
FKSPRMMRSAIQLIASNWPYWNRTEGADHFFVVPHDFGACFHYQEEKAIGRGILPLLQR
ATLVQTFGQRNHVCLKEGSITVPPYAPPQKMQSHLIPEKTPRSIFVYFRGLFYDVGNDPE
GGYYARGARAAVWENFKDNPLFDISTEHPTTYYEDMQRAIFCLCPLGWAPWSPRLVEA
VIFGCIPVIIADDIVLPFADAIPWEDIGVFVDEKDVPYLDTILTSIPPEVILRKQRLLANPSM
KQAMLFPQPAQPGDAFHQVLNGLARKLPHERSVYLRPGEKLLNWTAGPVADLKPW

[0059] The amino acid sequence of rice OsIRXl0 (as known as Os01g70200) is as follows:
(SEQ ID NO, 3)
MRRWVLAIAILAAAVCFFLGAQAQEVRQGHQTERISGSAGDVLEDDPVGRLKVYVYD
LPSKYNKKLLKKDPRCLNHMFAAEIFMHRFLLSSAVRTFNPEEADWFYTPVYTTCDLTP
SGLPLPFKSPRMMRSAIELIATNWPYWNRSEGADHFFVTPHDFGACFHYQEEKAIGRGI
LPLLQRATLVQTFGQKNHVCLKDGSITIPPYAPPQKMQAHLIPPDTPRSIFVYFRGLFYDT
SNDPEGGYYARGARASVWENFKNNPLFDISTDHPPTYYEDMQRSVFCLCPLGWAPWSP
RLVEAVVFGCIPVIIADDIVLPFADAIPWEEIGVFVAEEDVPKLDSILTSIPTDVILRKQRLL
ANPSMKQAMLFPQPAQAGDAFHQILNGLARKLPHGENVFLKPGERALNWTAGPVGDL
KPW

[0060] The amino acid sequence of *Physcomitrella patens*IRXl0 (PpIRXl0) is as follows:
(SEQ ID NO, 5)
MEHPLECADSCSL AMSWFCNKKCR GWGLMKRTVVASGLRSVVLL LLFIYFV
QDVTAEMGHQRISGSAGDVLEDNPVGRLKVFIYDIPSKYNTDWLKKDPRCL
THMFAVEEYLHDFLTESPVRTLNPEEADWFYTPVYTTCDLTPNGLPLPFKSPRVMRSAIS
YISSHWPYWNRTDGADHFFVVPHDFAACFHYQEEKAIERGILPLLKRATLIQTFGQNHH
VCLKEDSIVIPPYAPPERMQTRLNPPSTPRSIFAYFRGLFYDPGNDPEGGYYARGARAAI

WENFKDNPLFDISTEHPATYYEDMQRAIFCLCPLGWAPWSPRLVEGVIFGCIPVIIADDIV
LPFADAIPWEKIGVFVEEKDVPILDKILCTINHEEVLEKQRLLANPAMKQAMLFPRPAKP
GDAFHQILNGLARKLPHDPSIYLQPGQSFLNWTEGPPGDLYPWGNDL

[0061] To use isolated sequences in the above techniques, recombinant DNA vectors suitable for transformation of plant cells such as crop plant cells are prepared. Techniques for transformation are well known and described in the technical and scientific literature. For example, a DNA sequence encoding the polypeptide can be combined with transcriptional and other regulatory sequences which will direct the transcription of the sequence from the gene in the intended cells, e.g., grass or other crop plant cells. In some embodiments, an expression vector that comprises an expression cassette that comprises the polypeptide gene further comprises a promoter operably linked to the poly- peptide gene. In other embodiments, a promoter and/or other regulatory elements that direct transcription of the polypeptide gene are endogenous to the plant and an expression cassette comprising the polypeptide gene is introduced, e.g., by homologous recombination, such that the heterologous polypeptide gene is operably linked to an endogenous promoter and is expression driven by the endogenous promoter. Regulatory sequences include promoters, which may be either constitutive or inducible, or tissue-specific.

Tissue-Specific Promoters

[0062] In some embodiments, a plant promoter to direct expression of the polypeptide of the present invention in a specific tissue is employed (tissue-specific promoters). Tis- sue specific promoters are transcriptional control elements that are only active in particular cells or tissues at specific times during plant development, such as in vegetative tis- sues or reproductive tissues.

[0063] Examples of tissue-specific promoters under developmental control include promoters that initiate transcription only (or primarily only) in certain tissues, such as vegetative tissues, cell walls, including e.g., roots or leaves. A variety of promoters specifically active in vegetative tissues, such as leaves, stems, roots and tubers are known. For example, promoters controlling patatin, the major storage protein of the potato tuber, can be used (see, e.g., Kim, *Plant Mal. Biol.* 26:603-615, 1994; Martin, *Plant J.* 11:53- 62, 1997). The ORF13 promoter from *Agrobacterium rhizogenes* that exhibits high activity in roots can also be used (Hansen, *Mal. Gen. Genet.* 254:337-343, 1997). Other useful vegetative tissue-specific promoters include: the tarn promoter of the gene encoding a globulin from a major taro *(Colocasia esculenta L. Schott)* corm protein family, tarin (Bezerra, *Plant Mal. Biol.* 28:137-144, 1995); the curculin promoter active during taro corm development (de Castro, *Plant Cell* 4:1549-1559, 1992) and the promoter for the tobacco root-specific gene TobRB7, whose expression is localized to root meristem and immature central cylinder regions (Yamamoto, Plant Cell 3:371-382, 1991).

[0064] Leaf-specific promoters, such as the ribulose biphosphate carboxylase (RBCS) promoters can be used. For example, the tomato RBCSl, RBCS2 and RBCS3A genes are expressed in leaves and light-grown seedlings, only RBCSl and RBCS2 are expressed in developing tomato fruits (Meier, *FEES Lett.* 415:91-95, 1997). A ribulose bisphosphate carboxylase promoters

expressed almost exclusively in mesophyll cells in leaf blades and leaf sheaths at high levels (e.g., Matsuoka, *Plant J.* 6:311-319, 1994), can be used. Another leaf-specific promoter is the light harvesting chlorophyll alb binding protein gene promoter (see, e.g., Shiina, *Plant Physiol.* 115:477-483, 1997; Casal, *Plant Physiol.* 116:1533-1538, 1998). The *Arabidopsis thaliana* myb-related gene promoter (Atmyb5) (Li, et al., *FEES Lett.* 379:117-121 1996), is leaf-specific. The Atmyb5 promoter is expressed in developing leaf trichomes, stipules, and epidermal cells on the margins of young rosette and cauline leaves, and in immature seeds. Atmyb5 mRNA appears between fertilization and the 16 cell stage of embryo development and persists beyond the heart stage. A leaf promoter identified in maize (e.g., Busk et al., *Plant J.* 11:1285-1295, 1997) can also be used.

[0065] Another class of useful vegetative tissue-specific promoters are meristematic (root tip and shoot apex) promoters. For example, the "SHOOTMERISTEMLESS" and "SCARECROW" promoters, which are active in the developing shoot or root apical meristems, (e.g., Di Laurenzio, et al., *Cell* 86:423-433, 1996; and, Long, et al., Nature 379: 66-69, 1996); can be used. Another useful promoter is that which controls the expression of 3-hydroxy-3-methylglutaryl coenzyme A reductase HMG2 gene, whose expression is restricted to meristematic and floral (secretory zone of the stigma, mature pollen grains, gynoecium vascular tissue, and fertilized ovules) tissues (see, e.g., Enjuto, Plant Cell. 7:517-527, 1995). Also useful are knl-related genes from maize and other species which show meristem-specific expression, (see, e.g., Granger, *Plant Mal. Biol.* 31:373-378, 1996; Kerstetter, *Plant Cell* 6:1877-1887, 1994; Hake, Philos. Trans. R. Soc. Land. B. Biol. Sci. 350:45-51, 1995). For example, the *Arabidopsis thaliana* KNATl promoter (see, e.g., Lincoln, *Plant Cell* 6:1859-1876, 1994) can be used.

[0066] In some embodiments, the promoter is substantially identical to the native promoter of a promoter that drives expression of a gene involved in secondary wall deposition. Examples of such promoters are promoters from IRXl, IRX3, IRX5, IRX8, IRX9, IRX14, IRX7, IRXl0, GAUT13, or GAUT14 genes. Specific expression in fiber cells can be accomplished by using a promoter such as the NSTl promoter and specific expression in vessels can be accomplished by using a promoter such as VND6 or VND7. (See, e.g., PCT/US2012/023182 for illustrative promoter sequences).

[0067] One of skill will recognize that a tissue-specific promoter may drive expression of operably linked sequences in tissues other than the target tissue. Thus, as used herein a tissue-specific promoter is one that drives expression pref- erentially in the target tissue, but may also lead to some expression in other tissues as well.

Constitutive Promoters

[0068] A promoter, or an active fragment thereof, can be employed which will direct expression of a nucleic acid encoding a fusion protein of the invention, in all or most transformed cells or tissues, e.g. as those of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include those from

viruses which infect plants, such as the cauliflower mosaic virus (CaMV) 35S transcription initiation region (see, e.g., Dagless, *Arch. Viral.* 142:183-191, 1997); the 1'- or 2'-promoter derived from T-DNA of *Agrobacterium tumefaciens* (see, e.g., Mengiste supra (1997); O'Grady, Plant Mo!. Biol. 29:99-108, 1995); the promoter of the tobacco mosaic virus; the promoter of Figwort mosaic virus (see, e.g., Maiti, *Transgenic Res.* 6:143-156, 1997); actin promoters, such as the *Arabidopsis* actin gene promoter (see, e.g., Huang, *Plant Mal. Biol.* 33:125-139, 1997); alcohol dehydrogenase (Adh) gene promoters (see, e.g., Millar, *Plant Mal. Biol.* 31:897- 904, 1996); ACTll from *Arabidopsis* (Huang et al., *Plant Mal. Biol.* 33:125-139, 1996), Cat3 from *Arabidopsis* (Gen- Bank No. U43147, Zhong et al., *Mal. Gen. Genet.* 251:196- 203, 1996), the gene encoding stearoyl-acyl carrier protein desaturase from *Brassica napus* (Genbank No. X74782, Solocombe et al., *Plant Physiol.* 104:1167-1176, 1994), GPcl from maize (GenBank No. X15596, Martinez et al., *J. Mal. Biol.* 208:551-565, 1989), Gpc2 from maize (GenBank No. U45855, Manjunath et al., *Plant Mal. Biol.* 33:97-112, 1997), other transcription initiation regions from various plant genes known to those of skill. See also Holtorf, "Comparison of different constitutive and inducible promoters for the overexpression of transgenes in *Arabidopsis thaliana," Plant Mal. Biol.* 29:637-646, 1995).

Inducible Promoters

[0069] In some embodiments, a plant promoter may direct expression of the nucleic acids under the influence of changing environmental conditions or developmental conditions. Examples of environmental conditions that may effect transcription by inducible promoters include anaerobic conditions, elevated temperature, drought or other environmental stress, or the presence of light. Examples of developmental conditions that may effect transcription by inducible promoters include senescence and embryogenesis. Such promoters are referred to herein as "inducible" promoters. For example, the invention can incorporate drought- specific promoter such as the drought-inducible promoter of maize (Busk et al., *Plant J,* 11: 1285-95, 1997); or alternatively the cold, drought, and high salt inducible promoter from potato (Kirch *Plant Mal. Biol.* 33:897-909, 1997).

[0070] Suitable promoters responding to biotic or abiotic stress conditions include the pathogen inducible PRPl-gene promoter (Ward et al., *Plant. Mal. Biol.* 22:361-366, 1993), the heat inducible hsp80-promoter from tomato (U.S. Pat. No. 5,187,267), cold inducible alpha-amylase promoter from potato (PCT Publication No. WO 96/12814) or the wound-inducible pinII-promoter (European Patent No. 375091). For other examples of drought, cold, and salt-inducible promoters, such as the RD29A promoter, see, e.g., Yamaguchi-Shinozalei et al., *Mal. Gen. Genet.* 236:331-340, 1993 are also known

[0071] Alternatively, plant promoters which are inducible upon exposure to plant hormones, such as auxins, may be used to express the polypeptide gene. For example, the invention can use the auxin-response elements El promoter fragment (AuxREs) in the soybean *(Glycine max L.)* (Liu, *Plant Physiol.* 115:397-407, 1997); the auxin-responsive *Arabidopsis* GST6 promoter (also responsive to salicylic acid and hydrogen peroxide) (Chen, Plant J. 10: 955-966, 1996); the auxin-inducible parC promoter from tobacco (Sakai, 37:906-913, 1996); a plant biotin response element (Streit, *Mal. Plant Microbe Interact.* 10:933-937, 1997); and, the promoter responsive to the stress hormone abscisic acid (Sheen, *Science* 274:1900-1902, 1996).

[0072] In further embodiments, a plant can be engineered to overexpress the polypeptide using a positive feedback loop to express the polypeptide in a desired tissue. In some embodiments, a promoter for use in the polypeptide expression construct is responsive to a transcription factor that mediates expression in the desired tissue. The polypeptide expression construct is used in a genetically modified plant comprising an expression construct encoding a transcription factor where expression is also driven by a promoter that is responsive to the transcription factor. Examples of such expression systems are provided in PCT/US2012/023182, hereby incorporated by reference.

[0073] In some embodiments in which a positive feedback loop is employed, the plant is genetically modified to express a transcription factor that regulates the production of secondary cell wall. Examples of such transcription factors include NSTl, NST2, NST3, SND2, SND3, MYB103, MBY85, MYB46, MYB83, MYB58, and MYB63 (See, e.g., Mitsuda et al., *Plant Cell* 17:2993-3006 (2005); Mitsuda et al., *Plant Cell* 19:270-80 (2007); Ohashi-Ito et al., *Plant Cell* 22:3461-73 (2010); Zhong et al., *Plant Cell* 20:2763-82 (2008); Zhong et al., *Plant Cell* 19:2776-92 (2007); Ko et al., *Plant J.* 60:649-65 (2009); and McCarthy et al., *Plant Cell Physiol.* 50:1950-64 (2009)). Illustrative examples of gene and protein sequences and/or accession numbers for NSTl, NST2, NST3, SND2, SND3, MYB103, MBY85, MYB46, MYB83, MYB58, and MYB63 are provided in PCT/US2012/023182, hereby incorporated by reference.

[0074] In some embodiments, the polynucleotide encoding the transcription factor that regulates secondary cell wall production is operably linked to a promoter that is a down- stream target of the transcription factor. Similarly, the poly- peptide nucleic acid sequence is also linked to a promoter that is a downstream target of the transcription factor. The promoter may be the same promoter or different promoters. In such an embodiment, a promoter is suitable for use with the transcription factor that regulates secondary cell wall production if expression of the promoter is induced, directly or indirectly, by the transcription factor to be expressed, and if the promoter is expressed in the desired location, e.g., the stem of the plant.

[0075] In another embodiment, the polynucleotide encoding the polypeptide is expressed through a transposable element. This allows for constitutive, yet periodic and infrequent expression of the constitutively active polypeptide. The invention also provides for use of tissue-specific promoters derived from viruses including, e.g., the tobacco virus subgenomic promoter (Kumagai, *Proc. Natl. Acad. Sci. USA* 92:1679-1683, 1995); the rice tungro bacilliform virus (RTBV), which replicates only in phloem cells in infected rice plants, with its promoter which

drives strong phloem specific reporter gene expression; the cassava vein mosaic virus (CVMV) promoter, with highest activity in vascular elements, in leaf mesophyll cells, and in root tips (Verdaguer, *Plant Mal. Biol.* 31:1129-1139, 1996).

[0076] A vector comprising nucleic acid sequences encod- ing the polypeptide will typically comprise a marker gene that confers a selectable phenotype on the cell to which it is introduced. Such markers are known. For example, the marker may encode antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, and the like.

[0077] Nucleic acid sequences encoding the polypeptide of the invention are expressed recombinantly in plant cells as described. As appreciated by one of skill in the art, expression constructs can be designed taking into account such properties as codon usage frequencies of the plant in which the nucleic acid encoding the polypeptide is to be expressed. Codon usage frequencies can be tabulated using known methods (see, e.g., Nakamura et al. *Nucl. Acids Res.* 28:292, 2000). Codon usage frequency tables are available in the art (e.g., from the Codon Usage Database at the internet site www.kazusa.or.jp/codon/.)

[0078] Additional sequence modifications may be made that are also known to enhance gene expression in a plant. These include elimination of sequences encoding spurious polyadenylation signals, exon-intron splice site signals, transposon-like repeats, and other such well-characterized sequences that may be deleterious to gene expression. The G-C content of the sequence may be adjusted to levels average for a given cellular host, as calculated by reference to known genes expressed in the host cell. When possible, the sequence may also be modified to avoid predicted hairpin secondary mRNA structures.

Production of Modified Cells or Transgenic Plants

[0079] In some embodiments, the modified eukaryotic cell is a plant cell. Techniques for genetically modifying eukaryotic cells, such as plant, animal and fungal cells are well known to those skilled in the art. For example, U.S. Provisional Patent Application Ser. No. 61/676,811 teaches such methods for yeast.

[0080] In some embodiments, the plant is a grass plant. In some embodiments, the plant of plant cell is *Arabidopsis*, poplar, eucalyptus, rice, corn, cotton, switchgrass, sorghum, millet, miscanthus, sugarcane, pine, alfalfa, wheat, soy, barley, turfgrass, tobacco, hemp, bamboo, rape, sunflower, willow, or *Brachypodium*

[0081] The present invention provides for transgenic plants comprising recombinant expression cassettes either for expressing the polypeptide. It should be recognized that the term "transgenic plants" as used here encompasses the plant or plant cell in which the expression cassette is introduced as well as progeny of such plants or plant cells that contain the expression cassette, including the progeny that have the expression cassette stably integrated in a chromosome.

[0082] Once an expression cassette comprising a polynucleotide encoding the polypeptide has been constructed, standard techniques may be used to introduce the polynucleotide into a plant in order to modify gene expression. See, e.g., protocols described in Ammirato et al. (1984) Hand- book of Plant Cell Culture-Crop Species. Macmillan Pub!. Co. Shimamoto et al. (1989) Nature 338:274-276; Fromm et al. (1990) Bio/Technology 8:833-839; and Vasil et al. (1990) Bio/Technology 8:429-434.

[0083] Transformation and regeneration of plants is known in the art, and the selection of the most appropriate transformation technique will be determined by the practitioner. Suitable methods may include, but are not limited to: electroporation of plant protoplasts; liposome-mediated transformation; polyethylene glycol (PEG) mediated trans- formation; transformation using viruses; micro-injection of plant cells; micro-projectile bombardment of plant cells; vacuum infiltration; and *Agrobacterium tumeficiens* mediated transformation. Transformation means introducing a nucleotide sequence in a plant in a manner to cause stable or transient expression of the sequence. Examples of these methods in various plants include: U.S. Pat. Nos. 5,571,706; 5,677,175; 5,510,471; 5,750,386; 5,597,945; 5,589,615; 5,750,871; 5,268,526; 5,780,708; 5,538,880; 5,773,269; 5,736,369 and 5,610,042.

[0084] Transformed plant cells derived by any of the above transformation techniques can be cultured to regenerate a whole plant that possesses the transformed genotype and thus the desired phenotype such as enhanced drought-resistance. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker, which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans et al., Protoplasts Isolation and Culture, Handbook of Plant Cell Culture, pp. 124-176, MacMillilan Publishing Company, New York, 1983; and Binding, Regeneration of Plants, Plant Protoplasts, pp. 21-73, CRC Press, Boca Raton, 1985. Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally, e.g., in Klee et al. *Ann. Rev. of Plant Phys.* 38:467-486, 1987.

[0085] One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

[0086] The expression constructs of the invention can be used to increase the sugar content of cell walls of essentially any plant. The plant may be a monocotyledonous plant or a dicotyledonous plant. In some embodiments of the invention, the plant is a green field plant. In some embodiments, the plant is a gynmosperm or conifer. Thus, the invention has use over a broad range of plants, including species from the genera *Asparagus, Atropa, Avena, Brassica, Cannabis, Citrus, Citrullus, Camelina, Capsicum, Cucumis, Cucurbita, Daucus, Fragaria, Glycine, Gossypium, Helianthus, Heterocallis, Hordeum, Hyoscyamus, Lactuca, Linum,*

Lolium, Lycopersicon, Malus, Manihot, Majorana, Medicago, Nicotiana, Oryza, Panieum, Pannesetum, Persea, Pisum, Pyrus, Prunus, Raphanus, Secale, Senecio, Sinapis, Solanum, Sorghum, Trigonella, Triticum, Vitis, Vigna, and, *Zea.* In some embodiments, the plant is corn, switchgrass, sorghum, mis- canthus, sugarcane, poplar, pine, wheat, rice, soy, cotton, barley, turf grass, tobacco, potato, bamboo, rape, sugar beet, sunflower, willow, and eucalyptus. In further embodiments, the plant is reed canarygrass *(Phalaris arundinacea), Miscanthus* x *giganteus, Miscanthus* sp., sericea lespedeza *(Lespedeza cuneata*), millet, ryegrass *(Lolium multiflorum, Lolium* sp.), timothy, Kochia *(Kochia scoparia*), forage soybeans, alfalfa, clover, sunn hemp, kenaf, bahiagrass, bermudagrass, dallisgrass, pangolagrass, big bluestem, indiangrass, fescue *(Festuca* sp.), *Dactylis* sp., *Brachypodium distachyon,* smooth bromegrass, orchardgrass, or Kentucky bluegrass among others. In some embodiments, the plant is an ornamental plant. In some embodiments, the plant is a grass plant. In some embodiment, the plant is a vegetable- or fruit-producing plant. In some embodiments, the plant is a plant that is suitable for generating biomass, including plants as noted above, e.g., *Arabidopsis,* poplar, eucalyptus, rice, corn, switchgrass, sorghum, millet, miscanthus, sugarcane, pine, alfalfa, wheat, soy, barley, turfgrass, tobacco, hemp, bamboo, rape, sunflower, willow, *Jatropha,* and *Brachypodium.*

[0087] In some embodiments, the plant into which the expression construct comprising a nucleic acid sequence that encodes the polypeptide is introduced is the same species of plant from which the mutant IRXl0 sequence, and/ or the promoter driving expression of the mutant IRXl0 sequence, is obtained. In some embodiments, the plant into which the expression construct is introduced is a different species of plant compared to the species from which the mutant IRXl0 and/or promoter sequence was obtained.

[0088] Plants that overexpress the mutant IRXl0 can be identified using any known assay, including analysis of RNA, protein, or xylan composition. The xylan levels can be determined directly or indirectly, wherein such methods are well known in the art.

[0089] An expression cassette comprising a polynucleotide encoding the polypeptide operably linked to a promoter, as described herein, can be expressed in various kinds of plants. The plant may be a monocotyledonous plant or a dicotyledonous plant. In some embodiments of the invention, the plant is a green field plant. In some embodiments, the plant is a gynmosperm or conifer.

[0090] In some embodiments, the plant is a plant that is suitable for generating biomass. Examples of suitable plants include, but are not limited to, *Arabidopsis,* poplar, eucalyptus, rice, corn, switchgrass, sorghum, millet, miscanthus, sugarcane, pine, alfalfa, wheat, soy, barley, turfgrass, tobacco, hemp, bamboo, rape, sunflower, willow, *Jatropha,* and *Brachypodium.*

[0091] In some embodiments, the plant into which the expression cassette is introduced is the same species of plant as the promoter and/or as the polynucleotide encoding the polypeptide or

transcription factor (e.g., a vessel-specific promoter and/or transcription factor from *Arabidopsis* is expressed in an *Arabidopsis* plant). In some embodiments, the plant into which the expression cassette is introduced is a different species of plant than the promoter and/or than the polynucleotide encoding the polypeptide or transcription factor (e.g., a vessel-specific promoter and/or transcription factor from *Arabidopsis* is expressed in a poplar plant). See, e.g., McCarthy et al., *Plant Cell Physiol.* 51:1084-90 (2010); and Zhong et al., *Plant Physiol.* 152:1044-55 (2010).

[0092] Methods of enzymatic saccharification are also known in the art. Briefly, plants or plant biomass material (e.g., leaves and stems) are optionally pre-treated with hot water, dilute acid, alkali, or ionic liquid followed by enzymatic saccharification using a mixture of cellulases and hemicellulases and pectinases in buffer and incubation of the plants or plant biomass material with the enzymatic mixture. Following incubation, the yield of the saccharification reaction can be readily determined by measuring the amount of reducing sugar released, using a standard method for sugar detection, e.g. the dinitrosalicylic acid method well known to those skilled in the art. Plants engineered in accordance with the invention provide a higher sugar yield as compared to wild-type plants.

[0093] It is to be understood that, while the invention has been described in conjunction with the preferred specific embodiments thereof, the foregoing description is intended to illustrate and not limit the scope of the invention. Other aspects, advantages, and modifications within the scope of the invention will be apparent to those skilled in the art to which the invention pertains.

[0094] All patents, patent applications, and publications mentioned herein are hereby incorporated by reference in their entireties.

[0095] The invention having been described, the following examples are offered to illustrate the subject invention by way of illustration, not by way of limitation.

EXAMPLE 1

Dominant Suppression of Xylan Biosynthesis Using IRXl0

[0096] Potential catalytic residues in the xylan biosynthetic enzyme IRXl0 are identified and mutated. The over- expression of the mutated IRXl0 out competes the native form of the enzyme resulting in the suppression of the biosynthesis of the polymer.

[0097] FIG. **2** shows a comparison the amino acid sequences between *Arabidopsis* IRXl0, IRXl0-L, OsIRXl0, PpIRXl0, and HsEXOl. IRXl0 is well con- served within diverse homologs, including human EXOl. The indicated IRX10 mutants shown in FIG. **2** are generated and are overexpressed in plants. FIG. **3** shows that the mutant plants overexpressing the mutant IRXl0s have a phenotype consistent with reduced xylan.

[0098] Cell wall material from the plants in FIG. **3** are isolated and digested with Xylanase C, an enzyme that cleaves xylan specifically at glucuronic acid substitutions. The digestion

products are then fluorescently labeled and separated by size. FIG. **4** shows the distribution of glucuronic acid residues along the xylan chain. FIG. **4** shows that suppressors may alter the substitution pattern of the xylan backbone. Cell wall material from the basal stem of at least 3 biological replicates is fully hydrolyzed with TFA and the monosaccharides separated and quantified via HPAEC. FIG. **5** shows the monosaccharide analysis of T2 stems. The results indicate that the biosynthesis of xylan is reduced in the plants with the mutant IRXl0.

[0099] To eliminate the yield reduction associated with reduced xylan, expression of the suppressor is restricted specifically to the vessels. This can be accomplished by expressing the Cre recombinase under the vessel specific promoter pVND7. Cre recognizes cognate loxP sites, looping out any sequence between the sites from the genome. FIG. **6** shows a construct for restricting the expression of the suppressor in the plant vessels.

[0100] The results demonstrate the following: IRXl0 is at least 70% conserved in all land plants and potential catalytic residues can be identified. Mutations in some of the residues chosen are able to dominantly suppress xylan biosynthesis and reduce the amount of xylose in the plant by as much as about 55%. The mutation G283D, is the same mutation linked to cancer in the human homolog EXOl, exhibits significant suppression.

[0101] While the present invention has been described with reference to the specific embodiments thereof, it should be understood by those skilled in the art that various changes may be made and equivalents may be substituted without departing from the true spirit and scope of the invention. In addition, many modifications may be made to adapt a particular situation, material, composition of matter, process, process step or steps, to the objective, spirit and scope of the present invention. All such modifications are intended to be within the scope of the claims appended hereto.

SEQUENCE LISTING
<160> NUMBER OF SEQ ID NOS, 5

<210> SEQ ID NO 1
<211> LENGTH, 412
<212> TYPE, PRT
<213> ORGANISM, Arabidopsis thaliana

<400> SEQUENCE, 1

MetLysIleHisSerCysLeuSerAlaIleLeuLeuPheLeuPhePheSerAlaSerSerAla
LysGlnAsnValArgThrGluArgIleSerGlySerAlaGlyAspValLeuGluAspAspPro
ValGlyLysLeuLysValTyrValTyrGluLeuProSerLysTyrAsnLysLysLeuLeuGln
LysAspProArgCysLeuThrHisMetPheAlaAlaGluIlePheMetHisArgPheLeuLeu
SerSerProValArgThrArgAsnProAspGluAlaAspTrpPheTyrThrProIleTyrPro
ThrCysAspLeuThrProThrGlyLeuProLeuProPheLysSerProArgMetMetArgSer

SerIleGlnLeuIleSerSerAsnTrpProTyrTrpAsnArgThrGluGlyAlaAspHisPhe
PheValValProHisAspPheGlyAlaCysPheHisTyrGlnGluGluLysAlaIleGluArg
GlyIleLeuProLeuLeuGlnArgAlaThrLeuValGlnThrPheGlyGlnArgAsnHisVal
CysLeuAspGluGlySerIleThrIleProProPheAlaProProGlnLysMetGlnAlaHis
PheIleProProAspIleProArgSerIlePheValTyrPheArgGlyLeuPheTyrAspVal
AsnAsnAspProGluGlyGlyTyrTyrAlaArgGlyAlaArgAlaAlaValTrpGluAsnPhe
LysAsnAsnProLeuPheAspIleSerThrAspHisProThrThrTyrTyrGluAspMetGln
ArgAlaIlePheCysLeuCysProLeuGlyTrpAlaProTrpSerProArgLeuValGluAla
ValValPheGlyCysIleProValIleIleAlaAspAspIleValLeuProPheAlaAspAla
IleProTrpGluGluIleGlyValPheValAlaGluLysAspValProGluLeuAspThrIle
LeuThrSerIleProThrGluValIleLeuArgLysGlnArgLeuLeuAlaAsnProSerMet
LysArgAlaMetLeuPheProGlnProAlaGlnProGlyAspAlaPheHisGlnIleLeuAsn
GlyLeuAlaArgLysLeuProHisAspLysSerIleTyrLeuLysThrGlyGluLysAlaLeu
AsnTrpThrAlaGlyProValAlaAspLeuLysProTrp***

<210> SEQ ID NO 2
<211> LENGTH, 415
<212> TYPE, PRT
<213> ORGANISM, Arabidopsis thaliana

<400> SEQUENCE, 2
MetLysLeuSerSerCysValLeuIlePheLeuLeuCysAsnThrPheSerSerIleSerAla
PheArgLeuSerArgSerGlnProThrGluArgIleSerGlySerAlaGlyAspValLeuGlu
AspAspProValGlyArgLeuLysValPheValTyrGluLeuProSerLysTyrAsnLysLys
IleLeuGlnLysAspProArgCysLeuAsnHisMetPheAlaAlaGluIleTyrMetGlnArg
PheLeuLeuSerSerProValArgThrLeuAsnProGluGluAlaAspTrpPheTyrValPro
ValTyrThrThrCysAspLeuThrProAsnGlyLeuProLeuProPheLysSerProArgMet
MetArgSerAlaIleGlnLeuIleAlaSerAsnTrpProTyrTrpAsnArgThrGluGlyAla
AspHisPhePheValValProHisAspPheGlyAlaCysPheHisTyrGlnGluGluLysAla
IleGlyArgGlyIleLeuProLeuLeuGlnArgAlaThrLeuValGlnThrPheGlyGlnArg
AsnHisValCysLeuLysGluGlySerIleThrValProProTyrAlaProProGlnLysMet
GlnSerHisLeuIleProGluLysThrProArgSerIlePheValTyrPheArgGlyLeuPhe
TyrAspValGlyAsnAspProGluGlyGlyTyrTyrAlaArgGlyAlaArgAlaAlaValTrp
GluAsnPheLysAspAsnProLeuPheAspIleSerThrGluHisProThrThrTyrTyrGlu
AspMetGlnArgAlaIlePheCysLeuCysProLeuGlyTrpAlaProTrpSerProArgLeu
ValGluAlaValIlePheGlyCysIleProValIleIleAlaAspAspIleValLeuProPhe
AlaAspAlaIleProTrpGluAspIleGlyValPheValAspGluLysAspValProTyrLeu
AspThrIleLeuThrSerIleProProGluValIleLeuArgLysGlnArgLeuLeuAlaAsn
ProSerMetLysGlnAlaMetLeuPheProGlnProAlaGlnProGlyAspAlaPheHisGln
ValLeuAsnGlyLeuAlaArgLysLeuProHisGluArgSerValTyrLeuArgProGlyGlu
LysLeuLeuAsnTrpThrAlaGlyProValAlaAspLeuLysProTrp***

<210> SEQ ID NO 3
<211> LENGTH, 417

<212> TYPE, PRT
<213> ORGANISM, *Oryza sativa*

<400> SEQUENCE, 3

MetArgArgTrpValLeuAlaIleAlaIleLeuAlaAlaAlaValCysPhePheLeuGlyAla
GlnAlaGlnGluValArgGlnGlyHisGlnThrGluArgIleSerGlySerAlaGlyAspVal
LeuGluAspAspProValGlyArgLeuLysValTyrValTyrAspLeuProSerLysTyrAsn
LysLysLeuLeuLysLysAspProArgCysLeuAsnHisMetPheAlaAlaGluIlePheMet
HisArgPheLeuLeuSerSerAlaValArgThrPheAsnProGluGluAlaAspTrpPheTyr
ThrProValTyrThrThrCysAspLeuThrProSerGlyLeuProLeuProPheLysSerPro
ArgMetMetArgSerAlaIleGluLeuIleAlaThrAsnTrpProTyrTrpAsnArgSerGlu
GlyAlaAspHisPhePheValThrProHisAspPheGlyAlaCysPheHisTyrGlnGluGlu
LysAlaIleGlyArgGlyIleLeuProLeuLeuGlnArgAlaThrLeuValGlnThrPheGly
GlnLysAsnHisValCysLeuLysAspGlySerIleThrIleProProTyrAlaProProGln
LysMetGlnAlaHisLeuIleProProAspThrProArgSerIlePheValTyrPheArgGly
LeuPheTyrAspThrSerAsnAspProGluGlyGlyTyrTyrAlaArgGlyAlaArgAlaSer
ValTrpGluAsnPheLysAsnAsnProLeuPheAspIleSerThrAspHisProProThrTyr
TyrGluAspMetGlnArgSerValPheCysLeuCysProLeuGlyTrpAlaProTrpSerPro
ArgLeuValGluAlaValValPheGlyCysIleProValIleIleAlaAspAspIleValLeu
ProPheAlaAspAlaIleProTrpGluGluIleGlyValPheValAlaGluGluAspValPro
LysLeuAspSerIleLeuThrSerIleProThrAspValIleLeuArgLysGlnArgLeuLeu
AlaAsnProSerMetLysGlnAlaMetLeuPheProGlnProAlaGlnAlaGlyAspAlaPhe
HisGlnIleLeuAsnGlyLeuAlaArgLysLeuProHisGlyGluAsnValPheLeuLysPro
GlyGluArgAlaLeuAsnTrpThrAlaGlyProValGlyAspLeuLysProTrp***

<210> SEQ ID NO 4
<211> LENGTH, 449
<212> TYPE, PRT
<213> ORGANISM, *Physcomitrella patens*

<400> SEQUENCE, 4
MetGluHisProLeuGluCysAlaAspSerCysSerLeuAlaMetSerTrpPheCysAsnLys
LysCysArgGlyTrpGlyLeuMetLysArgThrValValAlaSerGlyLeuArgSerValVal
LeuLeuLeuLeuPheIleTyrPheValGlnAspValThrAlaGluMetGlyHisGlnArgIle
SerGlySerAlaGlyAspValLeuGluAspAsnProValGlyArgLeuLysValPheIleTyr
AspIleProSerLysTyrAsnThrAspTrpLeuLysLysAspProArgCysLeuThrHisMet
PheAlaValGluGluTyrLeuHisAspPheLeuThrGluSerProValArgThrLeuAsnPro
GluGluAlaAspTrpPheTyrThrProValTyrThrThrCysAspLeuThrProAsnGlyLeu
ProLeuProPheLysSerProArgValMetArgSerAlaIleSerTyrIleSerSerHisTrp
ProTyrTrpAsnArgThrAspGlyAlaAspHisPhePheValValProHisAspPheAlaAla
CysPheHisTyrGlnGluGluLysAlaIleGluArgGlyIleLeuProLeuLeuLysArgAla
ThrLeuIleGlnThrPheGlyGlnAsnHisHisValCysLeuLysGluAspSerIleValIle
ProProTyrAlaProProGluArgMetGlnThrArgLeuAsnProProSerThrProArgSer
IlePheAlaTyrPheArgGlyLeuPheTyrAspProGlyAsnAspProGluGlyGlyTyrTyr

AlaArgGlyAlaArgAlaAlaIleTrpGluAsnPheLysAspAsnProLeuPheAspIleSer
ThrGluHisProAlaThrTyrTyrGluAspMetGlnArgAlaIlePheCysLeuCysProLeu
GlyTrpAlaProTrpSerProArgLeuValGluGlyValIlePheGlyCysIleProValIle
IleAlaAspAspIleValLeuProPheAlaAspAlaIleProTrpGluLysIleGlyValPhe
ValGluGluLysAspValProIleLeuAspLysIleLeuCysThrIleAsnHisGluGluVal
LeuGluLysGlnArgLeuLeuAlaAsnProAlaMetLysGlnAlaMetLeuPheProArgPro
AlaLysProGlyAspAlaPheHisGlnIleLeuAsnGlyLeuAlaArgLysLeuProHisAsp
ProSerIleTyrLeuGlnProGlyGlnSerPheLeuAsnTrpThrGluGlyProProGlyAsp
LeuTyrProTrpGlyAsnAspLeu***

<210> SEQ ID NO 5
<211> LENGTH, 846
<212> TYPE, PRT
<213> ORGANISM, *Homo sapiens*

<400> SEQUENCE, 5
MetGlnAlaLysLysArgTyrPheIleLeuLeuSerAlaGlySerCysLeuAlaLeuLeuPhe
TyrPheGlyGlyLeuGlnPheArgAlaSerArgSerHisSerArgArgGluGluHisSerGly
ArgAsnGlyLeuHisHisProSerProAspHisPheTrpProArgPheProAspAlaLeuArg
ProPheValProTrpAspGlnLeuGluAsnGluAspSerSerValHisIleSerProArgGln
LysArgAspAlaAsnSerSerIleTyrLysGlyLysLysCysArgMetGluSerCysPheAsp
PheThrLeuCysLysLysAsnGlyPheLysValTyrValTyrProGlnGlnLysGlyGluLys
IleAlaGluSerTyrGlnAsnIleLeuAlaAlaIleGluGlySerArgPheTyrThrSerAsp
ProSerGlnAlaCysLeuPheValLeuSerLeuAspThrLeuAspArgAspGlnLeuSerPro
GlnTyrValHisAsnLeuArgSerLysValGlnSerLeuHisLeuTrpAsnAsnGlyArgAsn
HisLeuIlePheAsnLeuTyrSerGlyThrTrpProAspTyrThrGluAspValGlyPheAsp
IleGlyGlnAlaMetLeuAlaLysAlaSerIleSerThrGluAsnPheArgProAsnPheAsp
ValSerIleProLeuPheSerLysAspHisProArgThrGlyGlyGluArgGlyPheLeuLys
PheAsnThrIleProProLeuArgLysTyrMetLeuValPheLysGlyLysArgTyrLeuThr
GlyIleGlySerAspThrArgAsnAlaLeuTyrHisValHisAsnGlyGluAspValValLeu
LeuThrThrCysLysHisGlyLysAspTrpGlnLysHisLysAspSerArgCysAspArgAsp
AsnThrGluTyrGluLysTyrAspTyrArgGluMetLeuHisAsnAlaThrPheCysLeuVal
ProArgGlyArgArgLeuGlySerPheArgPheLeuGluAlaLeuGlnAlaAlaCysValPro
ValMetLeuSerAsnGlyTrpGluLeuProPheSerGluValIleAsnTrpAsnGlnAlaAla
ValIleGlyAspGluArgLeuLeuLeuGlnIleProSerThrIleArgSerIleHisGlnAsp
LysIleLeuAlaLeuArgGlnGlnThrGlnPheLeuTrpGluAlaTyrPheSerSerValGlu
LysIleValLeuThrThrLeuGluIleIleGlnAspArgIlePheLysHisIleSerArgAsn
SerLeuIleTrpAsnLysHisProGlyGlyLeuPheValLeuProGlnTyrSerSerTyrLeu
GlyAspPheProTyrTyrTyrAlaAsnLeuGlyLeuLysProProSerLysPheThrAlaVal
IleHisAlaValThrProLeuValSerGlnSerGlnProValLeuLysLeuLeuValAlaAla
AlaLysSerGlnTyrCysAlaGlnIleIleValLeuTrpAsnCysAspLysProLeuProAla
LysHisArgTrpProAlaThrAlaValProValValValIleGluGlyGluSerLysValMet
SerSerArgPheLeuProTyrAspAsnIleIleThrAspAlaValLeuSerLeuAspGluAsp
ThrValLeuSerThrThrGluValAspPheAlaPheThrValTrpGlnSerPheProGluArg
IleValGlyTyrProAlaArgSerHisPheTrpAspAsnSerLysGluArgTrpGlyTyrThr
SerLysTrpThrAsnAspTyrSerMetValLeuThrGlyAlaAlaIleTyrHisLysTyrTyr

HisTyrLeuTyrSerHisTyrLeuProAlaSerLeuLysAsnMetValAspGlnLeuAlaAsn
CysGluAspIleLeuMetAsnPheLeuValSerAlaValThrLysLeuProProIleLysVal
ThrGlnLysLysGlnTyrLysGluThrMetMetGlyGlnThrSerArgAlaSerArgTrpAla
AspProAspHisPheAlaGlnArgGlnSerCysMetAsnThrPheAlaSerTrpPheGlyTyr
MetProLeuIleHisSerGlnMetArgLeuAspProValLeuPheLysAspGlnValSerIle
LeuArgLysLysTyrArgAspIleGluArgLeu***

What is claimed is:

1. A polypeptide capable of dominant suppression of a first naturally occurring IRX10, wherein the polypeptide comprises an amino acid sequence having at least 70% identity as compared to a second naturally occurring IRX10 wherein the polypeptide comprises one or more of the conserved amino acid indicated in FIG. **2** substituted with a different amino acid residue.

2. A nucleic acid encoding the polypeptide of claim **1.** A host cell comprising the polypeptide of claim **1.**

3. A host cell comprising the nucleic acid of claim **2.**

4. A plant comprising the polypeptide of claim **1.**

5. A method of reducing xylan biosynthesis in a plant, comprising: (a) introducing a nucleic acid of claim **2** into a plant, and (b) culturing or goring the plant such that the nucleic acid expresses the polypeptide such that biosynthesis of xylan by the plant is reduced compared to the plant if the nucleic acid was not introduced the plant.

6. A method of reducing xylan biosynthesis in a plant, comprising: (a) introducing a nucleic acid of claim 2 into a plant, and (b) culturing or goring the plant such that the nucleic acid expresses the polypeptide such that the nucleic acid expresses the polypeptide such that biosynthesis of xylan by the plant is reduced compared to the plant if the nucleic acid was not introduced into the plant.

* * *

The following chapter is an Invited Review manuscript for publication in Frontiers in Plant Science: Plant Biotechnology as a part of the "Biofuels and Bioenergy" Research Topic
Chapter 3: Engineering of Bioenergy Crops: Dominant Genetic Approaches to Improve Biomass Properties and Composition

Andrew G. Brandon[1,2,3] and Henrik V. Scheller[1,2,3*]
[1] Department of Plant and Microbial Biology, University of California, Berkeley, CA, USA
[2] Joint BioEnergy Institute, Feedstocks Division, Emeryville, CA, USA
[3] Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

1 Introduction

Lignocellulosic plant biomass represents the largest renewable source of organic carbon on earth. Organic carbon that can be converted into a wide variety of compounds, including high-energy liquid fuel, thereby curbing our dependence on non-renewable sources and limiting the net production of carbon dioxide. The bulk of plant biomass is contained in the cell wall, specifically the thick secondary cell walls (SCW) of the vasculature and fiber tissues. The cell wall has evolved a highly complex and rigid structure to resist the mechanical forces of growth and protect the plant from various stresses. The most valuable component of the cell wall, from a bioenergy perspective, is the 6-carbon sugar glucose comprising the linear polysaccharide cellulose. Produced and extruded into the developing cell wall by cellulose synthase (CesA) complexes at the plasma membrane, individual cellulose chains coalesce through hydrogen bonding to form crystalline cellulose microfibrils. In all vascular plants, the SCW is composed of cellulose microfibrils embedded in a matrix of the aromatic polymer lignin. Hemicellulose like xylan coat the cellulose microfibrils and can form covalent linkages with cell wall proteins, lignin, and other hemicelluloses (Gírio et al., 2010; Meents et al., 2018; Scheller and Ulvskov, 2010). This natural complexity makes the deconstruction and recovery of useable sugars costly and resource-intensive.

The prevailing sources of bioenergy to date have been sucrose from crops like sugarcane or sugar beet and starch from corn. While much simpler and cheaper to process, their sustainability at a larger scale is dubious considering all are also major food and forage crops. The "food vs. fuel" competition for arable land could drive up the price of food and be socioeconomically detrimental. Therefore, modern approaches to sustainable bioenergy emphasize the development of dedicated bioenergy crops that can be grown on marginal land (Cai et al., 2011; Himmel and Bayer, 2009). Ideally, these crops will be fast-growing perennials, producing the maximum biomass per unit land over multi-year cycles and minimizing nutrient input needs (Sanderson and Adler, 2008). The additional constraints of varying soil quality, water availability, and average temperature mean biomass productivity will vary depending on where the crops are grown and that no single engineered species can meet bioenergy demands around the world (Chen and Peng, 2013; Somerville et al., 2010). For example, in tropical and sub-tropical climates, elephant or Napier grass (*Pennisetum purpureum*) produces more biomass per hectare annually than any other vegetation. Grasses like *Miscanthus x giganteus*, sorghum, and switchgrass, and tree species like poplar, aspen, and willow are capable of producing large

amounts of biomass in temperate regions like Europe and the USA (Guidi et al., 2013; Heaton et al., 2008; Junior et al., 2016). While the massive bioethanol productivity of Brazil comes from sugarcane sucrose, the lignocellulosic biomass left over could additionally be engineered for better bioenergy conversion. Crassulacean acid metabolism plants like *Agave spp*. have the highest water use efficiency of all plants, making them attractive bioenergy feedstock crops for cultivation on the 18% of the world's land area considered arid (<800 mm of rainfall per year) which is likely to increase as global temperatures rise (Borland et al., 2009; United Nations Environment Programme, 2007).

The wide variety of potential feedstock species means efforts to improve their quality through biotechnology should be as broadly useful as possible. Research into the fundamental cell biology of plants and the organisms that degrade them has revealed the causes of biomass recalcitrance and a variety of approaches to reduce it. Significant improvements in biomass have been demonstrated by suppressing or eliminating the expression of various genes related to the biosynthesis of specific polysaccharides, primarily in model species (Bhatia et al., 2017; Donev et al., 2018; Loqué et al., 2015; Wang et al., 2016). However, the usefulness of the same approaches in bioenergy crops can be limited by their comparative genetic complexity. Therefore, this review will focus on cell wall polysaccharide engineering strategies that act independent of genetic or genomic context, primarily via the overexpression of recombinant or native enzymes. Importantly, we include only enzymes that are active *in planta* concurrent with cell wall development. The engineering of plants to accumulate hyperthermophilic or inactive cell wall degrading enzymes has been well reviewed is other works (Damm et al., 2016; Mir et al., 2014; Park et al., 2016) .

2 Modulation of Polysaccharide Biosynthesis

2.1 Cellulose

Two major features of an "ideal" bioenergy crop are high cellulose content and a high ratio of C6:C5 sugars comprising the polysaccharide content. Increasing cellulose biosynthesis is an important goal for engineering efforts because cellulose consists entirely of the C6 sugar glucose and the genes involved in its biosynthesis are relatively well-understood and conserved among land plants. Secondary cell wall cellulose, which accounts for the bulk of cellulosic biomass in bioenergy-relevant crops, is synthesized at the plasma membrane by three, non-redundant cellulose synthase (CesA) proteins CesA4, CesA7 and CesA8 (McFarlane et al., 2014). Overexpression of CesAs is, therefore, a logical approach to generating transgenic plants enriched in cellulose. However, attempts to overexpress SCW CesAs in aspen and barley resulted in co-suppression and decreased cellulose content (Joshi et al., 2011; Tan et al., 2015). Greater success has been demonstrated in Arabidopsis (*Arabidopsis thaliana*) by overexpression of either of the primary cell wall (PCW) CesAs, *CesA2, CesA5* and *CesA6* (Hu et al., 2018). CesA2, CesA5 and CesA6 are each functional in a PCW cellulose synthase complex including CesA1 and CesA3. Transgenic plants overexpressing one of the three genes had 29-37% increase in crystalline cellulose content. Expression of both *CesA1* and *CesA3* was significantly higher in transgenic lines, indicating the possible secretion and activity of PCW CesA complexes even during SCW development. Transgenic lines also had more xylan and a slight, but significant, increase in lignin, potentially counteracting the benefits of the increase in cellulose. However, enzymatic saccharification efficiency of the transgenic biomass was not reported and further

work is required to determine the usefulness of PCW *CesA*-overexpression. A clever approach to increase SCW cellulose content that avoids the negative effects of co-suppression has been employed to generate transgenic sugarcane (*Saccharum spp.*). While plants, many bacteria and some fungi can produce cellulose, only one group of animals is known to do so: the marine invertebrates of subphylum *Tunicata*, otherwise known as sea squirts (Kimura and Itoh, 2007; Matthysse et al., 2004). A cellulose synthase cDNA from *Ciona savignyi* (*CsCesA*) was used to create transgenic sugarcane overexpressing a functional form of the protein (Ndimande, 2013). Use of this divergent gene sequence did not cause any co-suppression and the internode cellulose content was increased by up to 31%. Additionally, all tissues of *CsCesA*-overexpressing sugarcane lines had increased saccharification efficiency, with increases of 39% and 28% in young and mature internodes, respectively. Since sucrose is the primary source of bioenergy potential in sugarcane, total soluble sugars were also measured. Transgenic lines yielded up to 25% more, contrary to intuition that sucrose content would be depleted by increased conversion to UDP-glucose (UDP-Glc) for cellulose production. It could be that this depletion acts as a signal to source organs to increase production and/or transport of sucrose to sink organs. It has previously been demonstrated that induced depletion of sucrose from sugarcane stems increases photosynthetic productivity and phloem loading from leaves (McCormick et al., 2009; Wang et al., 2013). In the non-cellulosic polysaccharide fraction of the cell wall, Ndimande (2013) also observed increased glucose (<56%), galacturonic acid (<53%) and galactose (<22%). The author attributes the increase in glucose to increased biosynthesis of mixed-linkage glucan, which uses UDP-Glc as its substrate. An increased mixed-linkage glucan deposition driven by higher UDP-Glc has previously been reported in barley starch mutants (Christensen and Scheller, 2012). While sugarcane accumulates large amounts of soluble sucrose, the majority of the carbohydrate mass of the plant is still the cell wall, making it an interesting model for the study of carbon flux to cell wall biosynthesis.

Increased cellulose biosynthesis has been accomplished by overexpressing the enzyme responsible for producing its substrate. Sucrose synthase (*SuSy*) proteins catalyze the cleavage of sucrose to fructose and UDP-Glc, which is the sole substrate for the biosynthesis of glucans like cellulose, callose and mixed-linkage glucans. Evidence suggests that some *SuSy* isoforms interact directly with the *CesA* complex, channeling UDP-Glc directly to cellulose biosynthesis (Fig. 1) (Fujii et al., 2010; Stein and Granot, 2019). Overexpression of the hybrid poplar (*Populus simonii* × *Populus nigra)* gene *PsnSuSy2* in tobacco led to 25% thicker cell walls, containing up to 18% more cellulose and decrease in lignin of up to 28%, when compared to controls. Additionally, the degree of cellulose crystallinity (CrI) was reduced by 9-11% (Li et al., 2019; Wei et al., 2015). CrI reflects the degree of hydrogen bonding between individual cellulose chains and is a primary contributor to biomass recalcitrance by reducing the proportion of cellulose exposed to cellulolytic enzymes (Hall et al., 2010). A similar approach in transgenic rice (*Oryza sativa*) lines overexpressing *OsSuSy3* under a secondary cell wall-specific or a constitutive promoter and dramatic improvements in multiple bioenergy-relevant characteristics were observed (Fan et al., 2017, 2019). The effects on cell wall composition were similar using either promoter. Total plant biomass was only slightly increased in transgenic lines, but microscopic analysis of the cell wall revealed a 68% increase in cell wall thickness compared to controls. The increase in cell wall thickness corresponded to a 15-26% increase in cellulose and 11-13% increase in hemicelluloses, while lignin content was not significantly changed. Cellulose from transgenic plants had a 7-10% reduction in CrI, but an increased cellulose degree of polymerization (DP) by 8-15%. DP, like CrI, is negatively correlated with saccharification

efficiency. After pretreatment, transgenic lines yielded 13-23% total sugars, leading to a 20-49% greater ethanol yield and 22% higher conversion efficiency when compared to controls. The 2017 and 2019 studies distinguish themselves by investigating secondary effects of *OsSuSy3* overexpression on two important agronomic traits, respectively: lodging resistance and susceptibility to pathogen or insect attack. Lodging is a complex trait and crop susceptibility to lodging can dramatically reduce yield and increase the cost of harvesting. Lodging Index is a combination of several measurements taken from the plant used to determine its physicomechanical strength and resistance to lodging, All four independent transgenic lines overexpressing *OsSuSy3* showed increases of 17-50% in Lodging Index compared to controls, suggesting they could grow robustly and be high-yielding in the field. In the more recent study by Fan and coworkers, *OsSuSy3* transgenic lines were observed to be less susceptible to a number of biotic stresses including bacterial blight, fungal rice blast and herbivory by the brown planthopper. These resistances correlated with a significant increase in callose deposition upon initiation of infection or pest attack. Since both callose and cellulose are synthesized from UDP-Glc, it is reasonable to expect that *OsSuSy3* overexpression could also stimulate callose biosynthesis (Fig. 1). The β-(1,3)-Glc bonds in callose make it much less crystalline than cellulose, thus more amenable to saccharification. It would be intriguing to see if this accumulation of callose phenotype could be exploited, perhaps by triggering an immune response at senescence in order to rapidly accumulate low-recalcitrance callose before harvest. These works demonstrate that overexpression of a single gene involved in carbohydrate flux can simultaneously improve a variety of traits important in bioenergy crops.

Several genes that are important for, or even critical to, cellulose biosynthesis have been identified based on the phenotypes of their loss-of-function mutants. Their gene products often lack structural similarity to CesA proteins or even to glycosyltransferases in general, so the roles they play in cellulose biosynthesis can be difficult to determine. Despite this, a few studies have demonstrated that the overexpression of such proteins can lead to dramatic improvements to biomass composition and recalcitrance. A specific example is the engineered overexpression of a gene, *Domain of Unknown Function-266A (DUF266A),* in poplar (*Populus deltoides)* and Arabidopsis (Yang et al., 2017). DUF266-containing proteins are only present in land plants and are categorized as "non-classifiable GT" (Lao et al., 2014), although they are distantly related to GT family 14. GT14s have been characterized as having a range of activities, from arabinogalactan biosynthesis in plants to protein *O*-glycosylation in animals (Cantarel et al., 2009; Hansen et al., 2012; Knoch et al., 2013) The rice gene *Brittle Culm 10 (OsBC10)* is the only previously characterized gene encoding a DUF266-containing protein. *OsBC10* encodes a Type II transmembrane Golgi protein and loss-of-function mutant plants were dwarfed with reductions in cellulose, and increases in xylose and lignin, indicating a role in cell wall biosynthesis (Zhou et al., 2009). The transgenic poplar lines overexpressing *PdDUF266A* generated by Yang and co-workers had increased total biomass of 17-34% and cellulose content up to 37% greater than wild type. Their cellulose DP was also increased by 13%, although the CrI was not significantly altered. Total sugar release after enzymatic saccharification was increased in *PdDUF266A*-overexpressors by 38% compared to controls. These effects could be indirect, since several genes involved with SCW cellulose biosynthesis genes were found to be significantly upregulated. How a protein without a predicted function, residing in the Golgi, is able to alter gene expression can only be guessed at. However, there is a more direct effect *PdDUF266A* overexpression may have had on cell wall composition (Yang et al., 2017). When probing transverse section of the *Osbc10* mutant, Zhou et al. (2009) found a dramatic decrease in

signal using antibodies specific to arabinogalactan proteins. Compositional analysis revealed *Osbc10* mutants to have a 72% reduction in cell wall arabinogalactan proteins compared to controls. While it was not measured in the study, it is possible that overexpression of *PdDUF266A* modified the arabinogalactan protein profile of the cell wall, which could cause alterations in the organization or interaction of various wall polysaccharides. PdDUF266A could also be involved directly in the glycosylation of CesA proteins or other proteins critical to cellulose synthesis like KORRIGAN and COBRA. The latter two proteins play roles in microfibril deposition and require *N*-glycosylation to be fully functional (Liebminger et al., 2013; Roudier et al., 2005). Recently, a *COBRA-like* gene from cotton (*Gossypium hirsutum*), *GhCOBL9A*, was overexpressed in Arabidopsis, leading to dramatic increases in total biomass and cellulose content (Niu et al., 2018). The cell walls of cotton fibers consist almost entirely of cellulose and are an interesting model for high-level cellulose production. Of the 33 identified *COBL* genes in the cotton genome, *COBL9A* was expressed highly during SCW development and co-expressed with SCW *CesA* genes. COBL proteins are secreted and have a glycosylphosphatidylinositol anchor to the plasma membrane. COBLs also have a carbohydrate-binding domain (CBM) that preferentially binds to crystalline cellulose. COBLs are required for cellulose production, believed to direct the orderly deposition of nascent cellulose microfibrils but are not components of the cellulose synthase complex (Liu et al., 2013). Niu et al. (2018) found that overexpression of *GhCOBL9A* in Arabidopsis produced plants that grew taller and contained up to 59% more cellulose. The expressions of three SCW *CesA* genes, *CesA4*, *CesA7* and *CesA8*, were also dramatically increased, approximately 10-fold higher than controls. Interestingly, examination of transverse stem sections of transgenic plants revealed increased cellulose and wall thickening, not only in fiber and vessel cells, but also in pith and parenchyma cells. These tissues normally contain only cells with a thin primary wall. Increasing cellulose production in these cell types or even engineering them to deposit a SCW would be an interesting strategy with potential to increase total cellulose content in bioenergy crops.

2.2 Hemicelluloses and Pectin

Although it is by far the most abundant, cellulose is not the only C6-sugar polysaccharide in the cell wall. Co-overexpression of *Galactan Synthase1* (*GalS1*) with the enzyme that provides its substrate, *UDP-Glc/UDP-Gal-4-Epimerase2* (*UGE2*), in Arabidopsis increased galactose content in the cell wall by up to 80% (Gondolf et al., 2014). GalS1 transfers galactose to β-1,4-galactan side chains of the rhamnogalacturonan I (RGI) backbone domain of pectin. Overexpression of *UDP-Rha/UDP-Gal Transporter1 (URGT1*), responsible for transport of UDP-Gal from the cytoplasm to the Golgi, in combination with *GalS1* and *UGE2* further boosted galactose content in the stems to four times the levels in wild-type plants (Fig. 1) (Aznar et al., 2018). Another hexose polysaccharide found in the cell walls of grasses is mixed-linkage glucan. The accumulation of significant amounts of mixed-linkage glucans has been successfully engineered in Arabidopsis through overexpression of *Cellulose synthase-like F6* from rice *(OsCslF6)* (Vega-Sánchez et al., 2015).The cell walls of OsCslF6-overexpressor lines contained four time more non-cellulosic glucose. Additionally, the saccharification efficiency of plants producing mixed-linkage glucans was increased by 42% compared to wild type. The variety of linkages between glucose residues in mixed-linkage glucans make it much more amorphous and soluble than cellulose, and thus more amenable to saccharification. The greater solubility of mixed-linkage glucan means it is easily extracted from the biomass post-harvest which in turn

exposes more of the cellulose surface to hydrolytic enzymes. While high-level production of mixed-linkage glucans is, therefore, an attractive trait to engineer into bioenergy crops, the choice of promoter appears to be critical. Overexpression of *CslF6* in barley, Arabidopsis and tobacco with constitutive or SCW-specific promoters had severely adverse effects on plant growth (Burton et al., 2011; Vega-Sánchez et al., 2015). The successful outcomes of the study by Vega-Sanchez and coworkers depended on the use of a promoter active only during senescence.

In conjunction with increasing C6 sugars like glucose and galactose, a reduction in C5 sugars like xylose is also desirable in the cell walls of dedicated bioenergy crops. Mutants in xylan biosynthesis have been identified in model species, but genetic redundancy and greater genomic complexity in bioenergy crops make full knockouts or knockdowns difficult to generate (Brown et al., 2007, 2009; Chen et al., 2013; Lee et al., 2012; Mortimer et al., 2015; Wu et al., 2010). Recent work from our lab has successfully demonstrated a novel approach to reducing the amount of a specific polysaccharide with a protein-level, dominant knock-down of xylan biosynthesis (Brandon et al., 2019). The gene *Irregular Xylem 10 (IRX10)* and its partially redundant homolog *Irregular Xylem 10-like (IRX10-L)* encode GT47 enzymes that unambiguously exhibit xylan β-(1,4)-xylosyltransferase activity in recombinant systems (Jensen et al., 2014; Urbanowicz et al., 2014). However, other proteins, IRX9 and IRX14 in particular, play critical, likely structural roles in the functional xylan synthase complex (Ren et al., 2014; Wu et al., 2010). As of this publication, no crystal structure of a GT47 protein has been resolved, so the catalytic site of IRX10 is unknown. However, potentially important amino acids residues can be inferred to be involved in catalysis based on their very high degree of conservation in evolutionarily divergent species. Two of these residues, Gly-283 and Glu-293 drastically reduced or eliminated enzymatic activity when mutated. By overexpressing the mutated *IRX10* (*dnIRX10*) genes in wild type Arabidopsis, the mutant isoforms ($IRX10^{G283D}$ and $IRX10^{E293Q}$) out-competed the native IRX10 for its place in the proposed xylan synthase complex. The stems of *dnIRX10*-overexpressing plants had reductions in xylose content of up to 55% compared to wild type. In spite of its detrimental effects on recalcitrance and biomass conversion, xylan is critical to the strength of vessel and fiber SCWs, and mutants in xylan biosynthesis exhibit severe growth defects due primarily to collapsed xylem vessels and impaired water and nutrient transport (Brown et al., 2005, 2009; Lee et al., 2007; Wu et al., 2010). The phenotypes of *dnIRX10*-overexpressing lines, unsurprisingly, mimicked those of xylan knockout mutants. Previous work in our lab has demonstrated that the growth phenotype of xylan biosynthetic mutants can be rescued by expression of a functional copy of the gene under the control of a vessel cell-specific promoter (Petersen et al., 2012). Therefore, much of the xylose reduction could be maintained without growth penalty if expression of the transgene was abrogated in vessel cells or if a strong, fiber-specific promoter were used.

3 Polysaccharide Modification

3.1 Cellulose

Biomass quality can be improved by transgenic expression of enzymes that modify the polysaccharides of the cell wall prior to its full maturation. Plants express various glycosylhydrolases (GH) for building and remodeling the wall in many different tissues and stages of development (Barnes and Anderson, 2018). The GH9 β-1,4-endoglucanases in plants

are hypothesized to play a role in cellulose remodeling and biosynthesis by cutting specifically between glucose residues of a cellulose chain. The GH9 subfamily B is distinguished the absence of either a transmembrane domain or CBD (Hayashi et al., 2005). Recent work in rice has demonstrated that overexpression of two native GH9Bs (*OsGH9B1* and *OsGH9B3*) dramatically improved biomass quality without significant growth or developmental defects (Huang et al., 2019). The cell wall composition unchanged compared to the control and plant grew normally. However, the transgenic lines exhibited a 18-23% decrease in cellulose DP and an 11-23% reduction in CrI. After pretreatment and enzymatic saccharification of the biomass, both *GH9B1* and *GH9B3*-overexpressors released 63% more reducing sugars than control lines. Since total cellulose content in transgenic lines was unchanged, the authors posit that the enzymes directly decreased DP and CrI by cleaving microfibrils, thereby increasing accessibility to cellulases number of exposed cellulose ends (Fig. 2A).

Genes encoding GH enzymes from lignocellulose-degrading fungi can also be used to engineer bioenergy crops for heterologous expression. Work using the thermophilic endoglucanase gene from *Acidothermus cellulolyticus AcCel5A* suggests that even minimal cuts in cellulose chains can have dramatic effects on microfibril crystallinity, increasing enzymatic digestibility and sugar yield (Brunecky et al., 2011; Donohoe et al., 2017). AcCel5A is known to have high activity at high temperatures on synthetic substrates *in vitro*, but to be ineffective at degrading mature cell walls. Theorizing that *in planta* expression might be more effective, transgenic lines of maize and tobacco producing apoplast-targeted AcCel5A were generated. The high temperature optimum of the enzyme was a deliberate and important consideration, as it suggested activity would be low enough to avoid deleterious effects on normal plant growth. Indeed, both tobacco and maize plants grew normally and were less recalcitrant to bioconversion, yielding 10-15% more glucose from cellulose than untransformed plants under the same conditions (Brunecky et al., 2011). Importantly, adding recombinant AcCel5A to post-harvest cell wall material could not replicate these results, indicating that enzyme activity concurrent with cellulose production and deposition is key. In order to better understand the role of AcCel5A and build on prior work in maize and tobacco, the same endoglucanase gene was transformed into Arabidopsis to engineer overexpressor lines (Fig. 2B) (Donohoe et al., 2017). As in previous experiments using maize and tobacco, the composition of the cell wall was unchanged in transgenic plants and they grew normally. However, after closer inspection of the cell wall with scanning electron microscopy, large voids, pockets and other structural irregularities were observed in cell walls of *AcCel5A*-overexpressors.. Interestingly, these features mimicked the characteristics of electron micrographs of plant biomass after various chemical pretreatments. Thus, *in planta* expression of *AcCel5A* is able to mimic the role of pretreatment to increase the cellulose surface area accessible to hydrolytic enzymes. This presents the obvious benefit of likely reducing the chemical and enzyme input necessary downstream.

The choice of endoglucanase and the organism it is derived from is important. Overexpression and apoplast-targeting in tobacco of a similar GH from the mesophilic fungus *Trichoderma reesei*, TrCel5A, caused severe growth defects and a significant decrease in cellulose content (Klose et al., 2015). Presumably, TrCel5A would be more active than AcCel5A since *T. reesei* evolved in similar conditions to plants. Additionally, TrCel5A possesses a CBM that was truncated from the AcCel5A used by Brunecky and coworkers. CBMs are known to facilitate hydrolytic activity and reduce CrI by physically disrupting the hydrogen bonds between cellulose chains (Fig. 2C) (Abramson et al., 2010). This disruptive effect could be exploited for

the improvement of bioenergy crops, since CrI is strongly correlated with biomass recalcitrance. However, relatively few studies have been published exploring the engineered overexpression of CBMs, or CBM-containing proteins like expansins, to this end. Expansins are cell wall proteins that disrupt the intermolecular hydrogen bonds of cellulose and hemicelluloses, promoting the flexibility and extensibility of the primary cell wall. They play critical roles primary growth by maintaining the delicate balance between turgor pressure and cell wall integrity that drive cell expansion (Cosgrove, 2000, 2005). However, some expansins are specifically expressed in vessel and fiber cells during SCW development. Two such expansin genes from Chinese fir (*Cunninghamia lanceolata)* have been cloned and used to engineer overexpression lines in tobacco (Wang et al., 2011) Both *Expansin-A1 and Expansin-A2 (ClEXPA1 and ClEXPA2)* overexpressing lines grew taller and had thicker stems than wild type. The cell walls of xylem cells were 1.13 to 1.45 times thicker in transgenic plants and when the composition of stem cell wall material was analyzed, they contained 30-50% more cellulose. It is also interesting to explore the effects that CBDs alone can have on cell wall architecture. In Arabidopsis, the coding sequence of the CBD of *Starch Synthase III (SSIII)* was overexpressed and targeted to the cell wall (Grisolia et al., 2017). Though starch and cellulose are very different in structure, previous work indicated that the concatenated triplicate of CBDs (collectively referred to as SBD123) from SSIII had preferential affinity for the linear portions of the starch molecule. Thus, the rationale behind the work was that it may also bind to linear cellulose and modify the crystallinity of microfibrils. Like the tobacco plants expressing *ClEXPA* genes, Arabidopsis plants overexpressing *SBD123* grew significantly larger than untransformed lines. The average cell area was increased by 40% and dry biomass weight by 76%. Unlike the *ClEXPA*-expressing plants, Grisolia and co-workers observed a 27% reduction in cell wall thickness and similar amounts of cellulose in transgenic Arabidopsis stems. This suggests SBD123 loosened the components of the cell wall, stretching the cell wall to cover a greater cell volume. However, hemicellulose and pectin contents were significantly higher (50% and 30%, respectively) and dilute acid hydrolysis of cell wall material from transgenic plants yielded almost twice as much glucose. Dilute acid is generally insufficient to hydrolyze crystalline cellulose and there is relatively little xyloglucan in Arabidopsis stems. Thus, it is reasonable to conclude that the glucose released is derived from crystalline cellulose made amorphous by the disruptive effects of SBD123-cellulose interaction. Finally, an *in vitro* rumen digestability assay determined biomass from transgenic plants was 28% easily digested than that of control plants. However, other attempts to engineer plants expressing CBMs have been less successful.

3.2 Xylan

Xylan is the most abundant hemicellulose in most bioenergy crops and contributes significantly to biomass recalcitrance by enveloping cellulose microfibrils and forming covalent linkages to lignin and other hemicelluloses in the cell wall. Thus, modifications to xylan are important engineering goals in the development of dedicated bioenergy crops. The xylan backbone is β-(1,4)-linked xylose residues, many of which can be mono- or di-acetylated, and is decorated to varying degrees with glucuronic acid (GlcA), 4-*O*-methyl glucuronic acid (MeGlcA) and, in grasses, arabinose (Ara) side chains (Rennie and Scheller, 2014). Arabinose side chains can be further modified by esterification to ferulic acid (FA) or *p*-coumaric acid (*p*CA) moieties. The amount and distribution of these decorations to the xylan backbone

determine the properties of the polysaccharide *in muro*. Several engineering strategies targeting these sides chains have been successfully employed to reduce biomass recalcitrance.

Approximately 40-60% of xylose residues comprising xylan are acetylated at the O-2 or O-3 position (Busse-Wicher et al., 2014), and reductions in acetylation correlate positively with conversion efficiency. Additionally, acetic acid released from xylan during pretreatment is a strong inhibitor of microbial fermentation downstream. To engineer a reduction in xylan acetylation, an acetyl xylan esterase from the lignocellulose-degrading fungus *Aspergillus niger* (*AnAXE1*) was overexpressed in Arabidopsis (Fig. 3) (Pawar et al., 2016). After observing beneficial effects, the same gene was used to generate transgenic overexpressors in hybrid aspen (*Populus tremula L. x tremuloides Michx*) (Pawar et al., 2017). The transgenic trees developed normally with approximately 10% less xylan acetylation. This reduction in acetylation had the secondary effect of reducing the average xylan DP. Since no changes in xylan biosynthesis were observed, the authors propose that xylan with reduced acetylation may be more susceptible to cleavage by endogenous cell wall hydrolases. These modifications to xylan led to a 26% increase in saccharification efficiency compared to controls without pretreatment. The positive effect of *AnAXE1*-overexpression was confirmed after acid pretreatment of the biomass. Acid pretreatment removed xylan, and the difference in saccharification efficiency between transgenic and control lines was greatly reduced. The effects of xylan acetylation on recalcitrance, however, are not yet fully understood. The overexpression of a gene encoding a DUF231-containing protein in poplar (*PdDUF231A*) had beneficial effects on biomass composition, despite increasing xylan acetylation by 8% (Yang et al. 2017). It is likely that the effect on xylan acetylation is an indirect one, though, since the PdDUF231A is phylogenetically more closely related to PMR5, a pectin acetyltransferase (Chiniquy et al., 2019), than to the known xylan acetyltransferases. *PdDUF231A*-overexpressing lines exhibited an 8-21% increase in total cellulose content and 6-8% reduction in lignin, while cell wall content of other cell wall sugars was unchanged. Metabolite analysis of the transgenic lines additionally suggested the increase in carbon flux to cellulose biosynthesis came at the expense of lignin biosynthesis. While the precise role of *PdDUF231A* is unknown, the combination of beneficial biomass changes it produced makes it an interesting subject of further research.

The xylans of grasses are extensively arabinosylated, allowing the formation of many cross-linkages with other components in the cell wall and directly affect the characteristics of cellulose microfibrils (Li et al., 2015). Reducing the number of xylan arabinose side chains has great potential for improving the biomass quality of grasses, perhaps the most important group of dedicated biomass crops. Two native arabinofuranosidase genes from rice, (*OsARAF1* and *OsARAF3*), were used to engineer rice lines accumulating high levels of the protein in the cell wall (Fig. 3) (Sumiyoshi et al., 2013). Transgenic lines had 19-25% less arabinose in matrix cell-wall polysaccharides and a 28-34% increase in cellulose content. Its likely both factors combined led to a 46-70% increase in enzymatic sugar release compared to wild type. Many xylan arabinose side chains are esterified to ferulic acid, which can form extensive diferulate covalent bonds with other arabinoxylans. Ferulic acid can also act as a nucleation site for the polymerization of lignin (Grabber et al., 2004; Terrett and Dupree, 2019). These properties make ferulic acid a significant contributor to the recalcitrance of grass biomass. Diferulate cross-linking and the free-radical nucleation of lignin happen spontaneously and non-enzymatically. Thus, enzymes severing the connection between ferulic acid and arabinose are particularly useful for biomass engineering. Fortunately, we can again draw from ingenuity of lignocellulose-degrading fungi. *In planta* expression of a variety of ferulic acid esterases (FAEs) has

demonstrated beneficial effects to reduce recalcitrance and increase biomass digestibility, but in many cases there are trade-offs to be considered. Expression of *Aspergillus nidulans* FAE in Arabidopsis and *Brachypodium* enhanced saccharification but increased susceptibility to fungal pathogens (Pogorelko et al., 2011; Reem et al., 2016). In alfalfa (*Medicago sativa*), expression of *A. niger* FAE did reduce arabinoxylan feruloylation, but unexpectedly led to increased lignification and decreased digestibility (Badhan et al., 2014). The authors noted, however, that saccharification efficiency did increase compared to wild type after chemical delignification, possibly due to a reduction of inter-arabinoxylan cross-linking. The beneficial effect of *in planta* expression of FAEs has been particularly well studied in tall fescue (*Festuca arundinacea),* an important animal forage crop (Buanafina et al., 2008, 2010). As fescue is typically studied for its nutritional value to livestock, an *in vitro* model using rumen microorganisms was used to assess cell wall digestability and *AnFAE*-overexpressors demonstrated a 10-14% increase in digestibility. More recent work with fescue suspension culture cells showed that significantly less time was required for rumen microbes to digest the cell wall from cells overexpressing *AnFAE* than for control cells (Morris et al., 2017). Despite the use of very different analytical techniques, it is reasonable to believe these changes to cell wall composition could translate in bioenergy crops to improvements in enzymatic saccharification. Importantly, this increase in digestion rate could not be replicated by exogenous application of recombinant *AnFAE* to control cell wall material. Much like the previously described transgenics expressing endoglucanases, the beneficial effects are most significant when the enzyme is active during cell wall development. The observed negative side effects of FAE expression like increased susceptibility to pathogens could perhaps be mitigated by restricting expression of the transgene in the tissues most affected by the stress, like epidermal cells. Other labs have found success in rice and switchgrass using an alternative approach to reduce ferulate crosslinking in the cell wall by out-competing ferulic acid for arabinosyl side chains on xylan (Bartley et al., 2013; Li et al., 2018). The BAHD-type acyltransferase OsAT10 from rice is a putative *p*-coumaroyl transferase that increases *p*-coumaric acid esterification to arabinose moieties of xylan. Although it is also a phenylpropanoid-derived hydroxycinnamic acid, *p*-coumaric acid does not undergo oxidative coupling to monolignols in the polymerization of lignin (Marcia, 2009). In *OsAT10*-overexpressing lines of rice and switchgrass, the ratio of *p*-coumaroyl to feruloyl esterification of arabinoxylan increased by 150% and 75%, respectively. In rice, *OsAT10* overexpression had the additional effect of increasing cell wall glucose by 8-19% compared to wild type. Transgenic lines of both rice and switchgrass developed normally and both demonstrated an increase of up to 40% in total sugar yield after enzymatic saccharification.

3.3 Xyloglucan and Pectin

While most studies have focused on cellulose and xylan, some studies have targeted other polysaccharides, which are clearly important in the development of CSWs despite their lower abundance. Overexpression of *Aspergillus aculeatus* xyloglucanase in poplar (*Populus alba*) led to increased growth and cellulose deposition and substantially decreased recalcitrance (Kaida et al., 2009; Park et al., 2004). Pectin modification is likewise a useful strategy as shown in the knock-down studies mentioned above (Biswal et al., 2015; Li et al., 2019b). The modification of pectin by overexpression of a pectate lyase gene from *Populus trichocarpa* in hybrid aspen (*Populus tremula* x *tremuloides*) resulted in improved saccharification. The overall composition of the biomass did not differ significantly from wild-type plants, indicating that the enzyme

might act by loosening interactions between matrix wall components and increasing accessibility to cell wall-degrading enzymes. In another study, pectin methylesterase inhibitor (PMEI) was overexpressed in Arabidopsis and wheat, and resulted in improved growth and saccharification (Lionetti et al., 2010). The role of minor matrix polysaccharides in the development and structure of SCWs is poorly understood, but the successful results suggest that this would be a fertile ground for further research.

4 Conclusion

Plant biotechnological research can make a major contribution to meeting perhaps humanities' most common and critical goals. That is, mitigating the effects of climate change by reducing our net production of carbon dioxide and developing an alternative to the fundamentally unsustainable fossil resources on which we've become dependent. While traditional breeding techniques can, and likely will, be important in the development of dedicated bioenergy crops, genetic engineering allows for significant modifications to biomass to be made much more quickly. As our fundamental knowledge of plant cell biology and the regulation of cell wall development continues to grow, so to will our ability to approach the perfect balance of maximizing yield and quality while minimizing recalcitrance and deleterious side effects. Here we have described a number of dominant genetic engineering strategies to improve plant biomass for bioenergy conversion. Though they may have only been demonstrated in one or two species, they have potential for broad usability in the wide variety of dedicated bioenergy species currently being researched.

Bibliography

Abramson, M., Shoseyov, O., and Shani, Z. (2010). Plant cell wall reconstruction toward improved lignocellulosic production and processability. *Plant Science* 178, 61–72. doi:10.1016/j.plantsci.2009.11.003.

Aznar, A., Chalvin, C., Shih, P.M., Maimann, M., Ebert, B., Birdseye, D.S., et al. (2018). Gene stacking of multiple traits for high yield of fermentable sugars in plant biomass. *Biotechnology for Biofuels* 11, 2. doi:10.1186/s13068-017-1007-6.

Badhan, A., Jin, L., Wang, Y., Han, S., Kowalczys, K., Brown, D.C., et al. (2014). Expression of a fungal ferulic acid esterase in alfalfa modifies cell wall digestibility. *Biotechnology for Biofuels* 7, 39. doi:10.1186/1754-6834-7-39.

Barnes, W.J., and Anderson, C.T. (2018). Release, Recycle, Rebuild: Cell-Wall Remodeling, Autodegradation, and Sugar Salvage for New Wall Biosynthesis during Plant Development. *Molecular Plant* 11, 31–46. doi:10.1016/j.molp.2017.08.011.

Bartley, L.E., Peck, M.L., Kim, S.-R., Ebert, B., Manisseri, C., Chiniquy, D.M., et al. (2013). Overexpression of a BAHD acyltransferase, OsAt10, alters rice cell wall hydroxycinnamic acid content and saccharification. *Plant Physiology* 161, 1615–1633. doi:10.1104/pp.112.208694.

Bhatia, R., Gallagher, J.A., Gomez, L.D., and Bosch, M. (2017). Genetic engineering of grass cell wall polysaccharides for biorefining. *Plant Biotechnology Journal* 15, 1071–1092. doi:10.1111/pbi.12764.

Biswal, A.K., Hao, Z., Pattathil, S., Yang, X., Winkeler, K., Collins, C., et al. (2015).

Downregulation of GAUT12 in Populus deltoides by RNA silencing results in reduced recalcitrance, increased growth and reduced xylan and pectin in a woody biofuel feedstock. *Biotechnology for Biofuels* 8, 41. doi:10.1186/s13068-015-0218-y.

Borland, A.M., Griffiths, H., Hartwell, J., and Smith, J.A.C. (2009). Exploiting the potential of plants with crassulacean acid metabolism for bioenergy production on marginal lands. *Journal of Experimental Botany* 60, 2879–2896. doi:10.1093/jxb/erp118.

Brandon, A.G., Birdseye, D.S., and Scheller, H.V. (2019). A dominant negative approach to reduce xylan in plants. *Plant Biotechnology Journal*. doi:10.1111/pbi.13198.

Brown, D.M., Goubet, F., Wong, V.W., Goodacre, R., Stephens, E., Dupree, P., et al. (2007). Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. *The Plant Journal: for Cell and Molecular Biology* 52, 1154–1168. doi:10.1111/j.1365-313X.2007.03307.x.

Brown, D.M., Zeef, L.A.H., Ellis, J., Goodacre, R., and Turner, S.R. (2005). Identification of novel genes in Arabidopsis involved in secondary cell wall formation using expression profiling and reverse genetics. *The Plant Cell* 17, 2281–2295. doi:10.1105/tpc.105.031542.

Brown, D.M., Zhang, Z., Stephens, E., Dupree, P., and Turner, S.R. (2009). Characterization of IRX10 and IRX10-like reveals an essential role in glucuronoxylan biosynthesis in Arabidopsis. *The Plant Journal: for Cell and Molecular Biology* 57, 732–746. doi:10.1111/j.1365-313X.2008.03729.x.

Brunecky, R., Selig, M.J., Vinzant, T.B., Himmel, M.E., Lee, D., Blaylock, M.J., et al. (2011). In planta expression of A. cellulolyticus Cel5A endocellulase reduces cell wall recalcitrance in tobacco and maize. *Biotechnology for Biofuels* 4, 1. doi:10.1186/1754-6834-4-1.

Buanafina, M.M. de O., Langdon, T., Hauck, B., Dalton, S., and Morris, P. (2008). Expression of a fungal ferulic acid esterase increases cell wall digestibility of tall fescue (Festuca arundinacea). *Plant Biotechnology Journal* 6, 264–280. doi:10.1111/j.1467-7652.2007.00317.x.

Buanafina, M.M. de O., Langdon, T., Hauck, B., Dalton, S., Timms-Taravella, E., and Morris, P. (2010). Targeting expression of a fungal ferulic acid esterase to the apoplast, endoplasmic reticulum or golgi can disrupt feruloylation of the growing cell wall and increase the biodegradability of tall fescue (Festuca arundinacea). *Plant Biotechnology Journal* 8, 316–331. doi:10.1111/j.1467-7652.2009.00485.x.

Burton, R.A., Collins, H.M., Kibble, N.A.J., Smith, J.A., Shirley, N.J., Jobling, S.A., et al. (2011). Over-expression of specific HvCslF cellulose synthase-like genes in transgenic barley increases the levels of cell wall (1,3;1,4)-β-d-glucans and alters their fine structure. *Plant Biotechnology Journal* 9, 117–135. doi:10.1111/j.1467-7652.2010.00532.x.

Busse-Wicher, M., Gomes, T.C.F., Tryfona, T., Nikolovski, N., Stott, K., Grantham, N.J., et al. (2014). The pattern of xylan acetylation suggests xylan may interact with cellulose microfibrils as a twofold helical screw in the secondary plant cell wall of Arabidopsis thaliana. *The Plant Journal: for Cell and Molecular Biology* 79, 492–506. doi:10.1111/tpj.12575.

Cai, X., Zhang, X., and Wang, D. (2011). Land availability for biofuel production. *Environmental Science & Technology* 45, 334–339. doi:10.1021/es103338e.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research* 37, D233-8. doi:10.1093/nar/gkn663.

Chen, P., and Peng, L. (2013). "CHAPTER 6. The Diversity of Lignocellulosic Biomass Resources and their Evaluation for Use as Biofuels and Chemicals," in *Biological Conversion of Biomass for Fuels and Chemicals: Explorations from Natural Utilization Systems* Energy and environment series., eds. J. Sun, S.-Y. Ding, and J. D. Peterson (Cambridge: Royal Society of Chemistry), 83–113. doi:10.1039/9781849734738-00083.

Chen, X., Vega-Sánchez, M.E., Verhertbruggen, Y., Chiniquy, D., Canlas, P.E., Fagerström, A., et al. (2013). Inactivation of osirx10 leads to decreased xylan content in rice culm cell walls and improved biomass saccharification. *Molecular Plant* 6, 570–573. doi:10.1093/mp/sss135.

Chiniquy, D., Underwood, W., Corwin, J., Ryan, A., Szemenyei, H., Lim, C.C., et al. (2019). PMR5, an acetylation protein at the intersection of pectin biosynthesis and defense against fungal pathogens. *The Plant Journal: for Cell and Molecular Biology*. doi:10.1111/tpj.14497.

Christensen, U., and Scheller, H.V. (2012). Regulation of (1,3;1,4)-β-d-glucan synthesis in developing endosperm of barley lys mutants. *Journal of cereal science* 55, 69–76. doi:10.1016/j.jcs.2011.10.005.

Cosgrove, D.J. (2000). Loosening of plant cell walls by expansins. *Nature* 407, 321–326. doi:10.1038/35030000.

Cosgrove, D.J. (2005). Growth of the plant cell wall. *Nature Reviews. Molecular Cell Biology* 6, 850–861. doi:10.1038/nrm1746.

Damm, T., Commandeur, U., Fischer, R., Usadel, B., and Klose, H. (2016). Improving the utilization of lignocellulosic biomass by polysaccharide modification. *Process Biochemistry* 51, 288–296. doi:10.1016/j.procbio.2015.12.003.

Donev, E., Gandla, M.L., Jönsson, L.J., and Mellerowicz, E.J. (2018). Engineering Non-cellulosic Polysaccharides of Wood for the Biorefinery. *Frontiers in plant science* 9, 1537. doi:10.3389/fpls.2018.01537.

Donohoe, B.S., Wei, H., Mittal, A., Shollenberger, T., Lunin, V.V., Himmel, M.E., et al. (2017). Towards an understanding of enhanced biomass digestibility by in planta expression of a family 5 glycoside hydrolase. *Scientific Reports* 7, 4389. doi:10.1038/s41598-017-04502-1.

Fan, C., Feng, S., Huang, J., Wang, Y., Wu, L., Li, X., et al. (2017). AtCesA8-driven OsSUS3 expression leads to largely enhanced biomass saccharification and lodging resistance by distinctively altering lignocellulose features in rice. *Biotechnology for Biofuels* 10, 221. doi:10.1186/s13068-017-0911-0.

Fan, C., Wang, G., Wu, L., Liu, P., Huang, J., Jin, X., et al. (2019). Distinct cellulose and callose accumulation for enhanced bioethanol production and biotic stress resistance in OsSUS3 transgenic rice. *Carbohydrate polymers*, 115448. doi:10.1016/j.carbpol.2019.115448.

Fujii, S., Hayashi, T., and Mizuno, K. (2010). Sucrose synthase is an integral component of the cellulose synthesis machinery. *Plant & Cell Physiology* 51, 294–301. doi:10.1093/pcp/pcp190.

Gírio, F.M., Fonseca, C., Carvalheiro, F., Duarte, L.C., Marques, S., and Bogel-Łukasik, R. (2010). Hemicelluloses for fuel ethanol: A review. *Bioresource Technology* 101, 4775–4800. doi:10.1016/j.biortech.2010.01.088.

Gondolf, V.M., Stoppel, R., Ebert, B., Rautengarten, C., Liwanag, A.J., Loqué, D., et al. (2014). A gene stacking approach leads to engineered plants with highly increased galactan levels in Arabidopsis. *BMC Plant Biology* 14, 344. doi:10.1186/s12870-014-0344-x.

Grabber, J.H., Ralph, J., Lapierre, C., and Barrière, Y. (2004). Genetic and molecular basis of grass cell-wall degradability. I. Lignin-cell wall matrix interactions. *Comptes Rendus Biologies* 327, 455–465.

Grisolia, M.J., Peralta, D.A., Valdez, H.A., Barchiesi, J., Gomez-Casati, D.F., and Busi, M.V. (2017). The targeting of starch binding domains from starch synthase III to the cell wall alters cell wall composition and properties. *Plant Molecular Biology* 93, 121–135. doi:10.1007/s11103-016-0551-y.

Guidi, W., E., F., and Labrecque, M. (2013). "Short-Rotation Coppice of Willows for the Production of Biomass in Eastern Canada," in *Biomass Now - Sustainable Growth and Use*, ed. M. D. Matovic (InTech). doi:10.5772/51111.

Hall, M., Bansal, P., Lee, J.H., Realff, M.J., and Bommarius, A.S. (2010). Cellulose crystallinity--a key predictor of the enzymatic hydrolysis rate. *The FEBS Journal* 277, 1571–1582. doi:10.1111/j.1742-4658.2010.07585.x.

Hansen, S.F., Harholt, J., Oikawa, A., and Scheller, H.V. (2012). Plant glycosyltransferases beyond cazy: A perspective on DUF families. *Frontiers in plant science* 3, 59. doi:10.3389/fpls.2012.00059.

Hayashi, T., Yoshida, K., Park, Y.W., Konishi, T., and Baba, K. (2005). Cellulose metabolism in plants. *International Review of Cytology* 247, 1–34. doi:10.1016/S0074-7696(05)47001-1.

Heaton, E.A., Dohlemon, F.G., and Long, S.P. (2008). Meeting US biofuel goals with less land: the potential of Miscanthus. *Global change biology* 14, 2000–2014. doi:10.1111/j.1365-2486.2008.01662.x.

Himmel, M.E., and Bayer, E.A. (2009). Lignocellulose conversion to biofuels: current challenges, global perspectives. *Current Opinion in Biotechnology* 20, 316–317. doi:10.1016/j.copbio.2009.05.005.

Hu, H., Zhang, R., Feng, S., Wang, Y., Wang, Y., Fan, C., et al. (2018). Three AtCesA6-like members enhance biomass production by distinctively promoting cell growth in Arabidopsis. *Plant Biotechnology Journal* 16, 976–988. doi:10.1111/pbi.12842.

Huang, J., Xia, T., Li, G., Li, X., Li, Y., Wang, Y., et al. (2019). Overproduction of native endo-β-1,4-glucanases leads to largely enhanced biomass saccharification and bioethanol production by specific modification of cellulose features in transgenic rice. *Biotechnology for Biofuels* 12, 11. doi:10.1186/s13068-018-1351-1.

Jensen, J.K., Johnson, N.R., and Wilkerson, C.G. (2014). Arabidopsis thaliana IRX10 and two related proteins from psyllium and Physcomitrella patens are xylan xylosyltransferases. *The Plant Journal: for Cell and Molecular Biology* 80, 207–215. doi:10.1111/tpj.12641.

Joshi, C.P., Thammannagowda, S., Fujino, T., Gou, J.-Q., Avci, U., Haigler, C.H., et al. (2011). Perturbation of wood cellulose synthesis causes pleiotropic effects in transgenic aspen. *Molecular Plant* 4, 331–345. doi:10.1093/mp/ssq081.

Junior, H.J.E., Melo, R.X. de, Sartori, M.M.P., Guerra, S.P.S., and Ballarin, A.W. (2016). Sustainable use of eucalypt biomass grown on short rotation coppice for bioenergy. *Biomass and Bioenergy* 90, 15–21. doi:10.1016/j.biombioe.2016.03.037.

Kaida, R., Kaku, T., Baba, K., Oyadomari, M., Watanabe, T., Nishida, K., et al. (2009). Loosening xyloglucan accelerates the enzymatic degradation of cellulose in wood. *Molecular Plant* 2, 904–909. doi:10.1093/mp/ssp060.

Kimura, S., and Itoh, T. (2007). "Biogenesis and function of cellulose in the tunicates," in *Cellulose: molecular and structural biology*, eds. R. M. Brown and I. M. Saxena

(Dordrecht: Springer Netherlands), 217–236. doi:10.1007/978-1-4020-5380-1_13.
Klose, H., Günl, M., Usadel, B., Fischer, R., and Commandeur, U. (2015). Cell wall modification in tobacco by differential targeting of recombinant endoglucanase from Trichoderma reesei. *BMC Plant Biology* 15, 54. doi:10.1186/s12870-015-0443-3.
Knoch, E., Dilokpimol, A., Tryfona, T., Poulsen, C.P., Xiong, G., Harholt, J., et al. (2013). A β-glucuronosyltransferase from Arabidopsis thaliana involved in biosynthesis of type II arabinogalactan has a role in cell elongation during seedling growth. *The Plant Journal: for Cell and Molecular Biology* 76, 1016–1029. doi:10.1111/tpj.12353.
Lao, J., Oikawa, A., Bromley, J.R., McInerney, P., Suttangkakul, A., Smith-Moritz, A.M., et al. (2014). The plant glycosyltransferase clone collection for functional genomics. *The Plant Journal: for Cell and Molecular Biology* 79, 517–529. doi:10.1111/tpj.12577.
Lee, C., Zhong, R., Richardson, E.A., Himmelsbach, D.S., McPhail, B.T., and Ye, Z.-H. (2007). The PARVUS gene is expressed in cells undergoing secondary wall thickening and is essential for glucuronoxylan biosynthesis. *Plant & Cell Physiology* 48, 1659–1672. doi:10.1093/pcp/pcm155.
Lee, C., Zhong, R., and Ye, Z.-H. (2012). Arabidopsis family GT43 members are xylan xylosyltransferases required for the elongation of the xylan backbone. *Plant & Cell Physiology* 53, 135–143. doi:10.1093/pcp/pcr158.
Li, F., Zhang, M., Guo, K., Hu, Z., Zhang, R., Feng, Y., et al. (2015). High-level hemicellulosic arabinose predominately affects lignocellulose crystallinity for genetically enhancing both plant lodging resistance and biomass enzymatic digestibility in rice mutants. *Plant Biotechnology Journal* 13, 514–525. doi:10.1111/pbi.12276.
Li, G., Jones, K.C., Eudes, A., Pidatala, V.R., Sun, J., Xu, F., et al. (2018). Overexpression of a rice BAHD acyltransferase gene in switchgrass (Panicum virgatum L.) enhances saccharification. *BMC Biotechnology* 18, 54. doi:10.1186/s12896-018-0464-8.
Li, M., Wang, S., Liu, Y., Zhang, Y., Ren, M., Liu, L., et al. (2019a). Overexpression of PsnSuSy1, 2 genes enhances secondary cell wall thickening, vegetative growth, and mechanical strength in transgenic tobacco. *Plant Molecular Biology* 100, 215–230. doi:10.1007/s11103-019-00850-w.
Li, M., Yoo, C.G., Pu, Y., Biswal, A.K., Tolbert, A.K., Mohnen, D., et al. (2019b). Downregulation of pectin biosynthesis gene GAUT4 leads to reduced ferulate and lignin-carbohydrate cross-linking in switchgrass. *Communications Biology* 2, 22. doi:10.1038/s42003-018-0265-6.
Liebminger, E., Grass, J., Altmann, F., Mach, L., and Strasser, R. (2013). Characterizing the link between glycosylation state and enzymatic activity of the endo-β1,4-glucanase KORRIGAN1 from Arabidopsis thaliana. *The Journal of Biological Chemistry* 288, 22270–22280. doi:10.1074/jbc.M113.475558.
Lionetti, V., Francocci, F., Ferrari, S., Volpi, C., Bellincampi, D., Galletti, R., et al. (2010). Engineering the cell wall by reducing de-methyl-esterified homogalacturonan improves saccharification of plant tissues for bioconversion. *Proceedings of the National Academy of Sciences of the United States of America* 107, 616–621. doi:10.1073/pnas.0907549107.
Liu, L., Shang-Guan, K., Zhang, B., Liu, X., Yan, M., Zhang, L., et al. (2013). Brittle Culm1, a COBRA-like protein, functions in cellulose assembly through binding cellulose microfibrils. *PLoS Genetics* 9, e1003704. doi:10.1371/journal.pgen.1003704.
Loqué, D., Scheller, H.V., and Pauly, M. (2015). Engineering of plant cell walls for enhanced biofuel production. *Current Opinion in Plant Biology* 25, 151–161.

doi:10.1016/j.pbi.2015.05.018.
Marcia, M.D. (2009). Feruloylation in grasses: current and future perspectives. *Molecular plant* 2, 861–872.
Matthysse, A.G., Deschet, K., Williams, M., Marry, M., White, A.R., and Smith, W.C. (2004). A functional cellulose synthase from ascidian epidermis. *Proceedings of the National Academy of Sciences of the United States of America* 101, 986–991. doi:10.1073/pnas.0303623101.
McCormick, A.J., Watt, D.A., and Cramer, M.D. (2009). Supply and demand: sink regulation of sugar accumulation in sugarcane. *Journal of Experimental Botany* 60, 357–364. doi:10.1093/jxb/ern310.
McFarlane, H.E., Döring, A., and Persson, S. (2014). The cell biology of cellulose synthesis. *Annual review of plant biology* 65, 69–94. doi:10.1146/annurev-arplant-050213-040240.
Meents, M.J., Watanabe, Y., and Samuels, A.L. (2018). The cell biology of secondary cell wall biosynthesis. *Annals of Botany* 121, 1107–1125. doi:10.1093/aob/mcy005.
Mir, B.A., Mewalal, R., Mizrachi, E., Myburg, A.A., and Cowan, D.A. (2014). Recombinant hyperthermophilic enzyme expression in plants: a novel approach for lignocellulose digestion. *Trends in Biotechnology* 32, 281–289. doi:10.1016/j.tibtech.2014.03.003.
Morris, P., Dalton, S., Langdon, T., Hauck, B., and de Buanafina, M.M.O. (2017). Expression of a fungal ferulic acid esterase in suspension cultures of tall fescue (Festuca arundinacea) decreases cell wall feruloylation and increases rates of cell wall digestion. *Plant cell, tissue and organ culture* 129, 181–193. doi:10.1007/s11240-017-1168-9.
Mortimer, J.C., Faria-Blanc, N., Yu, X., Tryfona, T., Sorieul, M., Ng, Y.Z., et al. (2015). An unusual xylan in Arabidopsis primary cell walls is synthesised by GUX3, IRX9L, IRX10L and IRX14. *The Plant Journal: for Cell and Molecular Biology* 83, 413–426. doi:10.1111/tpj.12898.
Ndimande, S. (2013). Increasing Cellulosic Biomass in Sugarcane. [dissertation] [Stellenbosch, South Africa]: Stellenbosch University
Niu, E., Fang, S., Shang, X., and Guo, W. (2018). Ectopic expression of GhCOBL9A, a cotton glycosyl-phosphatidyl inositol-anchored protein encoding gene, promotes cell elongation, thickening and increased plant biomass in transgenic Arabidopsis. *Molecular Genetics and Genomics* 293, 1191–1204. doi:10.1007/s00438-018-1452-3.
Park, S.-H., Ong, R.G., and Sticklen, M. (2016). Strategies for the production of cell wall-deconstructing enzymes in lignocellulosic biomass and their utilization for biofuel production. *Plant Biotechnology Journal* 14, 1329–1344. doi:10.1111/pbi.12505.
Park, Y.W., Baba, K., Furuta, Y., Iida, I., Sameshima, K., Arai, M., et al. (2004). Enhancement of growth and cellulose accumulation by overexpression of xyloglucanase in poplar. *FEBS Letters* 564, 183–187. doi:10.1016/S0014-5793(04)00346-1.
Pawar, P.M.-A., Derba-Maceluch, M., Chong, S.-L., Gandla, M.L., Bashar, S.S., Sparrman, T., et al. (2017). In muro deacetylation of xylan affects lignin properties and improves saccharification of aspen wood. *Biotechnology for Biofuels* 10, 98. doi:10.1186/s13068-017-0782-4.
Pawar, P.M.-A., Derba-Maceluch, M., Chong, S.-L., Gómez, L.D., Miedes, E., Banasiak, A., et al. (2016). Expression of fungal acetyl xylan esterase in Arabidopsis thaliana improves saccharification of stem lignocellulose. *Plant Biotechnology Journal* 14, 387–397. doi:10.1111/pbi.12393.
Petersen, P.D., Lau, J., Ebert, B., Yang, F., Verhertbruggen, Y., Kim, J.S., et al. (2012).

Engineering of plants with improved properties as biofuels feedstocks by vessel-specific complementation of xylan biosynthesis mutants. *Biotechnology for Biofuels* 5, 84. doi:10.1186/1754-6834-5-84.

Pogorelko, G., Fursova, O., Lin, M., Pyle, E., Jass, J., and Zabotina, O.A. (2011). Post-synthetic modification of plant cell walls by expression of microbial hydrolases in the apoplast. *Plant Molecular Biology* 77, 433–445. doi:10.1007/s11103-011-9822-9.

Reem, N.T., Pogorelko, G., Lionetti, V., Chambers, L., Held, M.A., Bellincampi, D., et al. (2016). Decreased polysaccharide feruloylation compromises plant cell wall integrity and increases susceptibility to necrotrophic fungal pathogens. *Frontiers in plant science* 7, 630. doi:10.3389/fpls.2016.00630.

Ren, Y., Hansen, S.F., Ebert, B., Lau, J., and Scheller, H.V. (2014). Site-directed mutagenesis of IRX9, IRX9L and IRX14 proteins involved in xylan biosynthesis: glycosyltransferase activity is not required for IRX9 function in Arabidopsis. *Plos One* 9, e105014. doi:10.1371/journal.pone.0105014.

Rennie, E.A., and Scheller, H.V. (2014). Xylan biosynthesis. *Current Opinion in Biotechnology* 26, 100–107. doi:10.1016/j.copbio.2013.11.013.

Roudier, F., Fernandez, A.G., Fujita, M., Himmelspach, R., Borner, G.H.H., Schindelman, G., et al. (2005). COBRA, an Arabidopsis extracellular glycosyl-phosphatidyl inositol-anchored protein, specifically controls highly anisotropic expansion through its involvement in cellulose microfibril orientation. *The Plant Cell* 17, 1749–1763. doi:10.1105/tpc.105.031732.

Sanderson, M.A., and Adler, P.R. (2008). Perennial forages as second generation bioenergy crops. *International Journal of Molecular Sciences* 9, 768–788. doi:10.3390/ijms9050768.

Scheller, H.V., and Ulvskov, P. (2010). Hemicelluloses. *Annual review of plant biology* 61, 263–289. doi:10.1146/annurev-arplant-042809-112315.

Somerville, C., Youngs, H., Taylor, C., Davis, S.C., and Long, S.P. (2010). Feedstocks for lignocellulosic biofuels. *Science* 329, 790–792. doi:10.1126/science.1189268.

Stein, O., and Granot, D. (2019). An overview of sucrose synthases in plants. *Frontiers in plant science* 10, 95. doi:10.3389/fpls.2019.00095.

Sumiyoshi, M., Nakamura, A., Nakamura, H., Hakata, M., Ichikawa, H., Hirochika, H., et al. (2013). Increase in cellulose accumulation and improvement of saccharification by overexpression of arabinofuranosidase in rice. *Plos One* 8, e78269. doi:10.1371/journal.pone.0078269.

Tan, H.-T., Shirley, N.J., Singh, R.R., Henderson, M., Dhugga, K.S., Mayo, G.M., et al. (2015). Powerful regulatory systems and post-transcriptional gene silencing resist increases in cellulose content in cell walls of barley. *BMC Plant Biology* 15, 62. doi:10.1186/s12870-015-0448-y.

Terrett, O.M., and Dupree, P. (2019). Covalent interactions between lignin and hemicelluloses in plant secondary cell walls. *Current Opinion in Biotechnology* 56, 97–104. doi:10.1016/j.copbio.2018.10.010.

United Nations Environment Programme (2007). *Global Environment Outlook: Environment for Development*. 4th ed. , ed. UNEP New York Available at: https://na.unep.net/atlas/datlas/sites/default/files/GEO-4_Report_Full_en.pdf [Accessed November 14, 2019].

Urbanowicz, B.R., Peña, M.J., Moniz, H.A., Moremen, K.W., and York, W.S. (2014). Two

Arabidopsis proteins synthesize acetylated xylan in vitro. *The Plant Journal: for Cell and Molecular Biology* 80, 197–206. doi:10.1111/tpj.12643.

Vega-Sánchez, M.E., Loqué, D., Lao, J., Catena, M., Verhertbruggen, Y., Herter, T., et al. (2015). Engineering temporal accumulation of a low recalcitrance polysaccharide leads to increased C6 sugar content in plant cell walls. *Plant Biotechnology Journal* 13, 903–914. doi:10.1111/pbi.12326.

Wang, G., Gao, Y., Wang, J., Yang, L., Song, R., Li, X., et al. (2011). Overexpression of two cambium-abundant Chinese fir (Cunninghamia lanceolata) α-expansin genes ClEXPA1 and ClEXPA2 affect growth and development in transgenic tobacco and increase the amount of cellulose in stem cell walls. *Plant Biotechnology Journal* 9, 486–502. doi:10.1111/j.1467-7652.2010.00569.x.

Wang, J., Nayak, S., Koch, K., and Ming, R. (2013). Carbon partitioning in sugarcane (Saccharum species). *Frontiers in plant science* 4, 201. doi:10.3389/fpls.2013.00201.

Wang, Y., Fan, C., Hu, H., Li, Y., Sun, D., Wang, Y., et al. (2016). Genetic modification of plant cell walls to enhance biomass yield and biofuel production in bioenergy crops. *Biotechnology advances* 34, 997–1017. doi:10.1016/j.biotechadv.2016.06.001.

Wei, Z., Qu, Z., Zhang, L., Zhao, S., Bi, Z., Ji, X., et al. (2015). Overexpression of poplar xylem sucrose synthase in tobacco leads to a thickened cell wall and increased height. *Plos One* 10, e0120669. doi:10.1371/journal.pone.0120669.

Wu, A.-M., Hörnblad, E., Voxeur, A., Gerber, L., Rihouey, C., Lerouge, P., et al. (2010). Analysis of the Arabidopsis IRX9/IRX9-L and IRX14/IRX14-L pairs of glycosyltransferase genes reveals critical contributions to biosynthesis of the hemicellulose glucuronoxylan. *Plant Physiology* 153, 542–554. doi:10.1104/pp.110.154971.

Yang, Y., Yoo, C.G., Guo, H.-B., Rottmann, W., Winkeler, K.A., Collins, C.M., et al. (2017). Overexpression of a Domain of Unknown Function 266-containing protein results in high cellulose content, reduced recalcitrance, and enhanced plant growth in the bioenergy crop Populus. *Biotechnology for Biofuels* 10, 74. doi:10.1186/s13068-017-0760-x.

Zhou, Y., Li, S., Qian, Q., Zeng, D., Zhang, M., Guo, L., et al. (2009). BC10, a DUF266-containing and Golgi-located type II membrane protein, is required for cell-wall biosynthesis in rice (Oryza sativa L.). *The Plant Journal: for Cell and Molecular Biology* 57, 446–462. doi:10.1111/j.1365-313X.2008.03703.x.

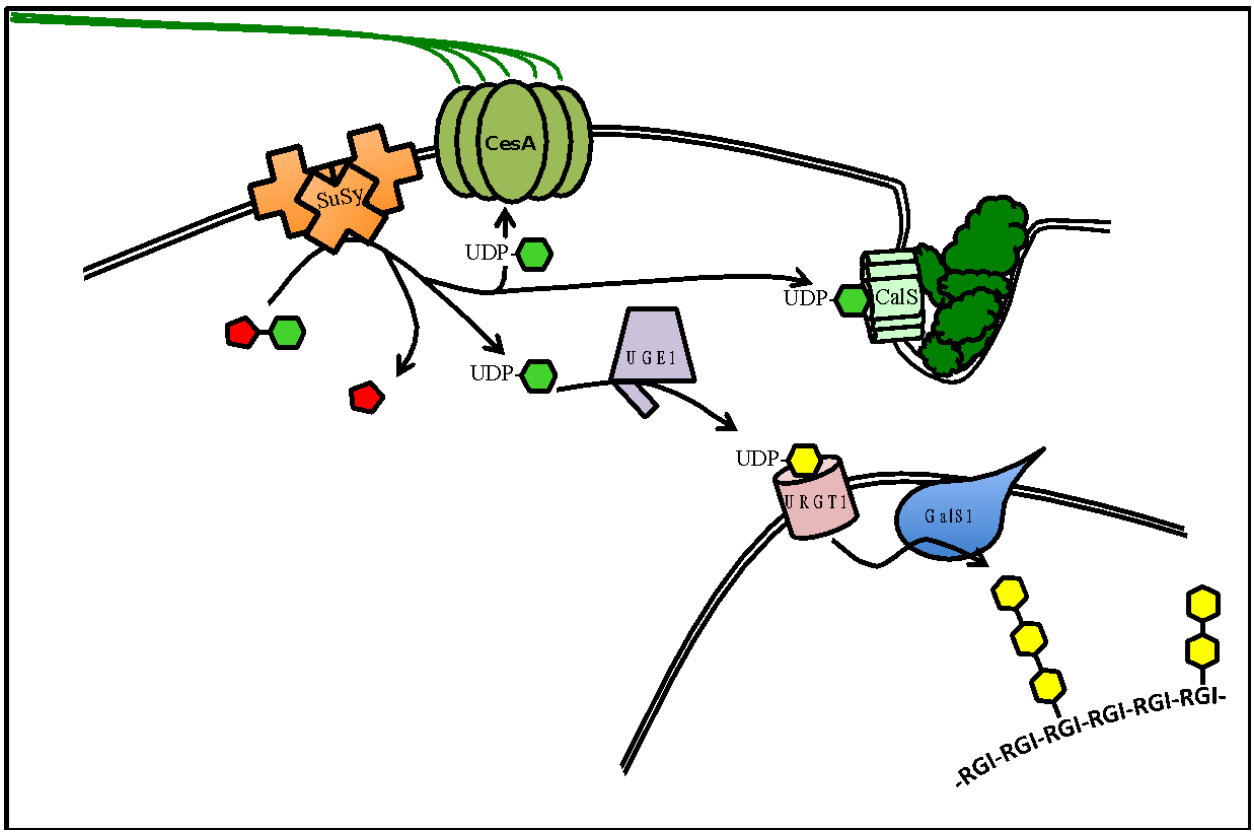

Figure 1 Schematic representation of the role of sucrose synthase (SuSy) in increasing the biosynthesis of several C6-sugar polysaccharides Sucrose synthase isoforms can be localized to the cytoplasm, tightly associated with the plasma membrane, some interacting directly with CesA, or in the cell wall. All catalyze the reversible conversion of sucrose into fructose (red) and UDP-Glu (green). UDP-Glu can be used to make cellulose by CesA or callose by CalS at the plasma membrane. It can also be converted to UDP-Gal (yellow) by UGE1, imported to the Golgi by URGT1, and used to synthesize the galactan side chains of pectic RGI

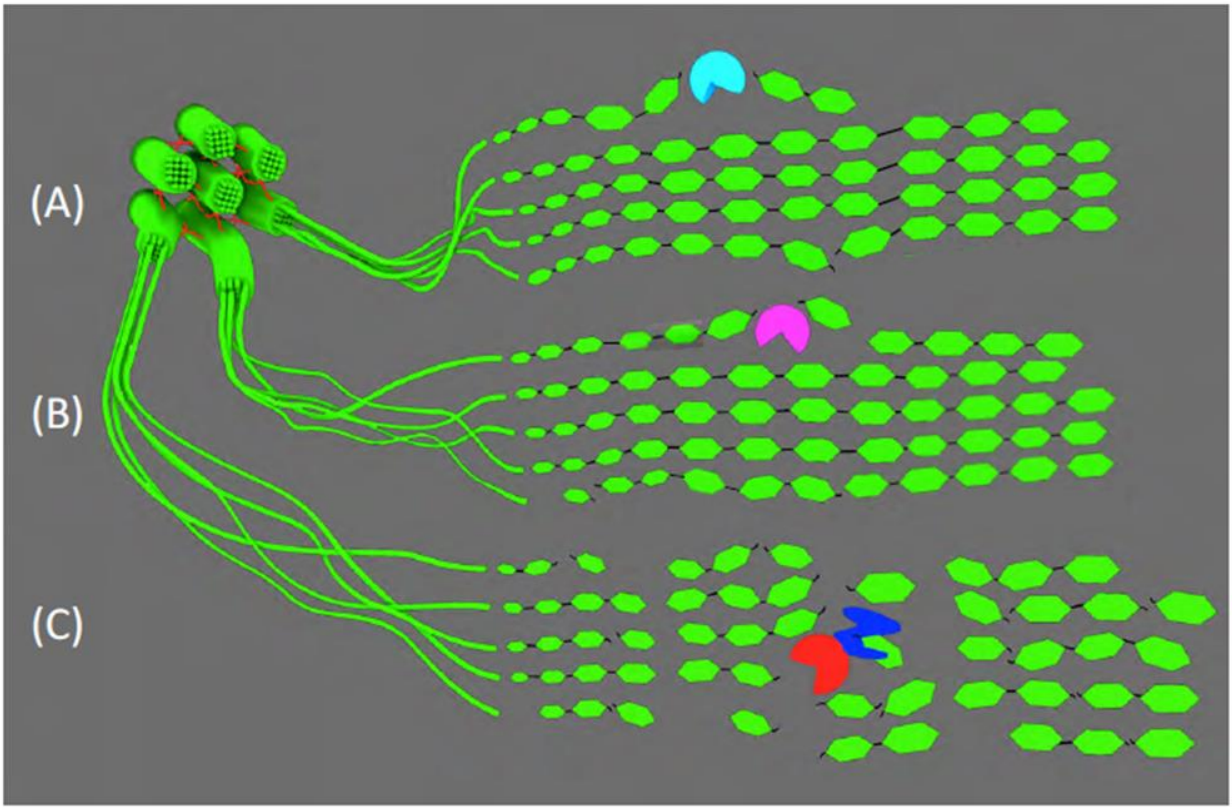

Figure 2 GH cuts to cellulose microfibrils (A) GH9B1 and GH9B3 and (B) thermophilic, CBM-truncated AcCel5A make relatively few cuts to superficial strands in the microfibril. (C) Mesophilic TrCel5A binds to cellulose via its CBM (blue), while the endoglucanase domain (red) makes cuts with significantly higher frequency due to its temperature optimum being similar to plant growth conditions

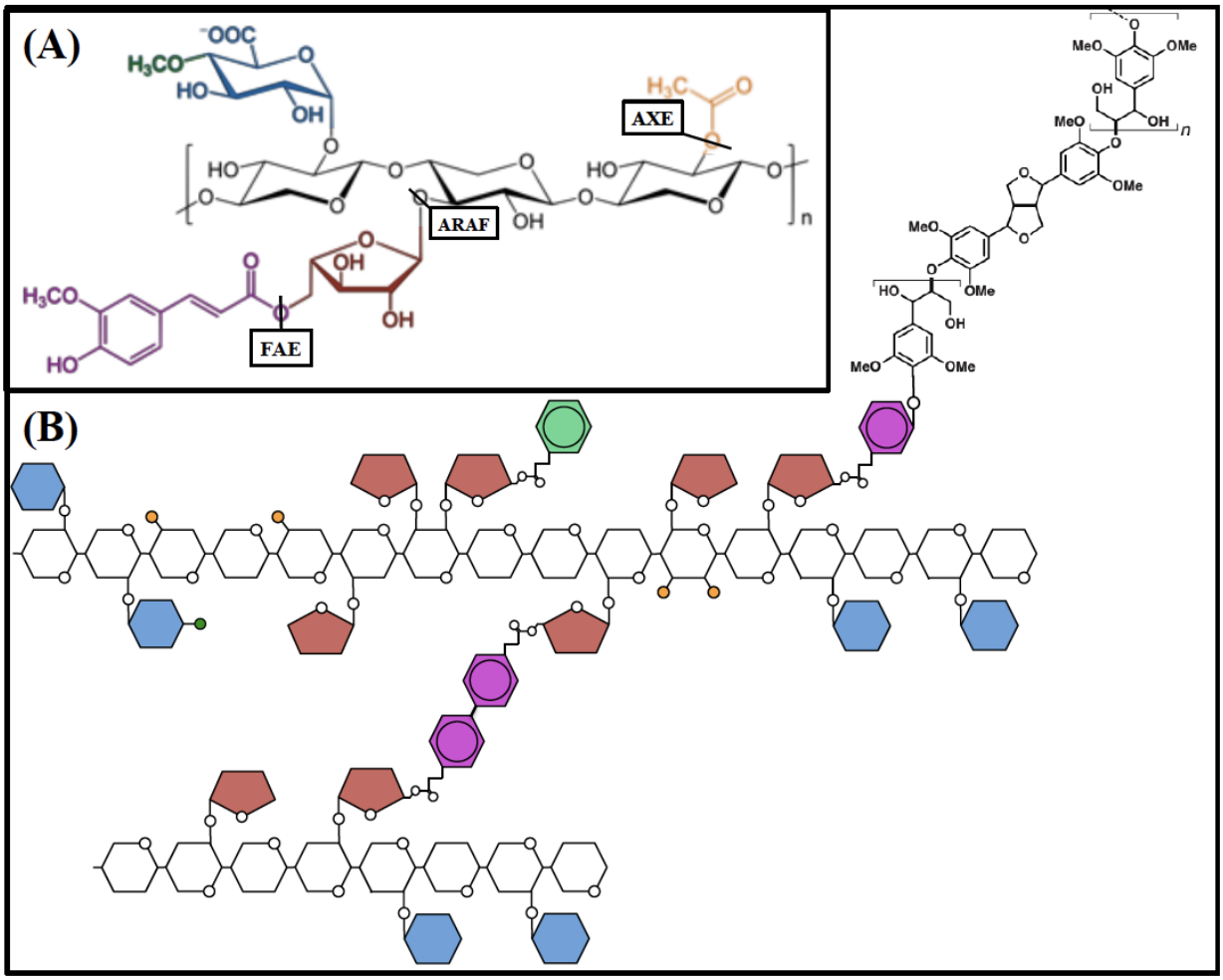

Figure 3 The hemicellulose xylan and interactions (A) Representative molecular structure of xylan module with side chains β-(1,4)-linked xylose residues (black) of the xylan backbone with acetyl (orange), arabinose (red) and glucuronic acid (blue). Arabinose is esterified to ferulic acid (magenta) and glucuronic acid is 4-*O*-methylated (green). Acetyl xylan esterase (AXE), arabinofuranosidase (ARAF) and ferulic acid esterase (FAE) indicating the bonds they hydrolyze (B) Schematic representation of the xylan chain, xylan-xylan diferulate cross-linking and ferulic acid-mediated lignin polymerization.