Identifying novel genes to improve lignocellulosic biomass as a feedstock for bioethanol

Poppy Marriott

PhD

University of York

Biology

September 2014

Abstract

The dwindling reserves of fossil fuels, coupled with the environmental consequences of burning these fuels, mean that a more sustainable alternative is required. Bioethanol produced from lignocellulosic biomass is an attractive candidate for the replacement of liquid transportation fuels. Lignocellulosic biomass is composed of plant cell walls, which are extremely resistant to digestion. Converting this biomass to fermentable sugars for bioethanol production therefore requires energetic pretreatment and expensive enzyme applications. To make lignocellulosic bioethanol a commercial reality, the conversion efficiency needs to be improved and one approach of doing so is to produce crops that are more susceptible to hydrolysis. To this end, this study used a forward genetic approach with the objective of identifying genes that affect the digestibility of plant biomass. A chemically mutagenised population of the model grass *Brachypodium distachyon* was screened for improved saccharification with industrial cellulases. This revealed 12 mutant lines with heritable increases in saccharification. Characterisation of these 12 mutants revealed a range of different alterations in cell wall composition. Interestingly, a number of the mutant lines showed no change in lignin content, which is thought to be the major contributor to cell wall recalcitrance. These results show that saccharification can be significantly improved through a number of distinct modifications of the cell wall, giving the potential for combining more than one of these modifications in biofuel crops to obtain even higher ethanol yields. Furthermore, the mutations seem to have little effect on plant growth, development or stem strength, important traits for crop field performance. Candidate causal mutations of three of the high saccharification mutants have been identified and characterisation of the mutated genes has begun. In the long term, this will enable subsequent examination of orthologous genes in the relevant cereal and grass crops used for biofuel production.

Table of contents

6

List of figures

11

List of tables

Acknowledgements

Declaration

The work presented in this thesis is the sole effort of the author, except where explicitly stated. Reference to the work of others has been duly acknowledged. No portion of this work has been submitted for any other degree. Any reference to this work should be acknowledged.

Chapter 1 – Introduction

1.1 World energy

The burning of fossil fuels is responsible for producing approximately 81% of the world's energy (1) and a rapidly expanding and industrialising world population is driving energy demand ever higher. It is estimated that the world population will increase from 7.2 billion in mid-2013 to 9.6 billion in 2050 and 10.9 billion in 2100 (2), while global energy demand is predicted to increase by 40% between 2009 and 2035 (1). This burning of fossil fuels is contributing significantly to climate change through the release of large amounts of greenhouse gases (GHGs). It has been estimated that fossil fuels are responsible for 80% of anthropogenic GHG emissions (3), while global energy use is responsible for 65%, a figure that is predicted to increase to 72% by 2035 (1). Not only is the combustion of these fuels the main source of anthropogenic GHG emissions but supplies are finite and are running out rapidly. Therefore, the need for sustainable and reliable renewable energy sources has become a pressing concern.

1.1.1 Our reliance on fossil fuels for transportation

Of the 81% global energy provided by fossil fuels, 41% comes from oil which is dominated by use in the transport sector (92%). Global demand for oil is predicted to increase by 18% in the period 2009 to 2035, with a 43% increase in energy used for transport. The number of road transport vehicles is expected to almost double by 2035 (1). As transportation is responsible for such a large proportion of global energy use, replacing the fuels used in this sector with renewable sources could greatly aid the drive to reduce our reliance on fossil fuels. Indeed, there are a number of policies in place across the globe to increase the proportion of energy used for transport that comes from renewable sources. The European Union, for example, has set a target of 10% of transport fuels to be derived from renewable sources by 2020. Similarly, the United States has set a target of producing 136 billion litres of renewable transport fuels by

2022, over five times the production in 2011 (4). Liquid biofuels such as bioethanol, produced by fermentation of plant-derived sugars, offer a renewable and hopefully more sustainable alternative supply of liquid transportation fuels.

1.1.2 First generation bioethanol

First generation bioethanol is produced from crops traditionally used for human or animal food, such as corn and sugar cane. It is produced from parts of the plant high in either starch or sucrose, which are relatively easy to break down into the monosaccharides required for fermentation into ethanol. The technology for producing first generation bioethanol is now mature, efficient and well established in some parts of the world. In Brazil, for example, a government-enforced program has been in place since 1975 to encourage use of sugar cane bioethanol as a transportation fuel, using subsidies and incentives. This has led to the development of a highly effective and competitive bioethanol industry that now competes with gasoline on a cost basis without the need for subsidies (5). Use of bioethanol as a transportation fuel is increasing rapidly and total world production tripled between 2000 and 2007 (Figure 1), aided by a number of targets, policies and subsidies (4). In 2007 biofuel provided 1.5% of the world's transport fuel, with the United States being the biggest producer, followed by Brazil and then the European Union (Figure 1) (4). Although first generation bioethanol can be produced efficiently, there is growing concern over its widespread development for a number of reasons, detailed below.
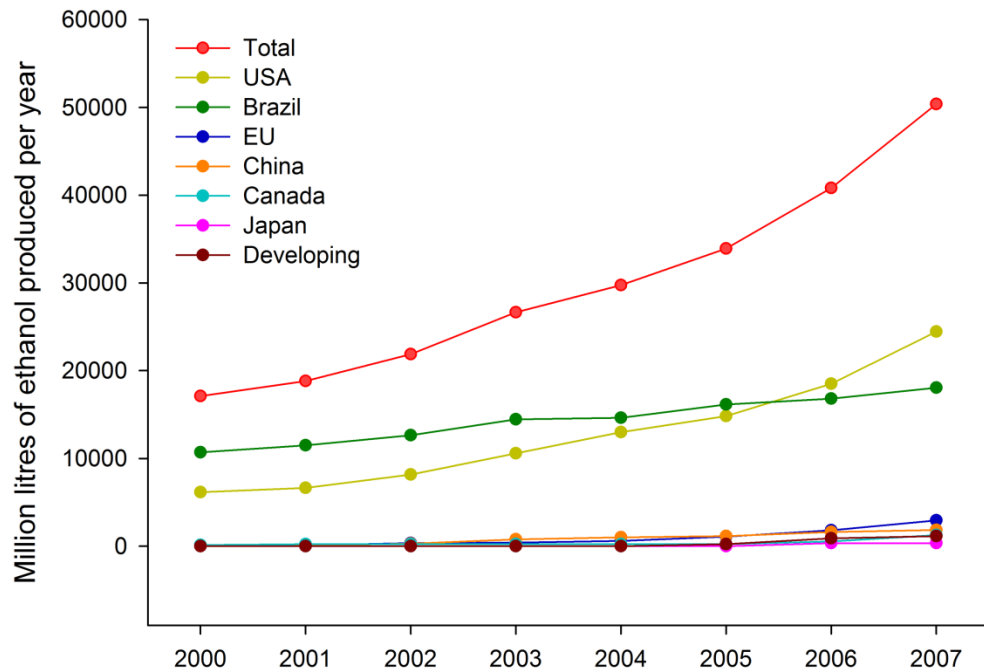
Figure 1: World bioethanol production by region 2000 - 2007. Data taken from the International Energy Agency (4).

1.1.2.1 Food versus fuel

Production of first generation bioethanol competes directly with human and animal food supply and there is only a limited amount of arable land to grow these crops. Subsequently, considerable increases in the production of first generation ethanol are likely to put added pressure on food supplies at a time when the world population is increasing rapidly and there is an upward trend in global food consumption (6). A study published in 2007 estimated that even if the entire global production of wheat, barley, corn, sorghum and sugar cane was used to produce bioethanol, it would only provide the equivalent of 19.4% of the oil used for transportation in 2007 under current production efficiencies (7). This demonstrates the huge amount of land and raw material we would require to meet world demands. According to the same study, if 16% of existing pasturelands were converted for agriculture and crop yields were increased by 60% (a plausible scenario according to the authors), there would be a 1.9 billion ton

surplus of grain which could be used to produce 23% of transportation fuel in 2050, taking into account population growth predictions and current ethanol production efficiencies (7). However, this assumes that there is no increase in food consumption per person. This is unlikely due to increasing per capita income of developing countries and a continual increase in the consumption of meat and dairy products (a vegetarian diet requires 1.3 kg grain equivalent per day, compared to 2.4 kg for a moderate diet and 4.2 kg for an affluent diet). If the food demand increased to three times the 2007 level (a moderate diet scenario according to the authors) then there would be no surplus grain for bioethanol production, even with the increase in yields and conversion of pastureland detailed above (7).

There are also concerns over increases in production of first generation bioethanol leading to a rise in food prices (8). Indeed, between 2005 and 2008, when bioethanol production expanded rapidly, maize prices increased by 131% and wheat prices by 177% (4). These increases were not solely due to the use of these cereals for bioethanol production and other factors included reduced supply, increased production costs due to high energy and fertiliser costs, and a weak US dollar. However, increased bioethanol production did have an impact, which has been estimated to vary widely between as little as 3% to as much as 75% (9, 10).

1.1.2.2 First generation bioethanol and the environment

In addition to the concerns over the effect of first generation bioethanol production on the cost and supply of human food, there are also concerns about the effect on the environment. Firstly, it is thought that an increase in bioethanol production will lead to an increase in deforestation, due to the limited amount of surplus arable land mentioned above (11). As well as the negative impact of deforestation on biodiversity, the conversion of land for agriculture is often associated with large GHG emissions (12). Furthermore, food crops require high inputs of water, fertilisers and pesticides, and

therefore have a high carbon footprint (13). One of the motivations for switching from fossil fuel-based transportation fuels to bioethanol is to provide a more environmentally sustainable energy source, in which the carbon fixed by the plants during growth helps to offset the carbon released during production and combustion of the fuel. However, a number of studies suggest that GHG emissions are reduced very little, if at all, by exchanging petroleum for first generation bioethanol (12, 14, 15).

1.1.3 Second generation bioethanol

Second generation bioethanol is produced from non-food crops or parts of the plant that are not used for food. It therefore has the potential to avoid many of the issues associated with first generation bioethanol discussed above. The raw material for second generation ethanol production is lignocellulose, which comes from plant secondary cell walls, the thick, strengthening layer of the cell wall that is laid down inside the primary cell wall after cell elongation has terminated. Approximately 70 - 75% of lignocellulose is comprised of polysaccharides which can potentially be converted into monosaccharides for fermentation (16). Lignocellulose can be acquired from the stems of food crops which have few current applications, resulting in the excess biomass often being burnt or left to rot in the field. For example, it was estimated in a 2008 study that in the UK there are 5.5 million tonnes surplus cereal straw per year (17) which could be used to produce 275 million gallons of ethanol at current production efficiencies (18). Alternatively, dedicated biomass crops that produce large yields of lignocellulose can be grown for bioethanol production, such as switchgrass or Miscanthus. These are low input crops that can be grown on degraded land that is unsuitable for food crops and therefore have the potential to have little impact on human food supply (19, 20). Life cycle analyses of second generation biofuels show a reduction in GHG emissions of 60 - 120% compared to petroleum fuels (4), generally a much larger reduction than that achieved with first generation biofuels (21-23). However, second generation bioethanol

currently only makes up around 0.1% total bioethanol production (Figure 2) (4) due to production challenges which will be discussed below.
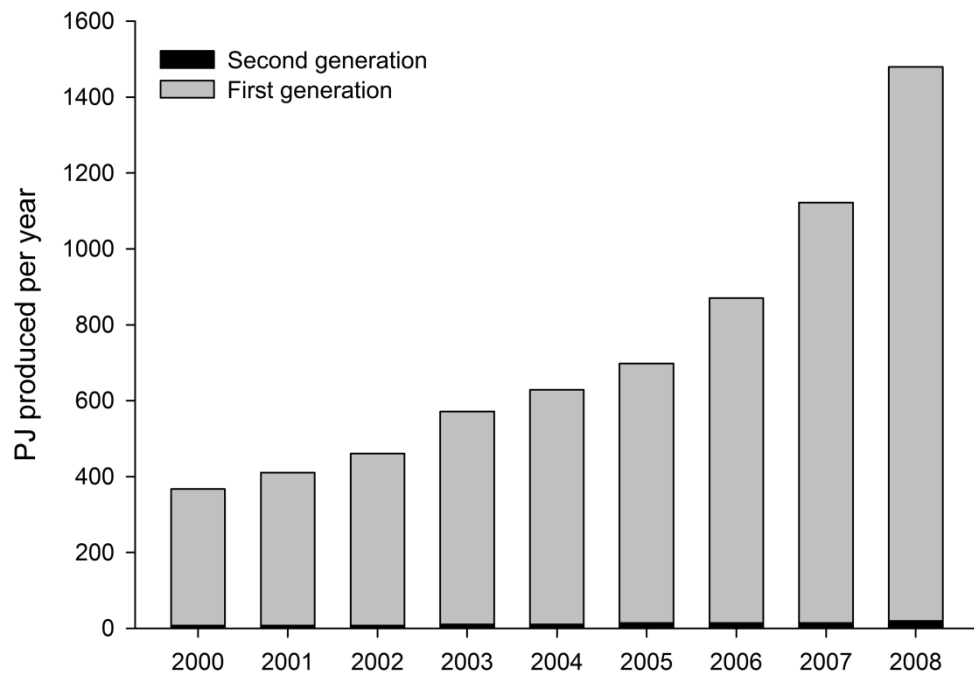
Figure 2: First and second generation bioethanol production 2000 - 2008. PJ: petajoules. Data from Sims et al. (24).

1.2 Components of the cell wall

The main challenge in producing bioethanol from lignocellulosic biomass is that the polysaccharides occur as part of a complex and indigestible macromolecular material within the cell wall, which is extremely resistant to digestion (25). The bioethanol production process therefore requires energetic pretreatment, in order to open up or break down the cell wall structure so that the enzymes can more easily access the polysaccharides, and high inputs of expensive enzyme. This makes the production of second generation bioethanol a costly process which is currently commercially unviable (4, 26). The main components that make up the plant cell wall, their synthesis and how they contribute to the indigestible nature of lignocellulosic biomass are discussed below.

1.2.1 Cellulose

One of the major constituents of the cell wall is cellulose which comprises around one third of the total mass of most plants (27). Cellulose is composed of linear β-1,4 glucans which are stiff, rigid, insoluble polymers of glucose that lack side chains. Sequential glucose residues in the glucans are rotated 180 ° so that cellobiose is the repeating unit (Figure 3A) (28). Multiple glucan chains are arranged in parallel crystalline arrays to form cellulose microfibrils (29), with the glucan chains held together by hydrogen bonds and Van derWaals forces (30, 31). Cellulose is the most important component of the cell wall for bioethanol production, containing large amounts of glucose which can easily be fermented to ethanol once released from the polysaccharide structure. However, the crystallinity of cellulose makes it extremely inaccessible to digestive enzymes and also water which is required for hydrolysis of the glucosidic bonds, making it resistant to digestion.

Cellulose is synthesised by cellulose synthases (CESA) which occur in hexameric complexes located in the plasma membrane (29, 32-34). Each of the six subunits of the CESA complex consists of multiple CESA proteins, enabling the simultaneous synthesis of all the glucan chains in a single microfibril to produce the parallel chain organisation (35). For a number of years it was thought that the cellulose microfibrils in plants are made up of 36 glucan chains and so each subunit of the CESA complex contains six CESA proteins (Figure 3B) (29, 36, 37). However, recently it has been shown that a 24-chain cellulose microfibril is more likely (38, 39). Spectroscopic methods combined with diffraction methods allowed the diameter of the microfibrils to be accurately measured at 3.0 nm, a size that would only permit 24 chains. The most likely structure for these 24 chains is a rectangular shape with eight sheets composed of three glucan chains each, with hydrogen bonds linking the sheets (Figure 3C) (38, 39). It is thought that three different homologues of the CESA proteins are required to form the CESA complex and cellulose microfibrils, and that two different sets of three form

the complexes for cellulose production in primary and secondary cell walls (40-42). It has been proposed that formation of these multi-meric CESA complexes requires connection of the CESA proteins, which is mediated by two zinc fingers in the N-terminal region of the proteins (43). CESA proteins have been shown to form hetero- and homo-dimers through these zinc fingers (43).
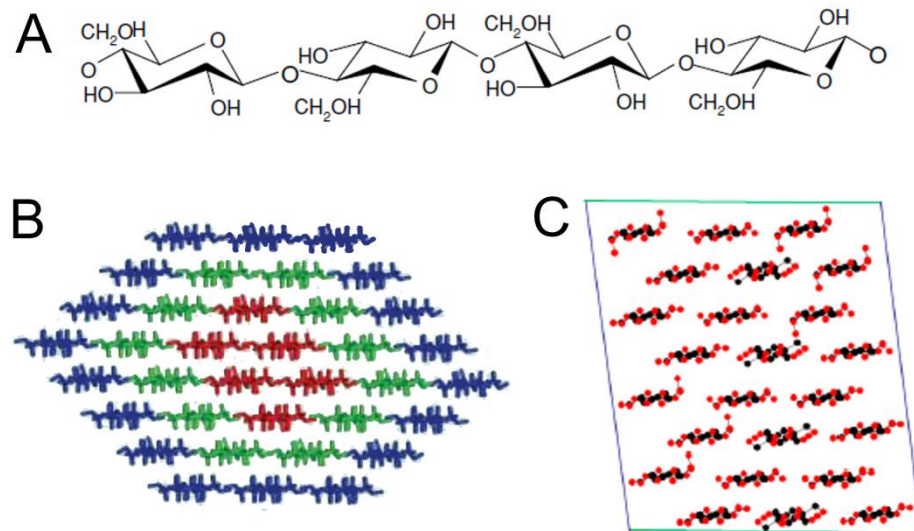
Figure 3: Structure of cellulose in the plant cell wall. A: structure of the β-1,4 glucan chains that make up the cellulose microfibrils, showing the alternating glucose residues rotated 180 °. B and C: two different models for the number and organisation of glucan chains within a cellulose microfibril. B: the 36 chain model in which the glucan chains are organised in a hexagonal configuration. This model was largely accepted until recently challenged by Fernandes et al. (39). Diagram taken from Himmel et al. (44). C: the more recent 24 chain model in which glucan chains are organised in eight sheets composed of three chains each and form a rectangular configuration. Diagram taken from Fernandes et al. (39).

The synthesis of cellulose also requires KORRIGAN (KOR), a membrane-bound β-1,4 endoglucanase (45). The function of KOR is not completely certain but it has been hypothesised that the endoglucanase may trim the growing glucan chain from a sitosterol-glucoside primer required for initiation of synthesis (46). Alternatively, it may trim the glucan chains to aid crystallization (47). *kor* mutants can synthesise β-1,4 glucan chains but they do not form proper crystalline microfibrils (48). Another protein

involved in cellulose synthesis is COBRA, a glycosyl-phosphatidylinositol (GPI)-anchored protein which is thought to be involved in orientation of cellulose microfibril deposition during cell elongation (49). KOBITO is also thought to be involved in microfibril organisation as *Arabidopsis thaliana* (hereafter, Arabidopsis) plants in which *KOBITO* has been mutated have randomly deposited microfibrils, compared to the parallel microfibrils orientated transversely to the elongation axis seen in wild-type (50). However, the exact function and mode of action of these three proteins, KOR, COBRA and KOBITO is uncertain.

1.2.2 Hemicellulose

Many cellulose microfibrils form the main network of the cell wall together with hemicellulose, which can be xylans, xyloglucans, mannans, glucomannans and β-(1-3,1-4)-glucans (also known as mixed linkage glucan (MLG)). Hemicellulose molecules generally consist of a β-1,4-linked backbone and have substantial side chains which prevent them from forming crystalline structures (27). Instead they form long chains that associate with cellulose by extensive hydrogen bonding (51-53). Hemicellulose forms a cohesive and strengthening network with cellulose by holding the microfibrils together but it can also act as a plasticiser to allow extensibility by keeping the fibrils apart from each other (54-56). The class, branching and abundance of hemicelluloses can vary widely between different species and cell types. The major hemicellulose in dicot primary cell walls is xyloglucan, making up 20 - 25% of the cell wall (54). The general structure of this xyloglucan is a β-1,4-glucan backbone with approximately 75% of the glucosyl residues being substituted with a single xylosyl residue at C-6. Some of these xylosyl residues are substituted with galactosyl or arabinosyl residues at the C-2 position and some of the galactosyl residues are substituted with fucosyl residues at the C-2 position (Figure 4A) (27, 57). Xyloglucan, however, only makes up a minor part of the cell wall in grasses (16) and the structure is also different, with much less branching, no fucosyl or arabinosyl side chains and only a small number of galactosyl substitutions

(27, 57). In dicot secondary cell walls the major hemicellulose is glucuronoxylan (GX), which has a xylan β-1,4-linked backbone with common substitutions of α-1,2 glucuronic acid (GlcA) and 4-*O*-methyl glucuronic acid (MeGlcA) residues and makes up 20 - 30% of the cell wall (Figure 4B) (54).

Alternatively, the major hemicellulose of grass cell walls is glucuronoarabinoxylan (GAX), making up 20 - 40% of primary cell walls and 40 - 50% of secondary cell walls (54). GAX consists of a xylan backbone with many substitutions of arabinose residues but also side chains of galactosyl, GlcA, MeGlcA, xylose and ferulic acid residues which can be on 10 - 90% of the backbone xylosyl residues (Figure 4C) (27). The hemicellulosic polysaccharide MLG is unique to Poales (e.g. grasses and cereals) and consists of β-linked glucosyl residues, with approximately 70% being 1,4-linked and 30% being 1,3-linked (27). It is dominated by strings of three or four 1,4-linked glucosyl residues followed by a single 1,3-linkage (Figure 4D). However, other structures have been shown to be present, including two, three or four continuous 1,3-linkages, blocks of more than four 1,4-linked glucosyl residues, and regions that have alternating 1,3- and 1,4-linkages (58). Interestingly the amount of MLG in the cell wall seems to be heavily growth-dependent; when cells are expanding MLG content is high (10 - 30%) but when growth terminates the MLG is broken down and occurs at a much lower percentage (0 – 4%) (16, 59, 60).

Mannans and glucomannans generally make up just a small proportion of the cell wall, especially in monocots (16). In dicots they make up about 3 – 10% of the cell wall and glucomannans are able to cross-link to the cellulose network (16, 27). The mannans and glucomannans also play a role in energy storage and hydration resistance in the endosperm of seeds in a number of plants (61).
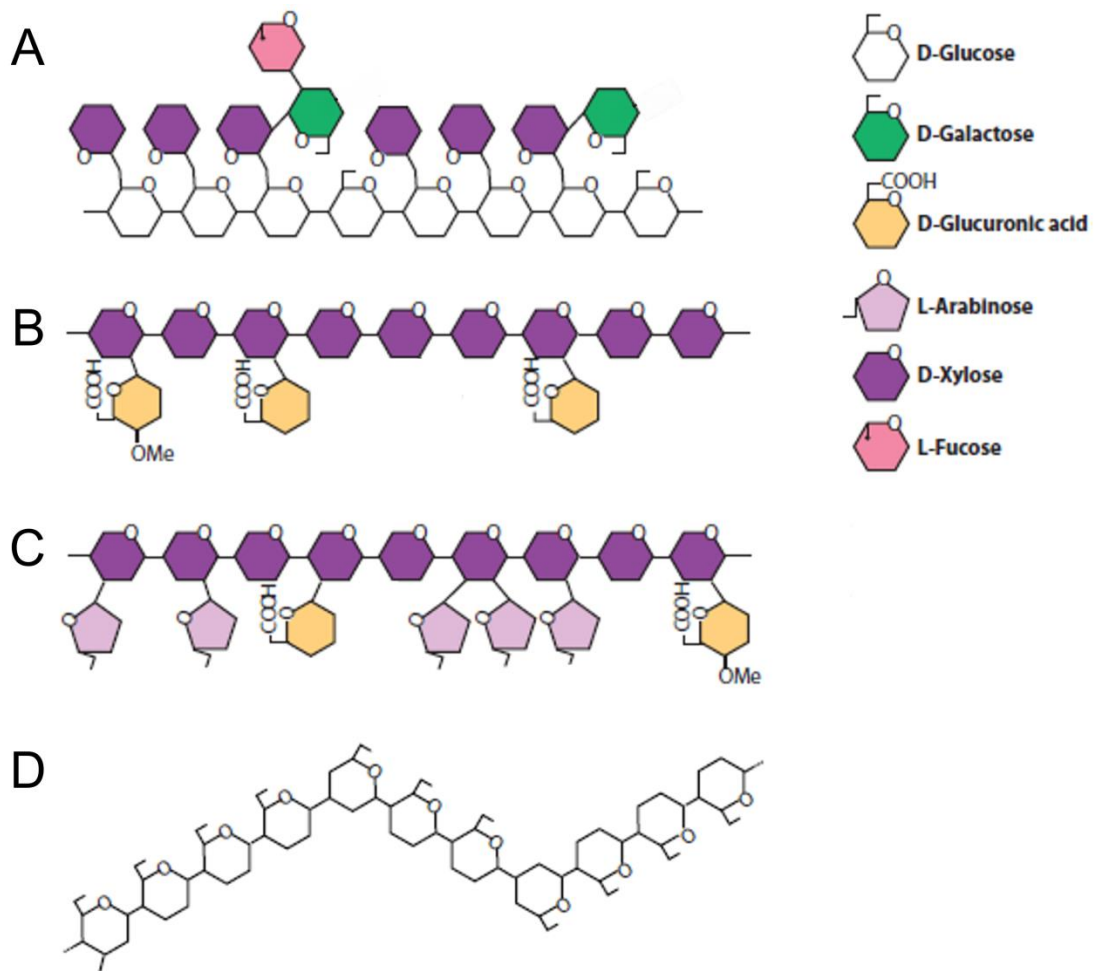
Figure 4: General chemical structure of the most common types of hemicellulose in plant cell walls. A: xyloglucan; B: glucuronoxylan (GX); C: glucuronoarabinoxylans (GAX); D: mixed linkage glucan (MLG). Diagram adapted from Scheller and Ulvskov (54)

1.2.2.1 Synthesis of the hemicellulose backbone

Hemicelluloses, unlike cellulose, are synthesised in the Golgi and travel to the membrane in vesicles (62, 63) (except for MLG which appears to be synthesised at the plasma membrane (64)). The backbones are synthesised by CELLULOSE SYNTHASE-LIKE (CSL) proteins, of which there are eight families named CSLA to CSLH. These proteins contain sequence motifs characteristic of β-glycosyltransferases, without the zinc fingers found on CESA (48). It is possible that the CSL proteins may interact with other glycosyltransferases in order to mediate the simultaneous synthesis

of the backbone and side chains of hemicelluloses (27). For example, the enzymes involved in synthesis of the glucan backbone and xylosyl residue substitution of xyloglucan are believed to act in a multi-enzyme complex. This is because when Golgi membranes were fed with UDP-glucose only, no elongation of xyloglucan molecules was observed, whereas when both UDP-glucose and xylose were added, substantial elongation did occur (65). Galactose and fucose residues however were shown to be incorporated onto the xyloglucan molecule when the molecule was not elongating, suggesting that these residues can be added independently of chain elongation and that the enzymes responsible for their substitution are not part of the multi-enzyme complex (66, 67).

An exception to the role of CSL enzymes in the synthesis of hemicellulose backbones is the β-1,4-xylan backbone, which is thought to be synthesised by glycosyltransferases in the GT43, GT8 and GT47 families. Specifically, IRX9/IRX9L and IRX14/IRX14L, family GT43, and IRX10/IRX10L, family GT47, are responsible for elongation of the xylan backbone (67-73). IRX9 and IRX14 act non-redundantly (73-75). Alternatively, IRX7/IRX7L from the GT47 family and IRX8 and PARVUS from the GT8 family have been implicated in synthesis of the tetrasaccharide at the reducing end of some dicot xylans (69, 76).

1.2.2.2 Side chain substitution of xylans

As discussed above, hemicelluloses can contain a large number of side chain substitutions, making their synthesis complicated, with a large number of enzymes involved. The main substitutions of GX and GAX, the major hemicelluloses of dicot and grass secondary cell walls, and the enzymes involved in their addition to the backbone will be discussed below.

1.2.2.2.1 Acetylation

A number of hemicelluloses are substituted with *O*-linked acetate and the degree of substitution varies between species and polymer. Dicot xyloglucan can be *O*-acetylated on the galactosyl side chains, while grass xyloglucan can be *O*-acetylated on the glucosyl backbone residues (77). The backbone of GAX and GX can also be *O*-acetylated, at the O-2 or O-3 position, on up to 55% of the xylosyl residues (77, 78). Recent work has found that a family of proteins with homology to proteins involved in acetylation in bacteria and fungi are likely involved in cell wall polysaccharide acetylation in plants. These proteins were named REDUCED WALL ACETYLATION (RWA) after the phenotype of mutants of this protein family in Arabidopsis (79-81). The mutants displayed reduced acetylation of pectins, xyloglucans and xylans and did not show complete loss of acetylation, suggesting that these enzymes are not polysaccharide-specific and that they act redundantly (79). However, closer inspection revealed that acetylation was reduced more in certain polysaccharides when different RWA proteins were knocked out, suggesting that they do have a certain amount of substrate preference. RWA2 appears to be more important for xyloglucan acetylation, whereas RWA3 and 4 appear to be more important for xylan acetylation (81).

A second protein, ESKIMO1 (ESK1), has also been shown to be involved in acetylation of xylan. Mutants of ESK1 in Arabidopsis showed a large reduction in *O*-acetylation of xylan (82, 83). Furthermore, measurement of acetylation activity in microsomes from wild-type and mutant plants revealed that incorporation of acetyl groups onto a xylooligosaccharide acceptor was reduced to 20% of wild-type in the mutant (82). The level of acetylation of other cell wall polysaccharides was not affected in the *esk1* mutant, suggesting that ESK1 acts specifically as a xylan acetyltransferase. However, ESK1 occurs as part of a large family of trichome birefringence-like (TBL) proteins which, in Arabidopsis, contains 46 members (82). This may suggest that different members of the TBL family are involved in the specific acetylation of different

polysaccharides. In fact, another member of the TBL family in Arabidopsis, ALTERED XYLAN4 (AXY4), has been implicated in the acetylation of xyloglucan. Knock-out mutants of *AXY4* completely lack *O*-acetylation of xyloglucan, whereas other polysaccharides are not affected (84).

In fungi, humans and some bacteria, the protein responsible for acetylation contains multiple transmembrane domains and a large loop that is predicted to face the Golgi lumen (85-87). It has been predicted that the transmembrane region is involved in translocation of the acetyl donor molecule across the membrane, while the loop is responsible for transfer of the acetyl group to the polysaccharide (88). However, in other bacteria there appears to be two separate proteins that act together to acetylate polysaccharides, one which contains multiple transmembrane domains and the other that contains a large loop (89, 90). Given the structures and the specificities of the RWA and TBL proteins, it seems likely that this two-protein system also occurs in plants; RWA would be responsible for the translocation of acetyl donor and TBL would be responsible for the polysaccharide-specific acetyl group transfer (77, 84). The location of cell wall polysaccharide acetylation is likely to be the Golgi, where the hemicellulose backbone is synthesised. This is because acetylated xyloglucan can be isolated from microsomes (91), radio-labelled acetate is found on polysaccharides when microsomal preparations are fed with radio-labelled acetyl-CoA (92), and RWA and ESK1 have been localised to the endomembrane system (80, 82).

1.2.2.2.2 (Methyl) Glucuronic acid

The backbone of xylans in the cell wall has also been shown to be decorated with α-1,2-linked GlcA and MeGlcA residues, collectively known as [Me]GlcA. These substitutions occur on about 10% of the xylosyl backbone residues of the GX of dicots but are generally less common on arabinoxylan in grasses (93). In Arabidopsis, three glycosyltransferases in the GT8 family have been implicated in [Me]GlcA substitution,

named GLUCURONIC ACID SUBSTITUTION OF XYLAN (GUX)1, 2 and 3 (94, 95). T-DNA knockouts of the genes individually resulted in a large reduction in both GlcA and MeGlcA, while a triple knockout resulted in complete loss of [Me]GlcA (94, 95). Furthermore, the GUX1, 2 and 3 proteins were shown to transfer GlcA residues from UDP-GlcA to a xylooligomer when expressed in tobacco cells (95). An additional set of genes, named *GLUCURONOXYLAN METHYLTRANSFERASE* (*GXM*)*1*, *2* and *3*, are thought to be involved in methylation of the GlcA residues (96). Mutants of these genes were shown to have a large reduction in MeGlcA but no change in the amount of GlcA substitution, while recombinant GXM proteins were able to transfer a methyl group on to GlcA-substituted xylooligomers (96). Both GUX and GXM proteins are localised to the Golgi (94, 96). The enzymes responsible for the addition of [Me]GlcA to xylans in other plant species, including grasses, have not been determined but it seems likely that they may be homologues of the GUX and GXM proteins.

1.2.2.2.3 Arabinose and ferulic acid

Arabinose residues can attach to C-3 or C-2 of the xylosyl backbone residues of xylans, via α-1,3 or α-1,2 linkages (54). It has recently been shown that members of the GT61 glycosyltransferase family are responsible for addition of these arabinose residues, named XYLAN ARABINOSYLTRANSFERASES (XATs) (97). *xat* RNAi lines in wheat were shown to have a large reduction in arabinose substitution of GAX, while heterologous expression of rice and wheat XATs in Arabidopsis, which lacks arabinosyl substitution of xylan in secondary cell walls, was shown to introduce these substitutions (97). XATs have been localised to the Golgi (97) and XAT activity has been found in Golgi membranes (98). Interestingly, the *xat* RNAi lines did not show an increase in unsubstituted xylose residues, but rather a general decrease in arabinoxylan. This suggests that arabinosyl substitution is in some way required for elongation of the xylan backbone (97). Enzymes in the GT47 and GT77 families have been shown to introduce

arabinosyl substitutions on to cell wall polysaccharides other than xylan, suggesting that this activity has evolved multiple times (99-102).

In grasses, the α-1,3 linked arabinose residues on arabinoxylan can be further substituted with ferulic acid via an ester linkage between the C5-hydroxyl of the arabinose and the carboxylic acid group of the ferulic acid (103). The role of ferulic acid in the cell wall will be discussed later. The specific enzymes, location and donor substrate involved in the feruloylation of GAX are not yet fully understood. The most likely donor substrate appears to be feruloyl-CoA. In maize cultures fed with radioactive cinnamic acid, the radioactivity was first incorporated into feruloyl-CoA and then into feruloyl-polysaccharides (104). Furthermore, when enzyme extract from rice cell suspension cultures was incubated with feruloyl-CoA and arabinoxylan-trisaccharide (AXX), feruloylated arabinoxylan was formed (105). A bioinformatics approach comparing the expression of genes between dicots and grasses proposed genes in a sub-clade of the BAHD acyl-CoA transferase superfamily, also known as Pfam family PF02458, to be involved in GAX feruloylation (106), consistent with the role of feruloyl-CoA as a substrate. The BAHD superfamily is named after the first four members of the family to be biochemically characterised (BEAT: benzylalcohol acetyltransferases, AHCT: anthocyanin hydroxycinnamoyl transferase, HCBT: anthranilate hydroxycinnamoyl/benzoyl transferase, DAT: deactylvindoline acetyltransferase) (107). RNAi of four members of the BAHD family in rice led to a significant reduction in ferulic acid content (108). More recently, members of the BAHD family have been experimentally validated more specifically in the transfer of hydroxycinnamic acids to arabinose (109).

It is not certain whether the BAHD enzymes are involved in the direct feruloylation of arabinoxylan, or whether these enzymes add ferulate to a precursor, such as UDP-arabinose, and the feruloyl-arabinosyl unit is subsequently added to the xylan backbone by a different enzyme. The Mitchell group believe the latter is the case (110) as BAHD

proteins are predicted to be localised to the cytosol (107), whereas GAX is synthesised in the Golgi. Furthermore, feruloylation of arabinose on a synthetic molecule was shown to occur in the cytosol and not in the membrane fraction of rice seedlings (105). Although UDP-arabinopyranose is synthesised in the Golgi, this is converted to UDP-arabinofuranose, the substrate of arabinosylation of GAX, in the cytosol (111). The feruloylated UDP-arabinose would subsequently be transferred back to the Golgi for substitution on to the xylan backbone (110). Most evidence suggests that feruloylation of GAX occurs in the cytoplast, as cell-suspension cultures have been shown to form ferulate-arabinose linkages long before the polysaccharides are transported to the cell wall (112, 113). However, it has been demonstrated that feruloylation can also occur within the cell wall because root segments in which the secretory system was blocked still incorporated feruloylated polysaccharides in to the cell wall (114).

The feruloyl-arabinosyl residues can often be further substituted with β-xylose, attached to O-2 of the arabinose (115, 116). It has recently been proposed that a member of the GT61 glycosyltransferase family is responsible for addition of this xylose, due to absence of this substitution in a loss-of-function mutant of the gene (117). The authors named this gene *XYLOSYL ARABINOSYL SUBSTITUTION OF XYLAN1* (*XAX1*). Interestingly, the *xax1* mutant also showed a large reduction in ferulic acid content, leading the authors to believe that xylosylated arabinose acts as the acceptor for ferulic acid substitution (117).

1.2.3 Pectin

Pectins are also polysaccharides but differ from hemicelluloses in that they contain relatively high proportions of D-galactosyluronic acid. They make up around 20 - 35% of the primary cell walls of dicots and non-graminaceous monocots but only a small proportion of grass primary cell walls (~5%) and a very minor (~0.1%) proportion of secondary cell walls of both dicots and grasses (16). As lignocellulosic biomass is

mainly composed of secondary cell walls, pectins will only briefly be discussed. There are three types of pectic polysaccharide in plant cell walls: homogalacturonan, rhamnogalacturonan I (RG-I) and rhamnogalacturonan II (RG-2). Homogalacturonan consists solely of 1,4-linked α-D-galactosyluronic acid and up to 70% of these residues are methyl esterified. Rhamnogalacturonan I has a backbone of a disaccharide repeating unit consisting of α-D-galactosyluronic acid and α-L-rhamnose which are 1-4 and 1-2 linked. Approximately 50% of the rhamnosyl residues are substituted at the C-4 position with side chains containing arabinosyl, galactosyl and, to a lesser extent, fucosyl and glucosyluronic acid residues. RG II has the same 1,4-linked α-D-galactosyluronic acid backbone as homogalacturonan but has more complex side chains consisting of 12 different sugars with over 20 different linkages (118). In general, pectin in plant cell walls consists of around 65% homogalacturonan, 20 - 30% rhamnogalacturonan I and around 10% rhamnogalacturonan II, with substantial cross-linking between the structures (118). When cellulose produced from the bacterium *Acetobacter* is grown in the presence of pectin, the cellulose is less stiff and more extensible. Interestingly, these properties of the cellulose remain even when the pectin is removed. This suggests that the pectin may aid the deposition of the cellulose microfibrils within the cell wall, acting as a spacer and preventing the microfibrils from forming large aggregates with each other (119). There is also evidence that pectins may covalently link to hemicelluloses in the cell wall (120, 121). Pectins are thought to be synthesised in the Golgi by a large number of glycosyltransferases, methyltransferases and acetyltransferases and then transported to the cell wall (reviewed in 118).

1.2.4 Structural proteins

Similar to pectins, structural proteins are present in dicot primary cell walls (~10%) but only present in minor amounts in the monocot primary cell wall and the secondary cell walls of both monocots and dicots (16). The role of these proteins in the cell wall will therefore only be briefly discussed. There are five main classes of structural proteins:

extensins (approximately 45% protein and 55% sugar), glycine-rich proteins (GRPs) (100% protein), proline-rich proteins (PRPs) (80 – 100% protein, 0 – 20% sugar), solanaceous lectins (approximately 55% protein and 45% sugar), and arabinogalactan proteins (AGPs) (92 - 10% protein, 2 – 10% sugar) (122). Extensins are thought to play a role in strengthening the cell wall and creating a barrier against pathogen infection as they have been shown to accumulate in the cell wall and cross-link in response to ethylene, mechanical wounding and pathogen invasion (123-125). They are thought to cross-link to each other and also to pectins in the cell wall (122, 126, 127). Similarly, PRP, GRP and solanaceous lectin gene expression has been shown to be upregulated in response to wounding (122). AGPs have been shown to be abundant in the middle lamella and to be membrane linked by a GPI (128). It is therefore possible that AGPs play a role in cell-cell recognition or cell signalling (122, 129).

1.2.5 Lignin

In secondary cell walls, once cell expansion has been completed, the polysaccharide network is impregnated and coated by lignin. This provides rigidity and strength to the cell wall and also acts as a barrier to pests and pathogens (130, 131). Lignin is secreted as monomers (monolignols) into the cell wall, where it is polymerised by free radical chemistry rather than an enzyme adding repeat structures. There is some debate over whether there is any biological control over this process (132) but the general view is that it is a strictly chemical, combinatorial process (133). This makes lignin a complex compound in which the bonds between the monomers can form at multiple positions, making it difficult for enzymes to digest (Figure 5) (133). The molecular weight of lignin has been estimated to be around 20,000 to 77,000, suggesting that each lignin molecule is made up of hundreds of monomers (27). Furthermore, lignin is a racemic polymer and able to form multiple isomers, meaning that lignin molecules are likely to never be identical (134). Lignin degraders therefore need a whole suite of different enzymes in order to break down this complex structure.
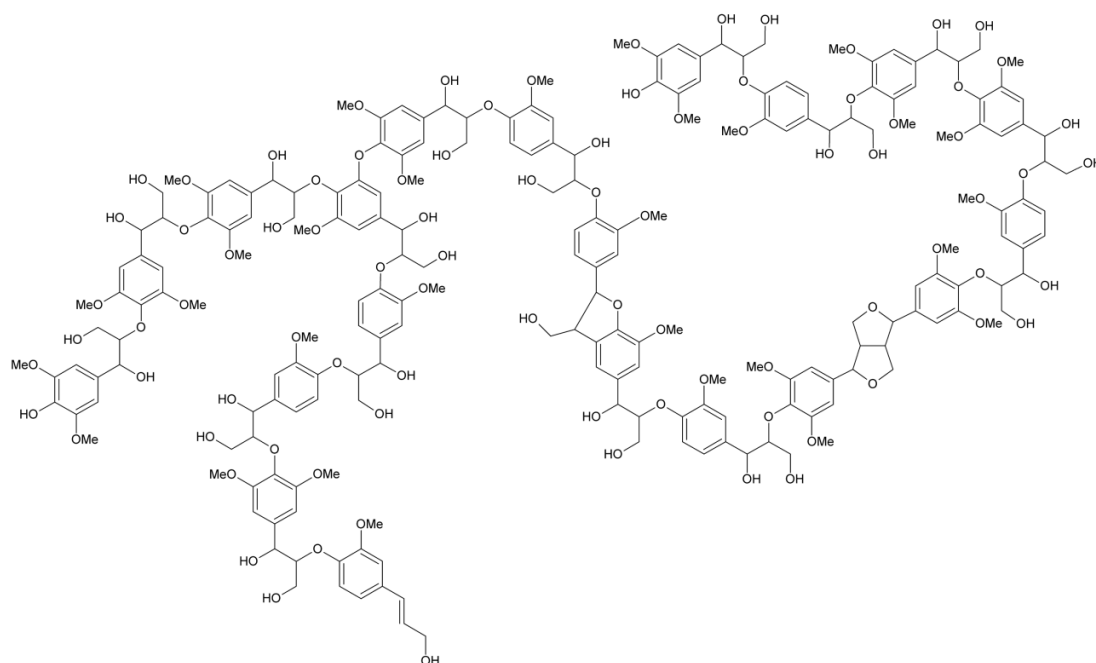
Figure 5: A hypothetical structure of a section of lignin. The model shows the multiple different bonds that can occur in the lignin polymer, resulting in a complex structure. Diagram taken from Albersheim et al. (27).

Lignin is synthesised by the polymerisation of cinnamyl alcohol monomers, known as monolignols. There are three major monolignols in plant cell walls, namely *p*-coumaryl, coniferyl and sinapyl alcohols, which are characterised by the number of methoxy side groups on the phenolic ring (*p*-coumaryl = 0, coniferyl = 1, sinapyl = 2) (Figure 6). When these monolignols are incorporated into the lignin molecule they are known as *p*-hydroxyphenyl (H), guaiacyl (G) and sinapyl (S) units, respectively (133). The monolignol biosynthesis pathway is reasonably well characterised; they are synthesised in the cytosol from phenylalanine through the phenylpropanoid pathway, via a number of different enzymes (133) (Figure 6). The phenylalanine is first deaminated by phenylalanine ammonia-lyase (PAL) to produce cinnamic acid, which is then hydroxylated at the C-4 position of the aromatic ring by cinnamate-4-hydroxylase (C4H) to produce *p*-coumaric acid. A CoA group is then added to the acid at the terminal carbon (γ) of the propanoid side chain by 4-coumarate: CoA ligase (4CL) to form *p*-coumaroyl-CoA. A combination of hydroxycinnamoyl transferase (HCT) and *p*-

35

coumarate 3-hydroxylase (C3H) are thought to convert *p*-coumaroyl-CoA to caffeoyl CoA; HCT takes the substrates *p*-coumaroyl-CoA and shikimic acid and converts *p*-coumaroyl-CoA to *p*-coumaroyl shikimate, C3H then hydroxylates *p*-coumaroyl shikimate to caffeoyl shikimate and HCT then acts in reverse to convert caffeoyl shikimate to caffeoyl-CoA and shikimic acid. Alternatively, in Arabidopsis it has been proposed that another enzyme, caffeoyl shikimate esterase (CSE), replaces the reverse activity of HCT by converting caffeoyl shikimate to caffeic acid (135). 4CL would then add a CoA group to the caffeic acid to produce caffeoyl-CoA. The role of CSE has not yet been investigated in other plant species. Caffeoyl-CoA is then *O*-methylated at the C-3 position of the aromatic ring by caffeoyl coenzyme A *O*-methyltransferase (CCoAOMT) to form feruloyl-CoA. The enzyme cinnamoyl-CoA reductase (CCR) can act on both *p*-coumaroyl-CoA and feruloyl-CoA to produce *p*-coumaraldehyde and coniferaldehyde respectively. Coniferaldehyde can then be further hydroxylated at the C-5 position of the aromatic ring by ferulate 5-hydroxylase (F5H) to produce 5-hydroxy-coniferaldehyde, which is then further *O*-methylated by caffeic acid *O*-methyltransferase (COMT) to produce sinapaldehyde. *p*-coumaraldehyde, coniferaldehyde and sinapaldehyde can then all be reduced by cinnamyl alcohol dehydrogenase (CAD) to form the corresponding alcohols, i.e. the monolignols. It has been suggested more recently that C3H and C4H may act as a multi-protein complex for the hydroxylation of cinnamic acid and *p*-coumaroyl shikimate (136). It has also been hypothesised that an additional enzyme, sinapyl alcohol dehydrogenase (SAD), performs the specific reduction of sinapaldehyde to sinapyl alcohol, due to immunolocalisation of this enzyme to cells with S unit-rich lignin and because SAD uses sinapaldehyde preferentially to other CAD substrates in aspen (137). However, it was shown more recently that downregulation of SAD in *Nicotiana tabacum* had no impact on lignin content or composition, whereas additional downregulation of CAD caused a large decrease in S units, suggesting that CAD is responsible for the reduction of sinapaldehyde (138).

Figure 6: Lignin monomer biosynthesis pathway.

The dotted arrows are suggested conversions and have not been proven. PAL: phenylalanine ammonia-lyase; C4H: cinnamate-4-hydroxylase; 4CL: 4-coumarate: CoA ligase; HCT: hydroxycinnamoyl transferase; C3H: *p*-coumarate 3-hydroxylase; CSE: caffeoyl shikimate esterase; CCoAOMT: caffeoyl coenzyme A O-methyltransferase; CCR: cinnamoyl CoA reductase; F5H: ferulate 5-hydroxylase; COMT: caffeic acid *O*-methyltransferase; CAD: cinnamyl alcohol dehydrogenase.

Once the monolignols are formed, they are secreted into the cell wall. There is currently debate over the mode by which they are transported across the plasma membrane, namely exocytosis via endoplasmic reticulum (ER) vesicles, passive diffusion or transporter-mediated active transport. The most convincing evidence has been presented for transport via active transporters. Firstly, a number of transcriptomic and proteomic studies have shown that membrane transporters are highly expressed in lignifying tissues (139-141). Furthermore, transport of monolignols across isolated plasma and vacuolar membranes from Arabidopsis and poplar was shown to be greatly reduced when ATP-binding cassette (ABC) transporter inhibitors were added and also highly dependent on the presence of ATP, suggesting that the transportation process requires energy (142). Recently, an ABC transporter has been shown to transport *p*-coumaroyl alcohol across the plasma membrane (143).

Once the monolignols have reached the cell wall they are polymerised by oxidative polymerisation. This involves enzymatic dehydrogenation/oxidation of the monolignols, resulting in the formation of a free radical that can then couple with another monolignol or a lignin oligomer/polymer (27). Polymerisation mainly occurs via the addition of monomers to a growing polymer, although linkages between two oligomers or polymers does also occur (27). The mechanism of polymerisation is not fully understood but it is thought to be catalysed by peroxidases and/or laccases due to co-localisation of these enzymes with lignifying tissues, both enzymes being able to oxidise monolignols in vitro, and lignin being polymerised in plant tissue after additions of H_2O_2 (144-150). Furthermore, downregulation of both peroxidases and laccases has been shown to reduce lignin content in plants (151-153). It has been proposed that peroxidases could oxidise the monolignols directly which, in turn, could transfer the radical to the lignin polymer or couple with it (depending on the oxidation state of the lignin polymer) (154). Alternatively, peroxidases may oxidise manganese which could then oxidise both monolignols and the lignin polymer (155).

Two main linkages, β-aryl ether (β-*O*-4) and carbon-carbon (C-C), can form between the monolignols during polymerisation of the lignin molecule. The β-*O*-4 linkages (Figure 7B) are the most abundant and are more easily cleaved; they make up about 80% of the linkages in angiosperms and 50% in gymnosperms. The C-C linkages consist of β-5, β-β, 5-5 and β-1 linkages, and a less common linkage is 4-*O*-5 (Figure 7B) (156). The number in the monolignol linkage nomenclature refers to the numbered carbons of the phenolic ring and the 'β' refers to the carbons on the propane side chain (Figure 7A). Abundance of the different linkages depends partly on the ratio of the different monolignols. For example lignin rich in S units have elevated concentrations of β-*O*-4 linkages and low concentrations of the β-5, 5-5 and 5-*O*-4 linkages because the C-5 position is not available for coupling due to the extra methoxy group (156).

Figure 7: Linkages within the lignin polymer.
A) The basic monolignol structure with the carbon atoms labelled. B) Major linkages that occur between monolignols within the lignin polymer.
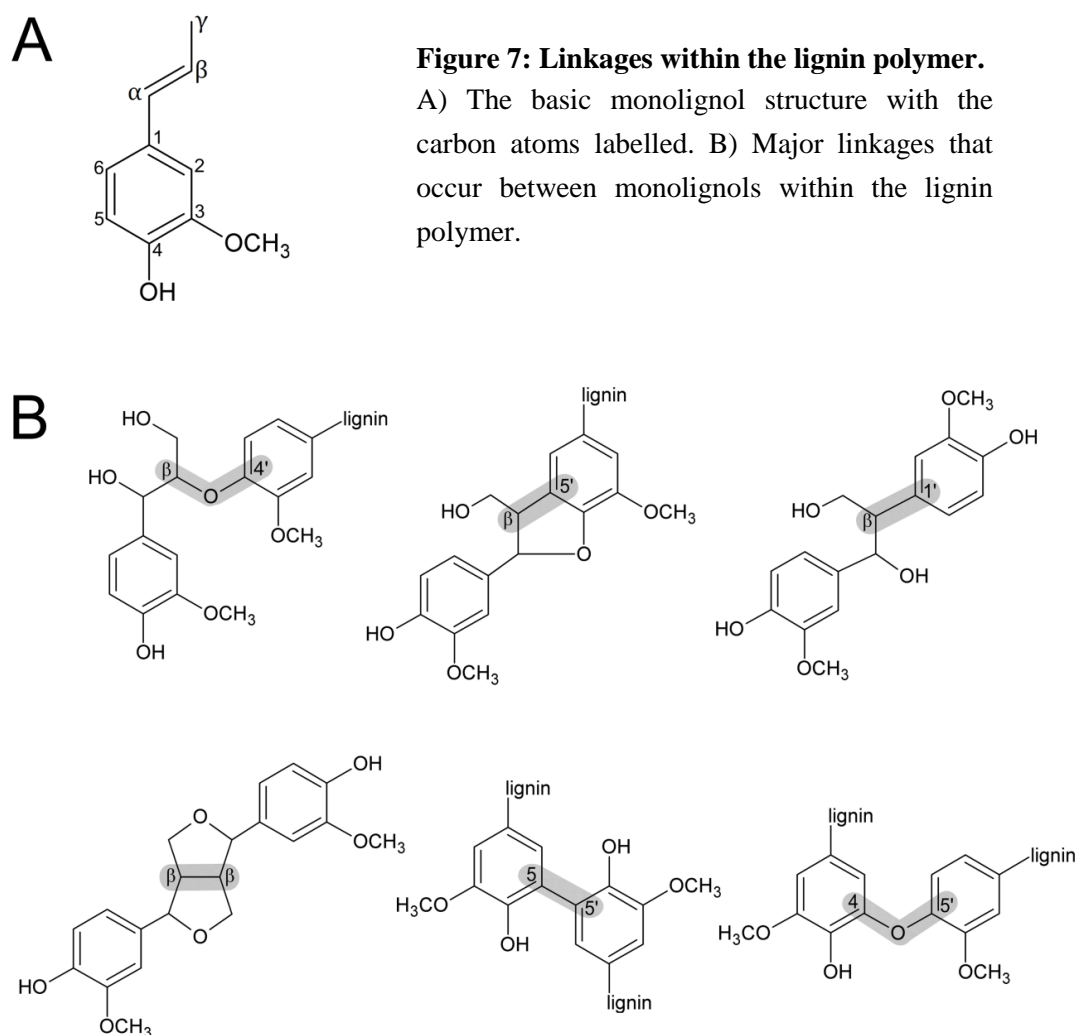
The composition of lignin varies widely between species and cell type, as with hemicelluloses. For example, gymnosperms generally contain lignin composed mostly of G monomers, with a small amount of H monomers and very little or no S monomers. In contrast, in angiosperms G and S units make up the majority of lignin, with H units being present at much lower levels (157). The difference between monocots and dicots is seen in the proportion of H units, which make up 4 - 15% of monocot lignin but are only found in trace levels in dicot lignin (16). As well as the three main monolignols, many other monomers can be incorporated into the lignin polymer, although normally in small amounts. These monomers include both intermediates of the monolignol synthesis pathway and derivatives of the main monolignols. The main compounds found to be incorporated are hydroxycinnamaldehydes, hydroxybenzaldehydes, 5-hydroxyconiferyl alcohol, dihydrocinnamyl alcohols, arylpropane-1,3-diol and hydroxycinnamoyl acetates (Figure 8) (156).
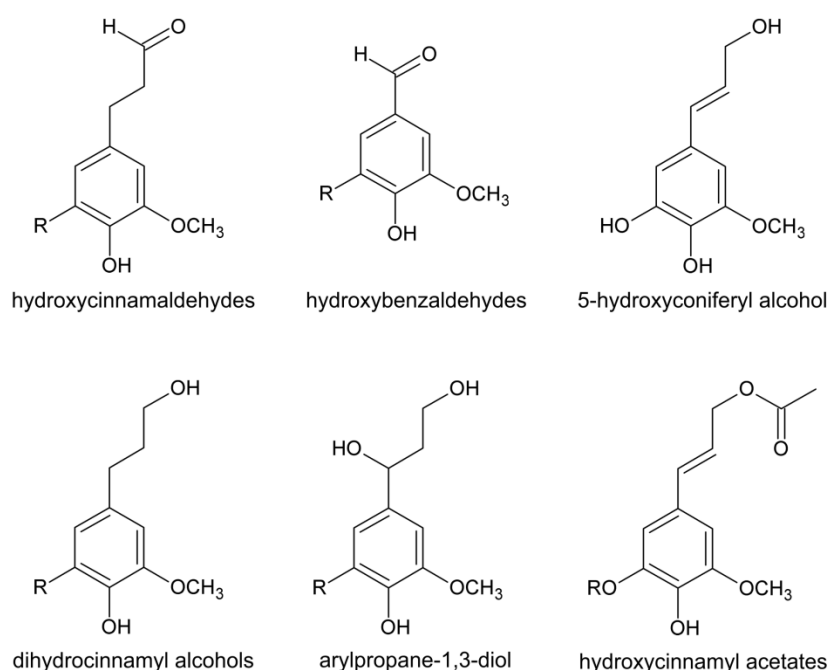
Figure 8: Non-traditional lignin monomers. Intermediates of the monolignol synthesis pathway and derivatives of the three main monolignols can be incorporated into the lignin polymer. The structures of the main non-traditional lignin monomers are shown here.

1.2.6 Ferulate esters

As mentioned earlier, ferulic acid, an intermediate of the phenylpropanoid pathway (Figure 6), can be found as a side chain on the arabinose residues of arabinoxylan in grasses. These ferulates make up $0.5 - 5\%$ of the cell wall (16) and can form dimers via peroxidase-mediated oxidation, which results in cross-linking of hemicelluloses (Figure 9A) (158, 159). Ferulic acid can also form a covalent bond via an ether linkage with lignin, creating a cross-link between hemicellulose and lignin (Figure 9B) (160). It has been proposed that GAX-bound ferulic acid could in fact act as a nucleation site for lignification. This is because the dominant bond that forms between ferulic acid and lignin is only able to form if the lignin unit has a free phenol and α,β-unsaturation. This could be a free monolignol or, less likely, a coniferyl alcohol dimer or a coniferyl alcohol linked to a lignin oligomer via a 5-5 or 5-O-4 bond (161).
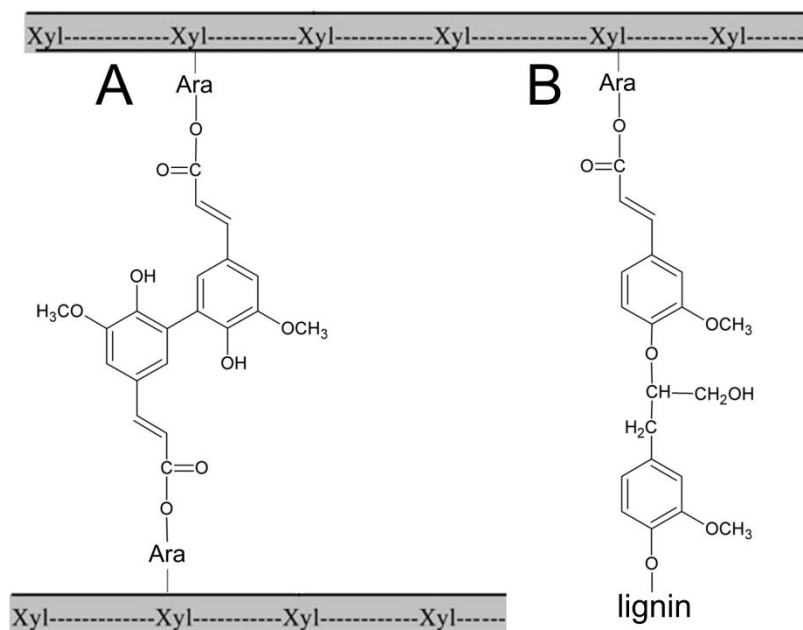
Figure 9: Ferulic acid cross-linking between hemicellulose molecules and between hemicellulose and lignin. A: hemicellulose to hemicellulose cross-link. Ferulic acid is linked to the arabinose side chains of hemicellulose by an ester linkage and the ferulic acid residues dimerise through a 5-5 linkage. B: hemicellulose to lignin cross-link. Ferulic acid is linked to the arabinose side chains of hemicellulose by an ester linkage and the ferulic acid links to a lignin monomer via an ether bond; the ferulic acid residues dimerise through a 5-5 linkage. Diagram adapted from Buanafina (103).

The majority of ferulic acid is deposited into cell walls after cells stop elongating, suggesting that it acts to strengthen the cell wall and prevent elongation (162, 163). Indeed, it has been shown that the amount of ferulic acid bound to cell wall polysaccharides correlated with a reduction in extensibility and an increase in rigidity of the cell wall (164, 165). Studies have also shown that ferulic acid accumulates in the cell wall following pathogen invasion and that ferulic acid content positively correlates with resistance to a number of pathogens (166-169). This provides further evidence that the ferulic acid cross-linkages act to strengthen the cell wall. Ferulic acid only makes up a very minor percentage of dicot cell walls and instead of linking to hemicellulose, links to pectin via ester linkages (170, 171).

1.3 Structure of the plant cell wall

The plant cell wall is made up of multiple layers (Figure 10) which are laid down at different stages of cell development (172). The outermost and first layer to be laid down is the middle lamella. This forms the junction between cells and is largely composed of pectin which plays a role in the adhesion of cells (173, 174). The primary cell wall is the next layer in and is formed at the start of cell differentiation. The cellulose microfibrils in the primary cell wall are orientated randomly (Figure 10) (175, 176). In certain cells, such as xylem vessels, fibre cells, phloem sieve elements and sclereids, after cell elongation has been completed a secondary cell wall is laid down inside the primary cell wall. This secondary wall is composed of three layers: S1, S2 and S3. The cellulose microfibrils within secondary cell walls have a parallel orientation to each other and, interestingly, the orientation is different in the different layers (Figure 10). In the S1 and S3 layers the cellulose microfibrils are almost transverse, orientated 60 - 80 ° to the cell axis. In the S2 layer however the microfibrils are almost longitudinal, orientated 5 - 30 ° to the cell axis (175, 176). This variation in orientation of the microfibrils may be a feature to allow strength in multiple directions (177). The S2 is the thickest layer, being

between 1 and 10 µM thick and accounting for 75 - 85% of the total thickness of the primary wall (178). The S1 and S3 layers are 0.1 - 0.35 µM and 0.5 - 1.1 µM thick respectively, while the middle lamella and primary cell wall are 0.5 - 1.5 µM and approximately 0.1 µM thick, respectively (178).
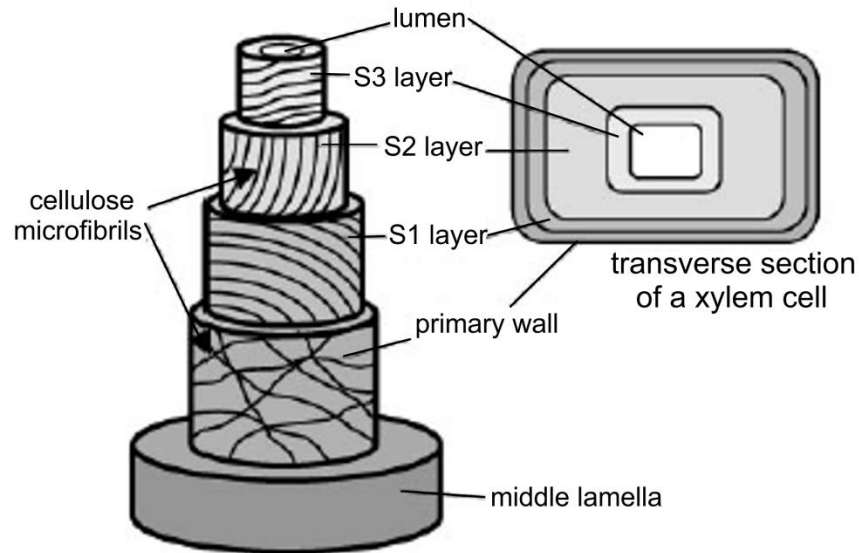
Figure 10: Layers of the plant cell wall. Cell walls are composed of multiple layers, namely the middle lamella, primary wall and, if secondary thickening is present, a secondary wall comprising of the S1, S2 and S3 layers. These layers have different orientations of their cellulose microfibrils. Diagram taken from Plomion et al. (178).

The various cell wall components are deposited into the cell wall in a series (Figure 11). Pectins and hemicelluloses are the first to be deposited, after cell division, and cellulose synthesis follows shortly afterwards, allowing the deposition of the microfibrils within the hemicellulose and pectin networks. Pectin is not deposited in the secondary cell wall, whereas hemicellulose and cellulose synthesis continues, allowing thickening of the cell wall. In dicots, hemicellulose synthesis switches from xyloglucans during primary wall formation, to xylans during secondary wall formation. In grasses, however, the main hemicellulose deposited throughout primary and secondary cell wall formation is xylan. Lignin deposition only initiates once the primary wall has been laid

down and elongation has terminated and it continues after cell death. Typically, lignification starts in the cell corner of the middle lamella and S1 layer and spreads into the primary cell wall and the S2 and S3 layers of the secondary cell wall (179). H monolignols are deposited in to the cell wall first, followed by G and then S units, resulting in a change from H-unit rich lignin in the cell corners to S- and G-rich lignin in the S2 and S3 layers (180).
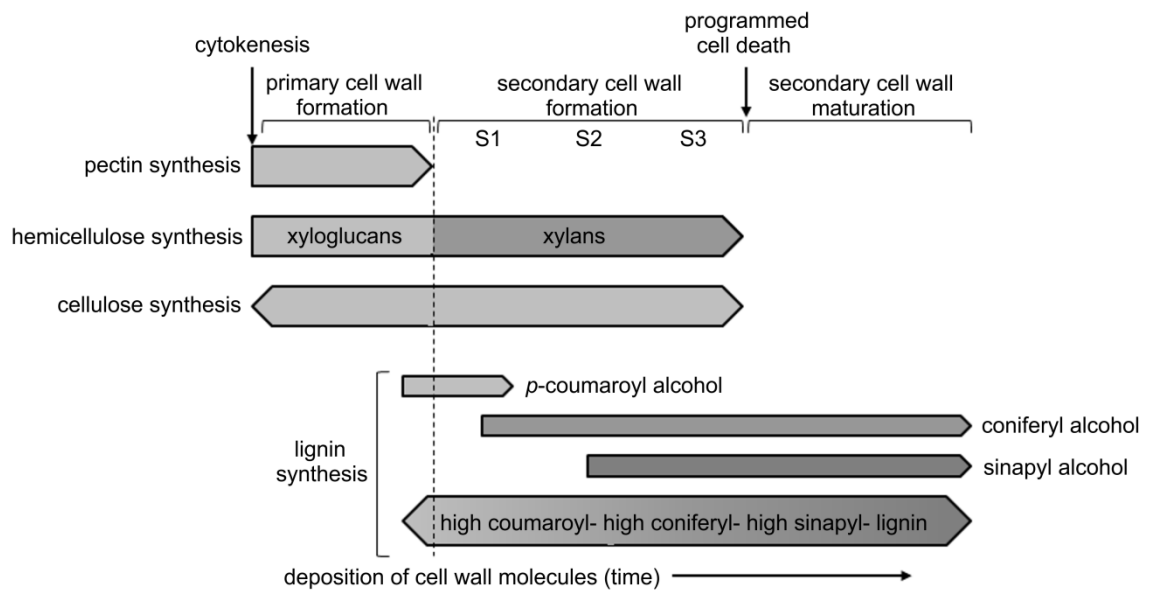
Figure 11: Time course of the deposition of cellulose, hemicellulose, pectin and lignin during cell wall formation. Taken from Albersheim et al. (27).

1.3.1 Interaction between components of the cell wall

Plant cell walls have been likened to two-component fibre composite materials such as ferroconcrete and fibreglass, in which the fibrous elements that are capable of resisting tension are embedded in a more amorphous material which gives the material the flexibility to withstand mechanical forces (27). In plant primary cell walls the cellulose microfibrils form the main structural feature and are responsible for the majority of the strength of the cell wall (181). The cellulose fibres are embedded in a matrix of hemicellulose and pectin in order to increase the extensibility of the cell wall and so

enable enlargement of the cell (48, 182). Experiments with cellulose produced from *Acetobacter xylinus* show that synthesis of the cellulose within medium containing xyloglucan or pectin greatly reduces the stiffness and increases the extensibility of the composite compared to cellulose alone (119, 181).

The exact organisation of the components within the cell wall is not yet fully understood, especially as there is so much heterogeneity between species, cell types and stages of development. However, there are a number of models to illustrate how the cell wall polymers may be arranged. These models have been formed from a combination of electron microscopy, biophysical and mechanical measurements, and chemical information about the cell wall components. Figure 12 shows a model of the plant primary cell wall, based on direct measurements taken from native and sequentially extracted onion cell walls (52). The cellulose microfibrils form layers one microfibril thick, in which the microfibrils are roughly parallel to each other and in a plane parallel to the plasma membrane (52, 183). This is because the microfibrils are formed by adjacent CESA rosettes on the plasma membrane and their deposition is strictly limited to the space between the existing cell wall and the plasma membrane which will be pushing against the cell wall due to turgor pressure (27). The orientation of microfibrils in neighbouring layers varies in primary cell walls, whereas in secondary cell walls the microfibrils tend to all be orientated in the same direction, as described earlier. The cellulose microfibrils are between 3 and 12 nm thick and are spaced approximately 20 - 40 nm apart, both within and between the layers (52). Therefore, a 100 nm thick primary wall could accommodate three to four layers of microfibrils, whereas a secondary cell wall could potentially accommodate hundreds of layers.

The cellulose microfibrils are linked to hemicelluloses through multiple hydrogen bonds (51). Conformational studies of arabinoxylan and xyloglucan show that these molecules can adopt a flat, ribbon-like structure, allowing them to run alongside the cellulose microfibrils (184, 185). It is thought that hemicelluloses form a monolayer coating of

the cellulose microfibrils but that they can also span between and cross-link the microfibrils (Figure 12) (52). This spanning between microfibrils is thought to hold the cellulose microfibrils apart and to act as a control over slippage between microfibrils. This provides resistance to mechanical forces but also allows enlargement of the cell (55, 56). Pectins form a gel-like network that is co-extensive with the cellulose-hemicellulose network. Pectins can cross-link to each other via covalent glycosidic bonds (186, 187) and potentially also to hemicelluloses (120). In secondary cell walls the whole polysaccharide network is coated and impregnated by lignin which is polymerised at the cell wall (133).
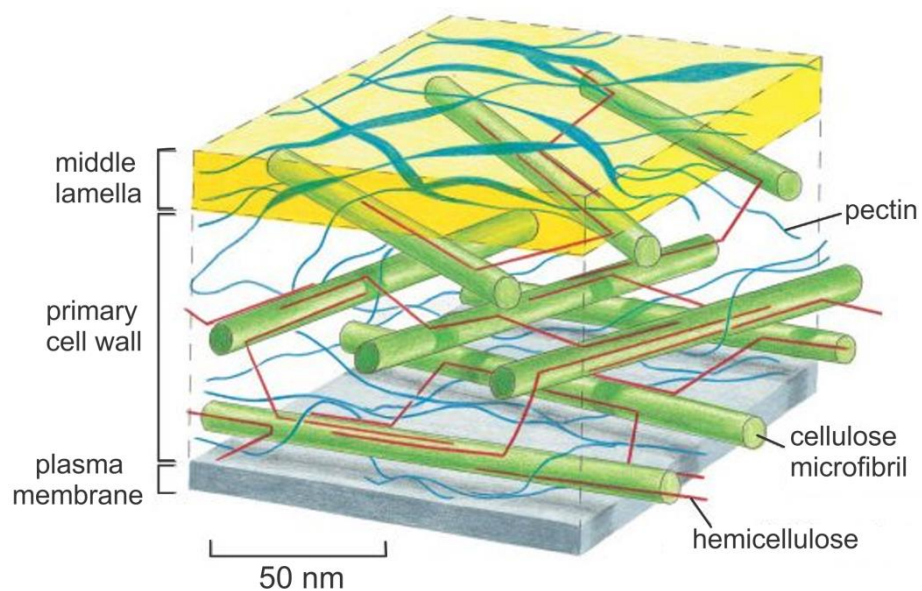
Figure 12: Model of a type I primary cell wall. The model is a simplified representation of how the cellulose, hemicellulose and pectin components are arranged in the cell wall. The size and spacing of the polymers are based on direct measurements of native cell walls. The model shows the layered structure of the cellulose microfibrils and the approximately parallel orientation of the microfibrils within each layer, the attachment of hemicellulose along lengths of cellulose microfibrils through hydrogen bonding and the cross-linking of two or more microfibrils with a hemicellulose molecule. The pectin network is closely associated with the cellulose-hemicellulose network. Taken from McCann and Roberts (188).

1.3.1.1 Type I and II primary cell walls

There are two broadly classified types of primary cell wall, type I and type II, although accumulating evidence suggests that cell walls actually fall into more of a continuum between type I and type II (25). Type I primary cell walls are present in dicots and liliaceous monocots and consist of cellulose, xyloglucan as the main hemicellulose, and pectin in approximately equal amounts (182), as shown in the cell wall model (Figure 12). Type II cell walls are found in Poales and other closely related monocots. They have the same cellulose architecture as type I cell walls but the main hemicellulose is GAX rather than xyloglucan, although small amounts of xyloglucan are often present (182). The GAX can hydrogen bond to and cross-link the cellulose microfibrils in the same way as xyloglucans (189). Type II cell walls also contain very little pectin, although they do contain small amounts of homogalacturonan (190). A highly substituted GAX has been shown to be closely associated with the pectin in grass cell walls (191) and it has been hypothesised that this composite has replaced the pectin network present in the type I cell wall, while the less highly branched GAX acts to cross-link the cellulose network (181). This is because highly branched GAX is extracted by dilute alkaline treatment, while less branched GAX requires more concentrated alkaline for extraction (182). Type II cell walls also contain MLG and ferulic acid linkages, neither of which are present in type I cell walls (discussed above).

1.4 The bioethanol production process

The production of bioethanol from lignocellulosic biomass on an industrial scale generally uses a three-step process: pretreatment, followed by enzymatic hydrolysis of the polysaccharides to monosaccharides, and finally fermentation of the resulting monosaccharides to ethanol. The methods used for these three processes are detailed below.

1.4.1 Pretreatment

The pretreatment step is used to open up and/or breakdown the cell wall structure in order to increase the accessibility of the sugars to enzymes (26). It can consist of a number of different treatments but often involves a milling step followed by a chemical or physico-chemical treatment. The milling step reduces the particle size and, although it has no effect on cell wall structure or sugar extractability itself, increases the surface area of the biomass to make the chemical treatment more effective (192). The chemical or physico-chemical treatment is used to alter the structure of the lignocellulosic composite in order to increase the extractability of the polysaccharides, particularly cellulose, and has been shown to greatly increase the level of saccharification of lignocellulosic material (193-196). The most common chemical pretreatments consist of soaking the biomass in acid or alkali. This is generally carried out at high temperatures (120 – 200°C) for timescales in the range of minutes to days (197). Alkali treatments primarily target and solubilise lignin in the cell wall, whereas acid treatments primarily target hemicellulose (193, 198). Organosolv pretreatments can also be used which employ organic or aqueous solvent mixtures, such as ethanol, methanol, acetone, ethylene glycol or tetrahydrofurfuryl, to solubilize lignin and produce relatively pure lignin as a by-product (199). Physico-chemical pretreatments include steam explosion and ammonia fibre explosion (AFEX). Steam explosion involves heating the lignocellulosic biomass with high-pressure steam (180 – 240°C and 1 – 3.5 MPa) (197). At this high temperature auto-hydrolysis of the acetyl groups on hemicellulose occurs, resulting in the formation of acetic acid and causing solubilisation of the hemicellulose (200). The pressure of the pretreatment is then suddenly reduced which causes explosive decompression and opens up the cell wall structure (201). AFEX uses a liquid anhydrous ammonia treatment at temperatures between 60 and 100°C and high pressure. Again, the pressure is suddenly released which causes rapid expansion of the ammonia as gas and results in swelling and physical disruption of the cell wall structure

(197). AFEX does not solubilise much of the biomass but has been shown to decrystallise cellulose and disrupt lignin-carbohydrate linkages (202).

1.4.2 Enzymatic hydrolysis

After pretreatment, enzymatic hydrolysis is carried out to convert the accessible polysaccharides to monosaccharides (saccharification). The majority of enzymes used industrially predominantly breakdown cellulose into glucose, a process that requires three groups of enzymes that are collectively known as cellulases: endo-β-1,4-glucanases, exo-β-1,4-glucanases and cellobiohydrolases. Endo-β-1,4-glucanases randomly hydrolyse internal β-1,4-glucosidic bonds. They require amorphous parts of the cellulose in order that the bond is accessible for hydrolysis. Exo-β-1,4-glucanases move progressively along β-1,4-glucans and cleave cellobiose units (a glucose disaccharide) from the end of the chain. Cellobiohydrolases then hydrolyse the cellobiose into glucose monomers. They can also cleave glucose from cellooligosaccharides (203).

1.4.3 Fermentation of the monosaccharides to ethanol

The final step of the bioethanol production process is fermentation of the monosaccharides that have been released during enzymatic hydrolysis to produce ethanol. This is carried out by microorganisms and the enzymatic hydrolysis and fermentation steps are now often carried out as a simultaneous reaction (simultaneous saccharification and fermentation (SSF)) (204, 205).

1.4.4 Improving the bioethanol production process

There are a number of elements of the bioethanol production process that are inefficient and expensive. The pretreatment step is currently very expensive due to high energy and chemical inputs (206). Furthermore, the enzymes used for hydrolysis are expensive (207) and inefficient due to product inhibition (208, 209), inhibition by pretreatment

degradation products (210, 211) and the adsorption of cellulases onto lignin (212). There are also issues with the fermentation process, including lack of natural microorganisms with the ability to ferment mixtures of pentose and hexose sugars to ethanol (213), inhibition of microorganisms by degradation products (24) and, for SSF, the difference in optimal temperatures of fermenting microorganisms and cellulases (203). There are three main approaches to improving the efficiency and reducing the cost of bioethanol production from lignocellulosic biomass. One option is to reduce the cost and/or improve the efficiency of the pretreatment step. For example, costs can be reduced by increasing the ability to recycle chemicals, performing the pretreatment at a reduced temperature or producing a useful by-product (214). New pretreatments such as ionic liquids are also being investigated which may provide more efficient structure disruption (215, 216). A second option is to improve the efficiency of the enzymes so that they can more easily depolymerise the complex nature of the cell wall, for example by characterisation of enzymes from novel lignocellulose-digesting organisms, random mutagenesis of fungal strains or genetic engineering of individual enzymes (207). Similar approaches can be taken to improve the fermentation process. A third option is to lower the recalcitrant nature of the lignocellulosic material itself, through modification of the plants. This thesis focuses on the third option, and the potential targets for reducing the recalcitrant nature of lignocellulosic biomass are discussed below.

1.5 Targets for reducing the recalcitrant nature of lignocellulosic biomass

Saccharification is the process of breaking down complex carbohydrates into the monosaccharide components. This is necessary for the production of bioethanol so that the monosaccharides can be fermented into ethanol. In order to make second generation

bioethanol a commercially viable product, it is necessary to improve the efficiency of saccharification of lignocellulose and, as mentioned above, reducing the recalcitrant nature of lignocellulose is one way to do so. It has been demonstrated that alteration of a number of cell wall components can affect the recalcitrance of lignocellulosic biomass and thus improve the saccharification and/or digestibility of lignocellulosic biomass; these will be discussed below. While saccharification of lignocellulose refers to the ability of cellulases to release monosaccharides from the material, digestibility refers to the ability of livestock to utilise the nutrients present in lignocellulose. This is generally tested by incubating the biomass with ruminal fluids in conditions resembling those of the stomach and measuring the resulting weight loss (217).

1.5.1 Lignin as a target

Due to the relatively complete understanding researchers have of the enzymes involved in the lignin monomer synthesis pathway, the effect of lignin on lignocellulose saccharification is the best researched cell wall component. Changes in the amount, structure and composition of lignin have all been shown to alter saccharification and digestibility.

1.5.1.1 Lignin content as a target

Many studies have shown that altering the expression of enzymes in the lignin monomer synthesis pathway leads to alteration in the amount of lignin deposited in the cell wall. This, in turn, has been shown to effect saccharification. Chen and Dixon (218) showed that in six transgenic lines in alfalfa, each downregulated in a different lignin biosynthetic gene, there was a strong negative correlation between lignin content and sugar released by enzymatic hydrolysis, with saccharification efficiency being increased by 56 – 100% in the transgenic plants compared to wild-type. A similar comprehensive study in Arabidopsis showed that in 20 mutants of 10 different genes of the lignin monomer synthesis pathway, lignin content was the main factor affecting

saccharification yield out of those the authors tested for (219). Activity of the CCR enzyme is considered to be an important regulatory point for the overall carbon flux toward lignin (220) and studies have shown that reduction in lignin content by downregulation of CCR in poplar caused an increase in biomass digestibility and a doubling of sugar release compared to wild-type (221, 222). Furthermore, CAD-downregulated switchgrass and tobacco plants showed a strong negative relationship between lignin content and saccharification (223, 224), and downregulation of 4CL and C3H individually in poplar resulted in reduced lignin content and increased saccharification, with no effect on the other cell wall components measured (225, 226). A study has also been carried out in which an artificial enzyme, monolignol 4-O-methyltransferase, which etherifies the 4-hydroxyls of monolignols and so prevents free radical formation, was expressed in Arabidopsis. The resulting plants had reduced lignin content and an increase in saccharification, with no effect on lignin composition or cellulose content (227). Lignin content has been correlated negatively with saccharification in natural populations of alfalfa (228) and guineagrass (229), and between different plant species (230). Removal of lignin from the cell wall by chemical or biological pretreatment has also been shown to increase saccharification (225, 231).

1.5.1.2 Lignin composition and structure as a target

Studies have also shown an effect of lignin composition on saccharification of lignocellulosic biomass. It is often hard to distinguish between the effect of lignin content and composition because modifications of the monolignol synthesis enzymes often result in alterations in both properties. For example, downregulation of C4H, 4CL and CCoAOMT all result in increased saccharification but the plants have an increase in S:G ratio as well as a reduction in lignin content (219, 232, 233). However, Studer et al. (234) showed that in a large population of natural accessions of poplar, although there was a strong negative correlation between lignin content and saccharification, this was only true for samples with an S:G ratio less than 2.0. For higher S:G ratios,

saccharification increased with an increasing proportion of S units and lignin content had much less effect. Furthermore, when the biomass was subject to a pretreatment before hydrolysis, lignin content no longer showed any correlation with saccharification, whereas S:G ratio still did. Another example of lignin composition affecting saccharification is highlighted in Arabidopsis, where down-regulation of F5H, an enzyme involved in the synthesis of S monomers, resulted in a reduction in S:G ratio and a decrease in saccharification. Accordingly, overexpression of F5H led to an increase in S:G ratio and an increase in saccharification (235). A similar observation was reported in poplar overexpressing F5H (226).

The increase in saccharification observed in plants in which the lignin contains a higher proportion of S units is thought to be due to the number of methoxy groups on the lignin monomers (S>G>H). An additional methoxy group means one less reactive site, which results in fewer possible combinations during lignin polymerisation. This means that lignin rich in S units has a simpler structure that is potentially more easily depolymerised (236). Furthermore, lignin rich in S units have elevated concentrations of the more easily digestible β-*O*-4 linkages and lower concentrations of the more recalcitrant β-5 and 5-5 C-C linkages because the C-5 position is not available for coupling due to the extra methoxy group (156). Alternatively, G lignin units are more likely to form C-C bonds during polymerisation. This is because the lower number of methoxy groups on the G monomer ring causes the unpaired electron density to be greatest on the carbon nuclei rather than the phenolic oxygen, resulting in a preference for formation of C-C linkages (237).

It is also thought that the proportion of H monomers in lignin can affect cell wall saccharification. The monolignol biosynthetic genes C3H and HCT are involved in the synthesis of S and G monomers but not H monomers (Figure 6) (133). In natural lignin, H units are present at very low levels but knockdown of these genes results in an increase in the proportion of H units and an increase in saccharification (218, 238-240)

and extractability of the lignin (236). Furthermore, rescuing the lignin deficiency of the C3H mutant by disruption of a transcriptional mediator showed that the increased saccharification remained when lignin content of the mutant was returned to wild-type levels, suggesting that the change in lignin composition was responsible for the increase in saccharification (240). Ziebell et al (236) hypothesised that the increased extractability of lignin enriched in H units, despite the lower number of methoxy groups on H units in comparison to S and G units, is because the methoxy groups can help to stabilize the free radical and aid oxidation. Therefore the lower stability of the free radical on H units, due to a lack of methoxy substitution, likely results in a lower number of reactive species and so smaller chain lengths (241). In fact, this is possibly the reason for H units being present in much smaller amounts in natural lignin than S and G monomers.

The effect of lignin composition on saccharification is however not clear cut. A study into the digestibility of natural maize accessions with variation in lignin content, structure and composition showed that amount of lignin and proportion of H units both affected digestibility but that S:G ratio did not (242). Alternatively, a previously mentioned study, that measured saccharification in a set of alfalfa mutants downregulated separately for six monolignol synthesis enzymes, showed an effect of lignin content on saccharification but failed to find any correlation between lignin composition and saccharification (218).

1.5.1.3 Problems associated with altered lignin

Despite the positive effects of altering lignin content or composition on saccharification, these modifications often result in severe growth and/or development defects (243). For example, C3H and HCT downregulation resulted in reduced biomass in alfalfa, tobacco and Arabidopsis (218, 238), C3H T-DNA lines in Arabidopsis displayed severe reductions in growth and fertility (244), and C3H suppression in poplar has been

reported to cause reduced biomass and collapsed xylem (245). Studies have also reported dwarfism, male sterility and reduced growth in C4H Arabidopsis mutants (246, 247), while CCR downregulation in poplar, tobacco and *Leucaena leucocephala*, a leguminous pulpwood tree, caused stunted growth (222, 248-250). Quadruple PAL knockout mutants in Arabidopsis were stunted and sterile (251), and vessel wall collapse and retarded growth has been reported in tobacco and poplar downregulated in 4CL (252, 253). These reductions in yield and fertility are obviously undesirable traits for biofuel crops. More research is therefore needed into how increased saccharification can be achieved without a yield or fitness penalty.

1.5.2 Cellulose as a target

Alterations in cell wall components other than lignin have also been shown to have an effect on biomass saccharification. Indeed, in the study described above by Studer et al. (234), which examined the effect of lignin content and S:G ratio on saccharification in natural poplar accessions, it was found that some samples with lignin content and S:G ratio similar to wild-type displayed much higher sugar release than expected. This suggests that factors other than lignin can influence lignocellulosic recalcitrance. Furthermore, a recent study showed that QTLs for lignin abundance were independent of those for saccharification in a maize recombinant inbred population (254). Altering cellulose particularly has been revealed to affect saccharification of plant biomass as it typically comprises around one third of the total mass of plants and the crystalline cellulose is naturally inaccessible to hydrolytic enzymes (27). In corn stover a negative correlation was observed between cellulose crystallinity and initial rate of hydrolysis (202), and it was shown that this was due to enzyme accessibility to the cellulose (255). The same was observed for Avicel, a purified crystalline cellulose (256). Increased saccharification has been observed in Arabidopsis and rice plants with a reduction in cellulose crystallinity caused by mutations in the CESA enzymes (257-259). Further evidence of cellulose crystallinity affecting saccharification comes from heterologous

expression of microbial carbohydrate binding modules (CBM), which disrupt cellulose crystallinity (260, 261) and have been shown to increase digestibility of rice biomass (262). In addition, expansins, which are the plant equivalent of CBMs and act to loosen the plant wall during growth and fruit softening, have been shown to increase saccharification of poplar biomass when used in conjunction with cellulases during enzymatic digestion (263). Similar to plants with reductions in lignin content, a reduction in cellulose content or crystallinity can have an effect on plant growth and development, including dwarfing, reductions in stem strength, reduced fertility and even embryo lethality (34, 35, 257, 258, 264-272).

1.5.3 Hemicellulose as a target

Relatively few studies have investigated the effect of hemicellulose content on lignocellulose saccharification but a link between the two has been demonstrated. For example, a number of Arabidopsis and poplar *irregular xylem* (*irx*) mutants, with T-DNA insertions or RNA silencing of glycosyltransferases involved in xylan synthesis, were shown to have reduced GX content in secondary cell walls and a significant increase in saccharification (273-275). Furthermore, poplar cell walls with a reduction in hemicellulose, due to overexpression of xylanase or xyloglucanase, showed an increase in enzymatic saccharification of almost double that of wild-type (276). Correlations have also been demonstrated between hemicellulose removal by chemical extraction and both cellulose accessibility (277) and enzymatic hydrolysis (278-280) in plant species with relatively low lignin contents. For lodgepole pine, which has high lignin content, a study has shown that enzymatic hydrolysis was only correlated to hemicellulose removal once approximately 20% of the lignin was removed (281). A reduction in hemicellulose content likely results in increased saccharification because of the role hemicelluloses have in coating and cross-linking the cellulose microfibrils, as discussed in section 1.3.1. This interaction reduces the accessibility of the cellulose to cellulases. However, plants with a reduction in hemicellulose content have also been

shown to have growth and development problems. They often have a large reduction in biomass, weak stems and a collapsed xylem phenotype (68-70, 273-275, 282).

Interestingly, increases in hemicellulose content have also been linked to increased saccharification. For example, comparing two Miscanthus accessions with different hemicellulose content but no difference in either lignin or cellulose content, revealed that the accession with high hemicellulose content had much higher saccharification (283). Furthermore, a positive correlation has been observed between hemicellulose content and saccharification in Arabidopsis (219). This positive relationship between hemicellulose content and saccharification is possibly due to the correlation of other cell wall components with hemicellulose content. For example, in the Miscanthus example, the sample with high hemicellulose content also showed a reduction in cellulose crystallinity (283), while in the Arabidopsis example hemicellulose content correlated negatively with lignin content (219).

1.5.3.1 Hemicellulose substitution as a target

The substitutions of hemicellulose can also impact on saccharification of lignocellulosic biomass. Indeed, treatment of corn stover with dilute acid followed by enzymatic hydrolysis left about 50% of the hemicellulose as oligosaccharides that were resistant to hydrolysis. These oligosaccharides were shown to be highly substituted with acetic acid, ferulic acid, arabinose and galactose (284). The effect of the most common substitutions of xylans will be discussed below.

1.5.3.1.1 Acetylation as a target

The xylosyl units of GX and GAX in dicot and grass cell walls are commonly substituted with O-linked acetate (77). This acetylation has been shown to inhibit enzymatic hydrolysis. For example, de-acetylation of both aspen and wheat using chemical treatment was shown to increase glucose release using cellulases by two- to

four-fold compared to non-treated material (285-287). This is likely because the acetyl groups prevent hydrolytic enzymes from accessing the glycosidic linkages through steric hindrance (288). Acetylation has also been shown to reduce the effect of acid pretreatment as deacetylation of corn stover biomass greatly increased the proportion of xylose monomers released by acid pretreatment compared to non-deacetylated controls (289, 290). Arabidopsis plants with a T-DNA insertion in an *O*-acetyl transferase gene, *ESK1*, showed an approximately 50% reduction in acetyl substitutions and an increase in the conversion of cellulose to glucose during enzymatic hydrolysis (83). However mutation of other *O*-acetyl transferases, RWAs, have been shown to have no impact on saccharification despite a reduction in acetic acid substitution of approximately one third compared to wild-type (80). Furthermore, recombinant expression of an acetylxylan esterase in Arabidopsis reduced acetic acid substitution by 20% but did not affect saccharification (291). In addition to the effect of hemicellulose acetylation on the efficiency of pretreatment and enzymatic hydrolysis, release of acetic acid into the fermentation medium during pretreatment can inhibit yeast growth and fermentation rates (292-294).

1.5.3.1.2 Glucuronic acid substitution as a target

The xylan backbones of GX and GAX are also commonly substituted with GlcA, which may or may not be methylated ([Me]GlcA). T-DNA insertions in the glycosyltransferases which add these substitutions to hemicellulose caused an almost complete lack of [Me]GlcA substitution of xylan in Arabidopsis. The altered xylan was easier to extract from the cell wall using alkali extraction and required fewer enzymes for its hydrolysis (94). Small increases in saccharification have been observed for double and triple mutants of these glycosyltransferases (95). In addition to the simplified xylan structure, there are a number of possible reasons for the positive effect of reduced [Me]GlcA substitution on saccharification. First of all, the [Me]GlcA may play a role in the association of hemicellulose with cellulose in the cell wall; it has been

suggested that the MeGlcA substitutions on xylans cause a change in conformation of the xylan to form a flat, ribbon-like structure the same shape as cellulose, allowing the xylan to interact with the microfibrils (184, 185). Furthermore, it has been shown that an ester linkage can form between [Me]GlcA and lignin (295), suggesting that [Me]GlcA also plays a role in lignin to hemicellulose interaction. Finally, acid pretreatment is unable to break the linkage between the [Me]GlcA and xylose, and so [Me]GlcA reduces the solubilisation of hemicellulose during pretreatment (296).

1.5.3.1.3 Ferulic acid substitution as a target

The content of ferulic acid in the cell wall has also been shown to affect saccharification and digestibility, due to the formation of ferulic acid cross-linkages between hemicellulose and lignin molecules, as described in section 1.2.6. Ferulic acid content has been shown to be negatively correlated with digestibility in forage grasses (297, 298), and switchgrass bred by divergent selection for high and low di-ferulate content showed that plants with a 16% higher di-ferulate content had a 12% reduction in digestibility (299). Furthermore, incorporation of differing concentrations of ferulates within a synthetically lignified maize cell wall showed that a 70% decrease in ferulate-lignin cross-links led to a 24 − 46% increase in sugar release from the biomass by enzymatic digestion (300). Gene manipulation has also demonstrated a role of ferulic acid in cell wall digestibility and saccharification. When the forage grass tall fescue was transformed to express a fungal ferulic acid esterase to reduce the number of ferulic acid residues in the cell wall, transformed plants had reduced levels of cell wall ester-bound ferulic acid and increased digestibility (301). Overexpression of a BAHD CoA-acyltransferase in rice, which is thought to be involved in the addition of ferulic acid and *p*-coumaric acid to GAX, resulted in a 60% reduction in cell wall ferulic acid content and a 300% increase in *p*-coumaric acid content. These plant lines showed a 20 − 40% increase in saccharification compared to wild-type, which the authors attributed to the reduction in ferulic acid rather than the increase in *p*-coumaric acid (109). Finally,

a knock-out mutant of a GT61 family gene, implicated in the addition of xylosyl residues on to the arabinosyl side chain of GAX, resulted in a large reduction in ferulic acid and an increase in saccharification (117). The authors believed that the increase in saccharification was principally as a result of the reduction in ferulic acid.

1.5.4 Pectins as a target

Finally, although pectin only makes up a small proportion of secondary cell walls, it can still act as a barrier to cellulolytic enzymes (118). Lionetti et al. (302) showed that a reduction in pectin content and an increase in the amount of methyl esterification of the pectin increased saccharification by 30 - 40% in wheat and tobacco. Methyl-esterification of HG has also been shown to have a positive impact on saccharification in Arabidopsis (303). This is likely because HG has been shown to cross-link via the backbone and methylation prevents the formation of these cross-linkages, resulting in a less complex structure (304, 305).

1.7 *Brachypodium distachyon*

Much of the research that has been done into the saccharification of lignocellulosic biomass has been carried out using Arabidopsis. The secondary cell wall composition and development of this annual dicot is quite different to the perennial monocots which are the main candidates for bioethanol crops, as discussed in section 1.3.1.1. The main differences are in the hemicellulose; the major hemicellulose in dicot secondary walls is GX, whereas in monocots it is GAX (16). GAX contains feruloyl side chains which can form cross-links with ferulates present in lignin and other GAX molecules (306). Dicots also contain large amounts of pectin in the cell wall which grasses do not. For this reason Arabidopsis is not a particularly suitable model for developing our knowledge of cell walls in bioethanol crops. A monocot that is currently being developed as a more

appropriate model for grass cell wall studies is *Brachypodium distachyon* (hereafter, Brachypodium), a member of the Pooideae subfamily of the grass family Poaceae. Brachypodium is a small, annual, temperate grass which has a small, diploid genome of 272 Mbp (compared to 135 Mbp for Arabidopsis) (307). It also has a relatively rapid generation time (8 - 12 weeks), is easy to grow, can be grown to relatively high density and is self-pollinating (308). The genome of Brachypodium has recently been fully sequenced (307) and, in addition, there are many other genetic tools being developed for this model plant, including bacterial artificial chromosome (BAC) libraries (309), physical maps (310), germplasm (311) and mutant collections (http://brachypodium.pw.usda.gov, http://www.brachytag.org), EST collections, genetic linkage maps (312), microarrays, and efficient *Agrobacterium tumefaciens* based transformation methods (313). Grass genomes have been shown to have significant co-linearity in gene organisation and Brachypodium is quite closely related to the major cereal crops (308).

1.8 Aims of the project

As I have demonstrated, there are a number of studies that have shown increases in biomass saccharification and digestibility upon alteration of various cell wall components. However, almost all of these studies have taken a reverse genetics approach, disrupting genes that play key roles in cell wall biosynthesis, or expressing genes that encode wall-modifying enzymes. The complexity of plant cell wall biosynthesis and the large number of genes involved, many of which are unknown, make it important that we explore saccharification potential in an empirical manner to gain a better understanding of factors that can impact this characteristic. Indeed, it has been estimated that $10 - 15\%$ of Arabidopsis genes (approximately $2500 - 4000$) are related to cell wall biology (314), while it has been estimated that only 121 genes have

been experimentally-validated as cell wall-related (315). Furthermore, studies on the digestibility of maize stover for animal feed show that an extremely large number of quantitative trait loci (QTLs) impact on this trait (316). In this context, classical (or forward) genetic screens benefit from a more empirical approach to gene discovery and have great potential to identify new or unexpected factors that impact on a phenotype.

This PhD project involves a forward genetic approach to develop further understanding into the genes and components involved in the recalcitrant nature and ease of saccharification of lignocellulosic biomass. This will be achieved by screening a large population of randomly mutagenised Brachypodium plants for alterations in saccharification of stem material compared to wild-type. Such a screen has the potential of identifying a range of genetic lesions that impact on lignocellulose saccharification and is not confined to predictions made from our current understanding of genes involved in cell wall synthesis. Furthermore, the direct screening for saccharification, rather than the disruption of particular cell wall components or using a forward screen to select for plants with visible cell wall phenotypes, will hopefully allow the identification of plants with altered saccharification potential but with normal growth and development that would not be affected in field performance.

Chapter 2 - Materials and Methods

2.1 Plant growth and material

The mutant *Brachypodium distachyon* population was provided by Dr. Richard Sibout and Prof. Herman Hofte at the Institut Jean-Pierre Bourgin, INRA-AgroParisTech, Versailles. The population consisted of 2500 M_2 plants in the Bd21 background which had been mutagenised with 3 mM or 10 mM sodium azide (NaN_3), according to the protocol described in (317). This population was made up from the progeny of 300 M_1 plants, with 12 seeds sown from each M_1 plant. The genetically effective cell number (GECN) of Brachypodium has been shown to be four (317). Therefore, a single M_1 plant should produce seeds with four different sets of mutations, as each cell in the mutagenised seed will be mutated randomly. All plants used in experimentation were sown in low nitrate soil (John Innes Seed Compost) and seeds were vernalised in the dark at 4°C for 21 days before being transferred to the greenhouse at 20°C with a 16 h light photoperiod. Watering was stopped when plants had reached full maturity and begun to senesce, and plants and seeds were harvested separately once they had completely dried. Internodes from the primary and secondary stems of mature and dried plants were used for saccharification and cell wall composition analyses. All internodes excluding the first were cut into 4 mm lengths and put into one 2 ml tube per plant with three ball bearings. Stem sections were ground for 60 sec using a custom-made robotic platform (Labman Automation, Stokesley, North Yorkshire, UK) (318).

2.2 Saccharification analysis

Ground stem material from above was weighed into four replicate 4 mg samples per plant in a 96 deep-well plate using the robotic platform above (318). Saccharification analysis of the plant material was performed using a liquid handling robotic platform

(Tecan Evo 200; Tecan Group Ltd. Männedorf, Switzerland), as described in Gomez et al. (318). In brief, plant material was pretreated with 0.5 M NaOH at 90°C for 20 min, followed by adjustment of the pH to equal that of the enzyme solution. Enzymatic digestion was then performed with a 4:1 mixture of Celluclast (Cellulase from *Trichoderma reesei*) and Novozyme 188 (cellobiose from *Aspergillus niger*) (Novozymes) and incubated at 50°C for 8 h with shaking. The Celluclast enzyme is produced by submerged fermentation of the fungus *Trichoderma reesei*. The key enzyme activity is cellulase, that hydrolyzes (1,4)-β-glucosidic linkages in cellulose and other β-D-glucans into glucose, cellobiose and higher glucose polymers, and it has an activity of ≥700 Endoglucanase units (EGU)/g (μmol reducing sugars released per gram of enzyme per min). The Novozyme 188 enzyme is a cellobioase enzyme preparation obtained by submerged fermentation of an *Aspergillus niger* microorganism. The key enzyme activity is cellobiase, that hydrolyzes cellobiose to glucose, and it has an activity of ≥250 Cellobiase units (CBU)/g (μmol of glucose released per gram of enzyme per min). The mixture of enzymes was diluted such that 7 FPU enzyme was added per gram biomass. Finally, saccharification potential was determined by measuring the amount of reducing sugars via a colourimetric assay, using 3-methyl-2-benzothiazolinone hydrazone (MBTH). Each plate contained standards of 50, 100 and 150 nmol glucose (three replicates each) and filter paper disks (four replicates) in order to monitor any change in enzyme concentration or conditions through time. Change in colour was read with a Tecan Sunrise microplate absorbance reader at 620 nm.

2.2.1 Selection of saccharification mutants

Mutant plants were selected from the first round screen if the saccharification value was 20% higher than the wild-type plant on the same 96-well plate and significant at $p \leq 0.01$ using a student's t-test, and if the saccharification value was outside the 95% confidence intervals of the wild-type population. To test for heritability of the high saccharification trait, twelve seeds were sown from each putative mutant selected in the

first screen and the resulting plant material tested for saccharification, as described above. The saccharification trait was considered to be heritable if all of the offspring showed a similar saccharification trait to that observed in the first screen.

2.3 Cell wall composition analyses

AIR was used for all cell wall composition analyses, unless otherwise stated. All cell wall composition analyses were carried out in biological triplicates.

2.3.1 Alcohol insoluble residue (AIR)

To obtain AIR, 100 mg stem material was ground to a fine powder using a tissue lyser with grinding jar set attachments at 20 1/s frequency for two min. The ground biomass was then subject to 30 min incubation with shaking at room temperature (RT), followed by centrifugation at 10,000 g at 4°C for 10 min and removal of the supernatant, with the following solutions: once with absolute ethanol, twice with chloroform:methanol (1:1 (v/v)), twice with 80% methanol and once with 100% methanol. The samples were then left to dry overnight at RT. This treatment provides a crude cell wall extract and solubilises and removes the starch in the biomass.

2.3.2 Xylan-enriched cell wall fraction

A xylan-enriched cell wall fraction was used for non-cellulosic monosaccharide analysis in addition to AIR. This was prepared by incubating 5 mg AIR in 50 mM 1,2-Diaminocyclohexanetetraacetic acid (CDTA) for 24 h at RT with rocking. The sample was then centrifuged at 17,000 g at 4°C for 20 min and the pellet washed twice with deionised water. The pellet was then incubated with 50 mM sodium carbonate (Na_2CO_3) containing 10 mM sodium borohydride ($NaBH_4$) for 24 h at 4°C with rocking, the sample centrifuged at 17,000 g at 4°C for 20 min and the pellet washed twice with

deionised water. The CDTA and Na_2CO_3 treatments remove the pectin from the biomass. The pellet was then incubated with 4 M KOH containing 10 mM $NaBH_4$ for 24 h at 4°C with rocking, the sample centrifuged at 17,000 g at 4°C for 20 min. This extracts mainly xylans. The supernatant was then adjusted to pH 5 with glacial acetic acid, dialysed against deionised water for 24 h to remove salts and lyophilised. The resulting pellet was re-suspended in 50 mM ammonium acetate buffer (pH 6), evaporated using a vacuum concentrator (Savant SPD131DDA, Thermo Scientific) and the pellet used for TFA hydrolysis, as described below (section 2.3.5).

2.3.3 Lignin content

Lignin content was quantified using Foster et al.'s acetyl bromide method (319), based on the method reported by Fukushima and Hatfield (320). Briefly, 3 mg AIR was weighed into a 5 ml volumetric flask and 250 µl freshly prepared acetyl bromide solution (25% (v/v) acetyl bromide in glacial acetic acid) added. Samples were incubated at 50°C for 2 h, followed by a further 1 h with vortexing every 15 min to solubilise the lignin. Samples were then cooled to RT before addition of 1 ml 2 M NaOH to hydrolyse excess acetyl bromide and 175 µl freshly prepared 0.5 M hydroxylamine hydrochloride to destroy bromine and polybromide. Samples were then taken to 5 ml with glacial acetic acid, mixed, and the absorption read using a Shimadzu UV-1800 spectrophotometer at 280 nm. Lignin content (µg mg^{-1} cell wall) was then determined using the following formula:

(absorbance ÷ (coefficient x path length)) x ((total volume x 100%) ÷ biomass weight))

The coefficient is specific to the type of plant that is being analysed and, for grasses, a coefficient of 17.75 is used (320).

2.3.4 Crystalline cellulose content

Crystalline cellulose content was analysed using Foster et al.'s method (321), based on the method reported by Updegraff (322). Briefly, 1 ml Updegraff reagent (acetic acid:nitric acid: water 8:1:2 (v/v/v)) was added to 4 mg AIR and heated at 100°C for 30 min. Samples were then cooled to RT and centrifuged at 10,000 rpm for 15 min. The pellet contains just crystalline cellulose. The pellet was then washed four times with 1.5 ml water, air dried and incubated with 175 μl 72% H_2SO_4 for 45 min at RT to hydrolyse the cellulose to glucose. Samples were then centrifuged at 10,000 rpm for 15 min after re-suspension of the pellet in 825 μl of water. Finally, the glucose content of the supernatant was quantified using the colourimetric anthrone assay: 10 μl of each sample was added in triplicate to a 96 well polystyrene microtiter plate with 90 μl water and 200 μl anthrone reagent (2 mg anthrone ml^{-1} concentrated H_2SO_4). A standard curve for glucose (0, 2, 4, 6, 8 and 10 μg) was made on each plate for quantification. The plate was then heated at 80°C for 30 min, allowed to cool and the absorption read at 620 nm using a Tecan Sunrise microplate absorbance reader.

2.3.5 Non-cellulosic monosaccharide analysis

Non-cellulosic monosaccharide analysis was performed using high performance anion exchange chromatography (HPAEC) (Carbopac PA-10; Dionex, Camberley, Surrey, UK). Analysis was performed on 3 mg AIR or a xylan-enriched cell wall fraction prepared from 5 mg AIR, as described in section 2.3.2. Non-cellulosic polysaccharides were hydrolysed to their monosaccharide constituents using 0.5 ml 2 M trifluoroacetic acid (TFA) for 4 h at 100°C. Samples were then cooled to RT and evaporated completely using a vacuum concentrator (Savant SPD131DDA, Thermo Scientific) with refrigerated vapour trap (Savant RVT4104, Thermo Scientific), rinsed twice with 200 μl isopropanol and once with 500 μl water using the vacuum concentrator. Samples were then re-suspended in 150 μl deionised water, centrifuged at 10,000 rpm for five min and the supernatants filtered with 0.45 μm PTFE filters (Millex™, Millipore).

Monosaccharides were separated and quantified by HPAEC using a Dionex ICS-3000 with integrated amperometry detection. Chromatographic separation was performed on a CarboPac PA20 (3 x 150 mm) column (Thermo) using a gradient elution. The mobile phase consisted of solution A: 100% water, solution B: 200 mM sodium hydroxide, and solution C: 0.1 M sodium hydroxide, 0.5 M sodium acetate. A flow rate of 0.5 ml min^{-1} was used and the gradient was as follows: 0 min: 100% A; 5 min: 99% A, 1% B; 15 min: 99% A, 1% B; 22 min: 47.5% A, 22.5% B, 30% C; 30 min: 47.5% A, 22.5% B, 30% C. The column was then washed as follows: 30.1 min: 100% B; 37 min: 100% B; 37.1 min: 99% A, 1% B; 50 min: 100% A; 55 min: 100% A. The separated monosaccharides were quantified by using external calibration with a mixture of seven monosaccharide standards at 100 μM (arabinose, fucose, galactose, glucose, mannose, rhamnose, and xylose) that were subjected to acid hydrolysis in parallel with the samples.

2.3.6 Ferulic acid content

Ferulic acid in the cell was quantified using a protocol based on Fry's method (323). To release the polysaccharide-bound ferulic acid, 1 ml 1 M NaOH was added to 10 mg AIR and incubated under argon at 25°C in the dark for 24 h. The resulting sodium ferulate was separated into the ferulate and sodium ion by addition of 2 M TFA to bring the pH down to 1.0. Partitioning of the ferulate in to the organic phase was then achieved by addition of 1ml butan-1-ol, vigorous shaking and removal of the upper organic phase for analysis. This partitioning was repeated twice. Finally, the organic phases were combined, the butan-1-ol evaporated and the residue redissolved in 200 μl 50% methanol (v/v). The extracted ferulic acid was analysed using high performance liquid chromatography (HPLC) on an activated reverse-phase C18 5 μm (4.6 x 250 mm) XBridge column (Waters Inc.) in 100% methanol-5% acetic acid, with a 20 - 70% methanol gradient over 25 min at a flow rate of 2 ml min^{-1}. Ferulic acid was detected and quantified with a SpectraSYSTEM® UV6000LP photo-diode array detector

(Thermo Scientific), with UV–visible spectra collected at 240 – 400 nm, and analysed against a ferulic acid standard. Samples were run on the HPLC by Valeria Gazda, CNAP, University of Yok.

2.3.7 Lignin composition

Analysis of lignin composition was carried out by Catherine Lapierre at INRA-AgroParisTech, Versailles, using thioacidolysis followed by gas chromatography-mass spectrometry (GC-MS), as described in Lapierre et al. (324). Briefly, 10 mg AIR was added to 5 ml freshly prepared thioacidolysis reagent (0.2 M boron trifluoride etherate in 8.75:1 (v/v) dioxane:ethanethiol) in a glass tube with a Teflon-lined screwcap. Thioacidolysis was achieved by incubating at 100°C in an oil bath for 4 h with occasional shaking. The reaction was then cooled to RT, poured into 30 ml dichloromethane containing 0.2565 mg C21 internal standard and 0.3109 mg C19 internal standard, the tube rinsed with 3 x 10 ml H_2O and added to the reaction mixture. The pH of the aqueous phase was adjusted to 3 - 4 with 0.4 M sodium carbonate, extracted with dichloromethane (3 x 30 ml) and the combined organic extracts dried over sodium sulphate, then evaporated under reduced pressure at 40°C. The resulting residue was redissolved in 0.5 ml dichloromethane and 5 μl of this solution silylated with 100 μl bistrimethylsilyltrifluoroacetamide (BSTFA) and 10 μl GC grade pyridine in a 200 μl silyvial by incubating at RT for 5 min. The resulting thioacidolysis products: H, G and S monomers, were quantified using GC-MS on a polydimethylsiloxane capillary column (SPB1, Supelco, 30 m x 0.32 mm I.D., film thickness 0.25 μm) working in the temperature program mode (from 90 to 260°C at +5°C or from 160 to 260°C at +2°C) with helium as carrier gas (0.5 bar inlet pressure), and a flame ionization detector. This was coupled to a quadrupole spectrometer operating in the electron impact mode (70eV), with a source at 100°C, an interface at 270°C and a 60 to 650 m/z scanning range. The H, G and S trimethylsilylated derivatives have m/z values of 239, 269 and 299, respectively

2.4 Histology

Transverse sections were cut by hand from the second internode of dry, mature stems of wild-type and *sac* mutant plants, using a double-edged razor blade. Sections were then cleared by autoclaving in 85% (v/v) DL-lactic acid. Wiesner staining was achieved by incubating the sections for 10 min in phloroglucinol-HCl (1:1 (v/v) ratio of 2% (w/v) phloroglucinol in ethanol: concentrated HCl). Sections were viewed immediately after staining using a Zeiss Axiovert 200 microscope and recorded with a Zeiss Axiocam Hrm digital camera and Zeiss Axiovision software.

2.5 Stem mechanical properties

The strength and stiffness of mutant and wild-type stems was assessed using three-point bending tests. Stem lengths of 5 cm were cut from the middle of the third internode of dried plants and bending tests were performed on this material using a Universal Testing Machine (model 3367, Instron, High Wycombe, UK). Stem segments were placed on two supports, separated by 2 cm, and a 2 mm pushing probe manually lowered until just in contact with the stem. The pushing probe was then set to automatically lower at a rate of 10 mm min^{-1}. Stem strength was measured by the maximum bending stress, λ_{max} ($\lambda_{max} = F_{max}Lr/4I$, where Fmax is the maximum force the sample can withstand before failure, L is the distance between supports, r is the radius of the stem, and I is the second moment of the cross-sectional area of the stem ($\pi r^4/4$)). Stem diameter was measured using a digital caliper. Stem stiffness was measured by the bending modulus (MPa) which is calculated by R/I, where R is the resistance of the stem to curvature ($R = L^3(dF/dY)/48$, where dF/dY is the initial slope of the force displacement curve acquired from the bending test).

2.6 Molecular biology techniques

2.6.1 DNA extraction from plant material

Unless specified otherwise, DNA was obtained by, firstly, grinding ≤100 mg leaf material to a fine powder in liquid nitrogen (LN) using a pestle and mortar, followed by DNA extraction using the DNeasy Plant Mini kit (Qiagen) according to the manufacturer's instructions.

2.6.2 Quantification of DNA concentration

Unless otherwise specified, DNA concentration was measured by the 260:280 ratio using a NanoDrop 8000 spectrophotometer.

2.6.3 Polymerase Chain Reaction (PCR)

Two different DNA polymerases were used for PCR reactions: DreamTaq and Phusion Hot Start II High Fidelity (Thermo Scientific). The reaction mix for a 20 µl reaction and cycling conditions for these polymerases are shown in Table 1 and Table 2, respectively. Table 3 shows all primers used in experimentation and their Tms.

Table 1: PCR reaction mixes for DreamTaq and Phusion Hot Start II High Fidelity DNA polymerases

	DreamTaq	Phusion Hot Start II
H_2O	11.2 µl	9.2 µl
10x DreamTaq Green Buffer	2 µl	-
5x Phusion HF Buffer	-	4 µl
10 mM dNTPs	0.4 µl each	0.4 µl each
10 µM forward primer	1 µl	1 µl
10 µM reverse primer	1 µl	1 µl
Template DNA	3 µl	3 µl
DNA polymerase	0.2 µl	0.2 µl

Table 2: PCR cycling conditions for DreamTaq and Phusion Hot Start II High Fidelity DNA polymerases

	DreamTaq		Phusion Hot Start II	
	Temperature	Time	Temperature	Time
Initial denaturation	95°C	3 min	98°C	30 s
Denaturation	95°C	30 s	98°C	10 s
Annealing	Tm	30 s	Tm + 3°C (or Tm if ≤ 20 nt)	20 s
Extension	72°C	1 min kb^{-1}	72°C	15-30 s kb^{-1}
Final Extension	72°C	5 min	72°C	10 min

Table 3: List of primers used in experiments, their sequences and their Tms.

Primer	F/R	Sequence	Tm (°C)
M13	F	TGTAAAACGACGGCCAGT	53.9
	R	AGGAAACAGCTATGACCAT	53.0
pGEX	F	GGGCTGGCAAGCCACGTTTGGTG	69.9
	R	CCGGGAGCTGCATGTGTCAGAG	67.9
Bd5g14720	F	GCAACCAGCCTAGCTATCGC	62.0
	R	CTGCGTCACAACAAGAGAGG	60.5
Bd5g14720_inner	F	TCGGCTTCACCTCTAACC	60.37

2.6.4 Agarose gel electrophoresis

Agarose gel electrophoresis was used to separate DNA molecules of different sizes. Gels were prepared by dissolving 1% (w/v) agarose in 0.5x TBE buffer (45 mM Tris-borate, 1 mM EDTA, pH 8.0) by heating in a microwave at full power for 2 min. The solution was then cooled to RT, 0.00002% (v/v) ethidium bromide added, poured into a cast, a well comb added and the gel left to set for 30 min. Once set, the gel was placed in a tank filled with 0.5x TBE buffer and the comb removed. Samples were mixed with

the appropriate volume of 5x loading buffer (50 mM Tris, 5 mM EDTA, 50% (v/v) glycerol, 0.3% (w/v) orange G, pH 8.0), loaded into the wells and the gel ran at 120 V for 45 min. 4 µl of 1 kb DNA ladder (Bioline) was ran alongside samples. DNA molecules were visualised using UV light.

2.6.5 Extraction and purification of DNA from agarose gels

Single bands from an agarose gel after electrophoresis were excised using a scalpel under UV light. The DNA was then extracted and purified using a QIAquick® Gel Extraction Kit (Qiagen) according to manufacturer's instructions. Briefly, three volumes buffer QG were added to one gel volume (100 mg = 100 µl) and the sample incubated at 50°C for 10 min to dissolve the gel. One gel volume isopropanol was then added, the sample mixed, applied to a QIAquick spin column in a 2 ml collection tube and centrifuged for 1 min at 13,000 rpm to bind the DNA. The flow-through was then discarded and, to wash the column, 750 µl buffer PE was added and centrifuged for 1 min at 13,000 rpm. Finally, to elute the DNA the spin column was placed in a microcentrifuge tube, 50 µl buffer EB (10 mM Tris-HCl, pH 8.5) added to the membrane of the column, incubated for 1 min and then centrifuged for 1 min at 13,000 rpm.

2.6.6 Purification of DNA from PCR reactions

DNA was purified from PCR reactions using a QIAquick® PCR purification kit (Qiagen) according to manufacturer's instructions. Briefly, five volumes buffer PB was added to one volume of the PCR reaction, the sample mixed, applied to a QIAquick spin column in a 2 ml collection tube and centrifuged for 1 min at 13,000 rpm to bind the DNA. The flow-through was then discarded and, to wash the column, 750 µl buffer PE was added and centrifuged for 1 min at 13,000 rpm. Finally, to elute the DNA the spin column was placed in a microcentrifuge tube, 50 µl buffer EB (10 mM Tris-HCl, pH

8.5) added to the membrane of the column, incubated for 1 min and then centrifuged for 1 min at 13,000 rpm.

2.6.7 DNA restriction digests

A typical restriction digest was carried out with 1 ug DNA, 5 U restriction enzyme, 1x buffer (recommended by the manufacturer), 1x BSA, and made up to 20 µl with H_2O. The reaction was incubated for 2 h at a temperature recommended by the manufacturer. Products were run on a 1% agarose gel to confirm digestion had occurred.

2.6.8 DNA ligation reactions

Ligation reactions were carried out in a total of 10 µl with 5 µl 2x rapid ligation buffer (promega), 1 µl T4 ligase and the remaining 4 µl made up with a 1:2 molar ratio of vector:insert DNA. Reactions were incubated at RT for 30 min.

2.6.9 Media

Lysogeny Broth (LB) was prepared by dissolving 25 g LB Broth Miller (Fisher Scientific) in 1 L deionised water, adjusting the pH to 7.0 and autoclaving at 121°C for 15 min. LB-agar was prepared as above but with the addition of 1.5% (w/v) agar before autoclaving.

2.6.10 Transformation of *E. coli*

Host *E. coli* cells were thawed on ice, 1 µl DNA (unless specified otherwise) added to 50 µl cells and gently mixed. Samples were incubated on ice for 30 min, heat-shocked at 42°C for 45 sec and incubated further on ice for 2 min. 450 µl SOC media (Bioline), pre-warmed to 37°C, was then added to the cells and the sample incubated at 37°C with shaking at 250 rpm for 1 h for the cells to recover. Spread plates were then produced with 50 µl and 150 µl cells on LB-agar plates supplemented with carbenicillin (50 mg L^{-1}) and incubated at 37°C overnight.

2.6.11 Plasmid DNA extraction

Single colonies were used to inoculate 3 ml LB supplemented with carbenicillin (50 mg L^{-1}) in 15 ml falcon tubes and incubated at 37°C overnight with shaking at 250 rpm. Plasmid DNA was then extracted from these cultures using Wizard® *Plus* SV miniprep kit (Promega), according to the manufacturer's instructions.

2.6.12 DNA cloning and Sanger sequencing

Cloning of PCR products was achieved using the StrataClone Blunt PCR Cloning Kit (Agilent Technologies) according to manufacturer's instructions with a few alterations. Briefly, 2 µl of a 1:2 dilution of a PCR reaction was mixed with 3 µl StrataClone Blunt Cloning Buffer and 1 µl StrataClone Blunt Vector Mix amp/kan. This mixture was then incubated at RT for 15 min to allow ligation to occur. The reaction was then placed on ice and transformation of StrataClone SoloPack competent cells carried out as above, using 1 µl of the ligation reaction and with 40 µl 2% X-Gal (5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside) spread on LB-agar plates for blue-white colour screening. Single white colonies were selected and a colony PCR carried out using M13 forward and reverse primers and DreamTaq DNA polymerase to check for insertion of the PCR product in to the plasmid. For positive colonies, plasmid DNA was purified and the DNA sequenced using the M13 forward and reverse primers (unless otherwise specified) by DNA Sequencing and Services (Dundee University). Sequence data was analysed using Vector NTI® software (Life technologies).

2.7 Phylogenetic analysis

Multiple alignment was performed on protein sequences using ClustalX2 (325). This alignment was used to construct a neighbour joining phylogenetic tree using ClusatlX2.

Bootstrap values were calculated for 1000 pseudosamples and values shown as percentages at nodes in the consensus tree which was visualised in TreeView (326).

2.8 Recombinant expression and purification of Brachypodium HCT proteins

2.8.1 Protein analysis techniques

2.8.1.1 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE gels were carried out using Mini-Protean Tetra cell apparatus (Bio-Rad, USA). The separating gel was composed of 2.5 ml 1.5 M Tris-HCl, pH 8.8; 100 µl 10% SDS; 3.3 ml 30% acrylamide; 4.05 ml H_2O; 100 µl 10% ammonium persulphate and 5 µl Tetramethylethylenediamine (TEMED). The solution was immediately transferred into a pre-assembled gel apparatus, isopropanol pipetted on top, and allowed to solidify. The isopropanol was then poured out of the gel apparatus, rinsed with deionised water and excess water blotted. The stacking gel was composed of 1.25 ml 1 M Tris-HCl, pH 6.8; 100 µl 10% SDS; 1.3 ml 30% acrylamide; 7.35 ml H_2O; 100 µl 10% ammonium persulphate and 10 µl TEMED. This was immediately transferred on top of the separating gel, a well-comb added, and allowed to solidify. The gel apparatus was then assembled into the tank and the tank filled with running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3). Samples were mixed with the appropriate volume of 5x SDS loading buffer (250 mM Tris/HCl pH 6.8, 50% glycerol, 0.5% bromophenol blue, 10% SDS, 500 mM 2-mercaptoethanol) and heated at 100°C for 5 min to denature the proteins. Wells were loaded with 8 µl sample alongside one well with 5 µl pre-stained 10 - 250 kDa protein ladder (PageRuler Plus, thermo Scientific). Gels were run at 200 V until the marker dye-front ran off the gel. Gels were stained with Instant*Blue* Coomassie stain (Expedeon).

2.8.1.2 Western Blotting

Proteins were blotted on to a nitrocellulose membrane (Protan™ BA85) using a Trans-Blot® semi-dry transfer cell (Biorad) according to the manufacturer's instructions. Once blotted, the membrane was stained with 0.1% (w/v) poncean S in 5% (v/v) acetic acid to check all protein bands had transferred to the membrane. The membrane was then incubated with blocking buffer (5% (w/v) skimmed milk powder in 1x Tris-buffered saline-tween (TBST) buffer (50 mM Tris, 150 mM NaCl, 0.05% Tween 20, pH 7.6)) for 75 min with gentle rocking. After blocking, the membrane was washed three times with 1x TBST buffer for 5 min with gentle rocking and then incubated with blocking buffer supplemented with anti-GST–Peroxidase Conjugate antibody (Sigma) at 1:10,000 dilution for 90 min. The membrane was then washed three times with 1x TBST buffer for 5 min with gentle rocking. Horseradish peroxidise (HRP) activity was detected using SuperSignal West Pico Chemiluminescent Substrate kit (Thermo Scientific) according to manufacturer's instructions.

2.8.2 Expression system

Recombinant expression of the Brachypodium HCT genes (Bradi3g48530, Bradi4g36830 and Bradi5g14720) was achieved using the pGEX vector system (GE Life Sciences). Gene expression in this system is controlled by the *tac* promoter which is induced by the addition of isopropyl α-D-thiogalactoside (IPTG), a lactose analogue. The system generates a Glutathione-S-Transferase (GST) tag at the N-terminal of the expressed protein which allows for affinity purification of the recombinant protein. In order to achieve good expression in *E. coli*, the coding sequences of all recombinantly expressed proteins were synthesised by GeneArt® (Life Technologies) with codon optimisation for expression in *E. coli,* and BamHI and EcoRI restriction enzyme sites introduced at the 5' and 3' end of the gene respectively. The GeneArt storage plasmid and pGEX plasmid were digested with the BamHI and EcoRI restriction enzymes and the synthesised gene separated from the storage plasmid by gel electrophoresis.

Subsequently, the synthesised gene and pGEX plasmid DNA were purified by gel purification and PCR purification respectively. The synthesised gene was then ligated into the pGEX-4T3 plasmid and the resulting pGEX-4T3 plasmid containing the synthesised genes transformed into *E. coli* solopack® gold competent cells (Agilent Technologies, UK), using 2 µl ligation mixture. Single colonies were selected and presence of the inserted gene was confirmed by colony PCR using pGEX primers and DreamTaq DNA polymerase. Plasmid DNA was extracted from positive colonies and the plasmids Sanger sequenced using pGEX forward and reverse primers to check the gene had been inserted in frame. Plasmids were then transformed into *E. coli* BL21 cells for protein expression, using 20 ng plasmid DNA. Again, single colonies were selected and presence of the inserted gene confirmed by colony PCR, using pGEX forward and reverse primers and DreamTaq DNA polymerase.

2.8.3 Testing for optimum conditions for protein expression

Different conditions for induction of protein expression were tested in order to optimise for expression of soluble protein. Conditions tested were different concentrations of IPTG (0.1, 0.5 and 1 mM), different temperatures (20, 30 and 37°C) and different incubation periods (3 and 14 h). For protein expression, a single colony of BL21 cells containing the pGEX plasmid with inserted gene was used to inoculate a 10 ml culture of LB containing carbenicillin (50 mg L^{-1}), and the culture grown overnight at 37°C with shaking at 250 rpm. This culture was used to inoculate 250 ml of fresh medium (LB containing carbenicillin (50 mg L^{-1})) with a 1:100 dilution, which was incubated at 37°C with shaking until it reached an OD at absorbance wavelength 600 nm of 0.6 - 0.7 (approximately 2 h). Once this OD had been achieved, the culture was separated into nine lots of 20 ml cultures in 100 ml conical flasks. To three flasks each, IPTG was added to a final concentration of 0.1, 0.5 and 1 mM. One flask of each IPTG concentration was then transferred to three different temperatures with shaking: 20, 30 and 37°C. After 3 or 14 h incubation, 1 ml culture was transferred from each flask into a

microcentrifuge tube and centrifuged at full speed for 10 min. A 1 ml sample was also taken from the 250 ml culture before protein expression had been induced as a control. The pellet was re-suspended in 0.5 ml 1x phosphate buffered saline (PBS), pH 7.3, the cells lysed by sonication (1 min: 3 sec on, 7 sec off) and centrifuged at full speed for 15 min. The supernatant (soluble fraction) was analysed by SDS-PAGE and Western blot.

2.8.4 Protein expression and purification for activity assays

A single colony of BL21 cells containing the pGEX plasmid with inserted gene was used to inoculate a 10 ml culture of LB containing carbenicillin (50 mg L^{-1}) and the culture grown overnight at 37°C with shaking at 250 rpm. This culture was used to inoculate 200 ml fresh medium (LB containing carbenicillin (50 mg L^{-1})) in a 2 L flask with a 1:100 dilution. This was incubated at 37°C with shaking until it reached an OD of 0.6 – 0.7 at absorbance wavelength 600 nm (approximately 2 h). Once this OD had been achieved, IPTG was added to a final concentration of 0.1 mM and the culture transferred to 20°C with shaking for 14 h to induce recombinant protein expression. The culture was then centrifuged at 4000 g for 15 min, the pellet re-suspended in 10 ml 1x PBS, the cells lysed by sonication (3 min: 3 sec on, 7 sec off) and centrifuged at 4000 g for 15 min. The supernatant (soluble fraction) was used for subsequent purification.

The recombinant protein was purified using batch purification with Glutathione Sepharose 4B (GE Life Sciences) which consists of a glutathione ligand coupled via a 10-carbon linker to highly cross-linked 4% agarose which is optimized to give high binding capacity for GST-tagged proteins. Sepharose beads were washed twice with 10 x volume of binding buffer (1x PBS, pH 7.3). The cell lysate was then incubated with the beads at 4°C for 2 h with end-over-end rotation. The bead mixture was then centrifuged at 500 g, the supernatant collected (flow-through) and the beads washed three times with binding buffer by centrifugation at 500 g. Finally, elution buffer (50 mM Tris-HCl, 10 mM reduced glutathione, pH 8.0) was added and the mixture

incubated at 4°C for 10 min with end-over-end rotation and eluted protein collected by centrifugation at 500 g. Protein concentration in the eluate was measured by the Bradford method (327) and elution steps were repeated until protein concentration was below 0.2 mg ml^{-1}. These eluates were pooled for HCT activity assays. The flow-through, wash and eluates were analysed by SDS-PAGE and Western blot to confirm presence and purity of the recombinant protein. BL21 cells containing an empty pGEX-4T3 vector were put the through the same expression and purification steps and the resulting protein used as a negative control for activity assays.

2.9 Protein extraction from plant material

Total protein was extracted from seven week old plants by grinding approximately 300 mg fresh stem material to a fine powder in LN. Three volumes extraction buffer (20 mM Tris-HCl, pH 7.5, 10 mM DTT, 1% polyvinylpolypyrrolidone (PVPP), 15% glycerol and 1 x cOmplete Mini Protease Inhibitor Cocktail) was then added and the proteins extracted on ice for 1 h. The mixture was then centrifuged for 10 min at 17,000 g at 4°C. The supernatant was removed, re-centrifuged at 17,000 g for 10 min at 4°C and the supernatant used for HCT activity assays. Protein concentration was measured by the Bradford method (327).

2.10 HCT activity assays

The reaction mixture for HCT activity assays consisted of: 20 mM Tris–HCl buffer (pH 7.5) 1 mM DTT, 100 μM p-coumaroyl-CoA, 100 μM shikimic acid and 1 μg purified recombinant protein or 10 μg extracted plant proteins, in a total volume of 100 μl. For the boiled enzyme negative control, plant extract was incubated at 100°C for 30 min

prior to adding to the reaction. The reaction was initiated by addition of enzyme, incubated at 30°C and terminated by heating for 5 min at 100°C. The reaction was then cooled on ice and centrifuged at 20,000 g for 5 min to pellet any protein.

2.10.1 Determination of kinetic parameters

In most cases Km and Vmax were determined by incubating 0, 10, 20, 50 or 100 µM of the varying substrate and 500 µM saturating substrate with 3 µg purified recombinant protein in the presence of 20 mM Tris–HCl buffer (pH 7.5), in a final volume of 100 µl. For BdHCT1, when *p*-coumaroyl-CoA was used as the varying substrate and quinic acid as the saturating substrate, and when quinic acid was used as the varying substrate and *p*-coumaroyl-CoA as the saturating substrate, the varying substrate was used at concentrations: 0, 10, 20, 50, 100 and 200 µM. For BdHCT2, when *p*-coumaroyl-CoA was used as the varying substrate and shikimic acid as the saturating substrate, and when quinic acid was used as the varying substrate and *p*-coumaroyl-CoA as the saturating substrate, the varying substrate was used at concentrations: 0, 10, 20, 50, 100 and 200 µM. The reaction was initiated by addition of enzyme and incubated at 30°C, with time points at 0, 1, 2, 3, 5, 7, 10 and 15 min. The reaction was terminated by heating for 5 min at 100°C, then cooled on ice and centrifuged at 20,000 g for 5 min to pellet any protein. After quantification of reaction products, Michaelis-Menton equations and graphs were calculated using the Enzyme Kinetics add-in in SigmaPlot. Glycerol was also tested as an acyl group acceptor with *p*-coumaroyl-CoA, caffeoyl-CoA and feruloyl-CoA as the donors by incubating 500 µM of the acyl donor and acceptor with 10 µg purified recombinant protein in the presence of 20 mM Tris–HCl buffer (pH 7.5) in a final volume of 100 µl, but no activity was observed.

2.10.2 Identification and quantification of reaction products

Reaction products were identified and quantified by liquid chromatography-mass spectrometry (LC-MS), using a Waters Acquity UPLC connected to a LTQ Orbitrap

Mass spectrometer (Thermo Scientific) by Dr. Tony Larson and David Harvey, CNAP, University of York. Chromatographic separation was performed on an Acquity BEH C18 1.7 μm (2.1 x 150 mm) column (Waters), using a gradient elution. The mobile phase consisted of solution A: 99% water, 1% acetonitrile, 0.1% formic acid, and solution B: 1% water, 99% acetonitrile, 0.1% formic acid. A flow rate of 0.35 ml min^{-1} was used and the gradient was as follows: 0 min: 95% A, 5% B; 10.5 min 79.2% A, 20.8% B; 11 min: 0% A, 100% B; 13 min: 0% A, 100% B; 13.1 min: 95% A, 5% B; 15 min 95% A, 5% B. After separation the eluent was directed to the mass spectrometer by heat electrospray ionisation (HESI) in negative mode. Reaction products were characterised by their m/z, retention time and fragmentation spectra. *p*-coumaroyl shikimate: m/z = 319, retention times = 8.00 min (*trans* isomer) and 9.69 min (*cis* isomer); caffeoyl shikimate: m/z = 335, retention time = 6.05 min; feruloyl shikimate: m/z = 349, retention time = 9.09 min; *p*-coumaroyl quinate: m/z = 337, retention time = 5.26 min. The retention times and MS spectra were confirmed with authentic reference compounds which were synthesised by Dr. Dharshana Padmakshan in Prof. John Ralph's lab (University of Wisconsin-Madison, USA).

2.11 Real time PCR (RT-PCR)

2.11.1 RNA extraction

For gene expression analysis, ~100 mg tissue was harvested per biological replicate from leaves, nodes, spikelet, basal internodes (first three internodes), middle internodes (next two internodes) and top internode (apical internode holding the spikelet) of 8 week old plants. Tissue was ground in LN using a pestle and mortar and total RNA extracted using an RNeasy Plant Mini Kit (Qiagen) according to the manufacturer's instructions. RNA was quantified and assessed for quality using a tapestation 2200 (Agilent Technologies) with RNA standard sensitivity screen tape. RNA was then

DNase treated using RQ1 RNase-free DNase (Promega). Briefly, 1 μg RNA in a volume of 8 μl was incubated at 37°C with 1 μl DNase and 1 μl 10x buffer for 30 min and then the reaction stopped by heating at 65°C for 10 min. cDNA was then synthesised from 1 μg DNA-free RNA using Superscript® II Reverse Transcriptase kit (Invitrogen), according to the manufacturer's instructions. Briefly, 1 μl 10 mM dNTPs and 1 μl 100 mM oligo dTs were added to the above reaction and incubated at 65°C for 5 min, followed by incubation on ice for 2 min. To the reaction, 4 μl 5x First-Strand buffer, 2 μl DTT and 1 μl Ribolock RNase inhibitor was added, incubated at 42°C for 2 min, then 1 μl Superscript® II Reverse Transcriptase added and the reaction incubated at 42°C for 50 min, followed by 70°C for 15 min. The resulting cDNA was then made up to 100 μl with nuclease-free water before being used as a template for RT-PCR.

2.11.2 RT-PCR

RT-PCR was carried out on a StepOnePlus™ System (Applied Biosystems) in 96-well plates. In each well was a 20 μl reaction consisting of 10 μl 2x Fast SYBR® Green Master Mix (Applied Biosystems), 1 μl primer pair at 10 mM, 3 μl cDNA from above and 6 μl nuclease-free water. Reactions with nuclease-free water in place of cDNA were used as negative controls. Samples were ran in triplicate with the following protocol: one holding stage at 95°C for 20 sec, 40 amplification cycles at 95°C for 3 sec followed by 60°C for 30 sec, and one melt curve consisting of 95°C for 15 sec, 60°C for 1 min and then an increase from 60 to 95°C at a gradient of 0.3°C. In order to check the efficiency of the primers, standard curves were carried out using a ten-fold dilution series. Primers were selected for experimentation that did not differ in efficiency by more than 10%. RT–PCR data for genes of interest were normalized against an endogenous control gene, Ubiquitin10 (Ub10), which has been shown to be stable across different tissues in Brachypodium (328). Primers for Ub10 bound to the 3' UTR in order to more specifically measure expression of this gene, as demonstrated by

Chambers et al (329). Primer information is shown in Table 4 and all primers were synthesised by Sigma.

Table 4: List of primers used for RT-PCR experiments, their sequences Tms and efficiencies

Primer	F/R	Sequence	Tm (°C)	Efficiency of pair
Bradi3g48530	F	GCCGAGCACATGGAGAAGTT	59	108.1
Bradi3g48530	R	GTGGCGTGTGCTGTTTGTTGGT	60	
Bradi5g14720	F	GTGTGTGTTCTTGCTCTTCTTGAGTAC	59	108.1
Bradi5g14720	R	AGTCTTATTTGTATACCACGTGCAGTTT	59	
Bradi4g36830	F	GCCCGTCTTCCAGCACAA	59	103.3
Bradi4g36830	R	ATCTGAACCTGCACCTACTACTCCAT	59	
Ub10	F	TGGACTTGCTTCTGTCTGGGTTCA	59	105.9
Ub10	R	TGGTAGACAGGCATAACACTGACG	60	

2.12 Crossing Brachypodium

Brachypodium plants were crossed using the protocol developed at the USDA (http://www.ars.usda.gov/SP2UserFiles/person/1931/BrachypodiumCrossing.pdf) with a few alterations. Plants were at the right stage for crossing approximately $4.5 - 5$ weeks after transferring to the greenhouse. Firstly, flowers at the correct stage for emasculation were identified using a magnifying visor; these would be as mature as possible but before anthers had dehisced. Mature stigmas are very feathery, see Figure 13C, D and E for flowers at the correct stage for emasculation and Figure 13A and B for flowers too young for emasculation. Once a flower at the correct stage was identified, all other flowers on the same inflorescence were removed. Using a dissecting microscope, the anthers were then removed using fine, curved end tweezers (7.S, ideal-tek, Switzerland), taking care not to damage the palea or the gynoecium. Removed anthers and the stigma

were examined to check that no pollen had been shed. Finally, it was ensured that the palea and lemma were tightly together and then a small rectangle of cling-film was folded around the emasculated flower to prevent drying out. Pollination was carried out the following day. Pollination was performed in the morning, between 9 and 11 am, and in a room heated to approximately 27°C to increase anther dehiscing. Again, flowers as mature as possible, but before anthers had dehisced (see Figure 13F for an anther at the right stage for pollination, note the yellow colour), were identified and the anthers removed under a dissecting microscope, being careful not to damage the anthers. Anthers were placed on a glass microscope slide and about 20% of the anthers would dehisce after 10 - 40 min. Once the anthers had dehisced (Figure 13G), they were immediately brushed across the stigma of a previously emasculated flower with fine tweezers (size 5, Dumont & Fils, Switzerland), taking care not to spill the pollen or damage the palea and gynoecium (see Figure 13H for a pollinated flower). The pollen only remains viable for about 15 min after dehiscing so this must be done quickly. If the stigma had dried out then it was not pollinated and the flower was removed. The palea and lemma were then closed and held together by a small piece of masking tape, the plants sprayed with water using a hand-sprayer and a transparent plastic bag put over the crossed plants in order to prevent drying out. Bags were removed after four days. The endosperm developed as normal (Figure 13I) and the seeds were harvested once plants were fully mature and dried.
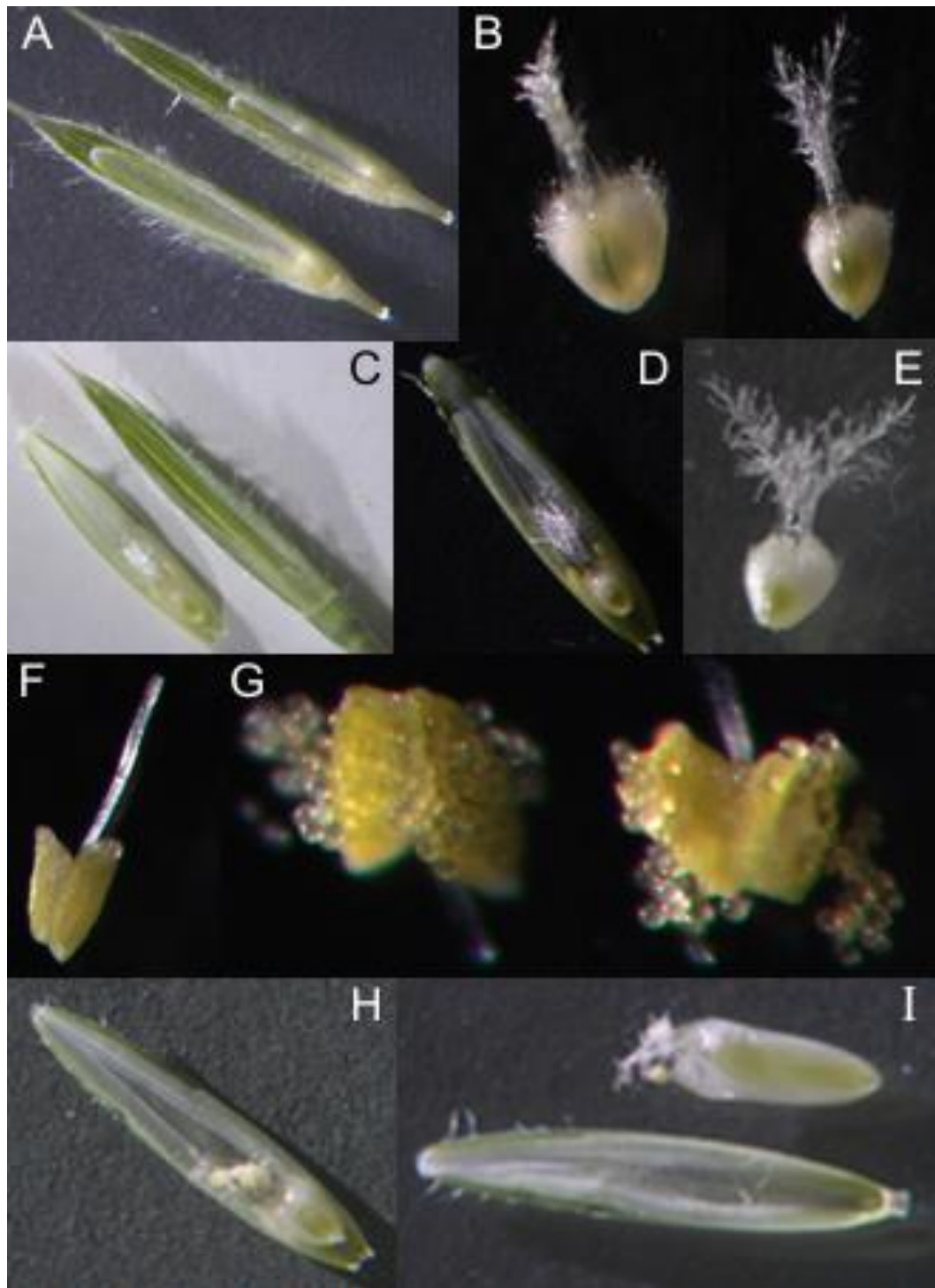
Figure 13: Stages in Brachypodium flower development, pollination and embryo growth.
A: under-developed floret, note non-fully grown palea; B: under-developed embryo and stigma;
C: floret at the right stage of development for crossing, separated into the lemma and palea; D:
close up of the palea in C, with embryo, pollen and stigma inside; E: Close up of embryo and
stigma in C, stigma are very feathery showing they are at the correct stage for crossing; F: Ripe
anthers at the right stage for crossing, note plump appearance and yellow colour; G: dehisced
anther; H: palea with embryo inside that has been pollinated, note emptied anthers within the
dried, desiccated stigma; I: developing embryo following pollination, separated from palea.

2.13 Whole genome sequencing of the *sac3*, *sac4* and *sac9* mutant lines

DNA was extracted from leaf material of *sac3*, *sac4* and *sac9* mutant plants. Paired-end whole-genome sequencing was performed on the DNA by The Genome Analysis Centre (TGAC), Norwich, using an Illumina HiSeq 2000 sequencer. A minimum of 4.7 Gbp data of high quality was achieved for each sample, giving a minimum coverage of 15.6x. Assembly of the sequence data to the reference genome (Bd21 accession) and identification of homozygous SNPs in the mutant plants that did not occur in the wild-type dataset was performed by Dr. Zhesi He and Dr. Yi Li, CNAP Bioinformatics group, University of York. The rest of the analysis was performed by myself. SNPs that did not resemble a mutation caused by sodium azide mutagenesis (i.e. G to A or C to T and vice versa) were removed. This left 1326 SNPs for *sac3*, 1197 SNPs for *sac4* and 1340 SNPs for *sac9*. An excel macro was then created to search the datasets and select for SNPs in genes related to cell wall synthesis and digestibility by key-word searching of the annotation. The annotation included both the annotation of the Brachypodium reference genome and the closest match to each gene in the Arabidopsis reference genome. SNPs were then removed if the SNP was located in an intron or 3'/5' UTR of a gene. The position of the amino acid encoded by the SNP-containing codon was then identified using a program created in Python. SNPs were only retained if they resulted in a change in amino acid. The web software 'SIFT sequence' (http://sift.jcvi.org/www/SIFT_seq_submit2.html) was then used, which predicts whether an amino acid substitution will have a phenotypic effect. This software works by performing a BLAST search for similar sequences and aligning these sequences. It then gives each amino acid within the protein sequence a score for how well a change to each of the 20 amino acids will be tolerated. This score depends on how well conserved the amino acid is among all the sequences and how different the amino acid properties are. The lower the SIFT score, the less likely that substitution will be tolerated and a SIFT score of below 0.05 for a substitution is predicted to affect the structure/function of a protein.

2.14 Confirming the SNP identified in Bradi5g14720 in *sac4*

In order to confirm the SNP in the Bradi5g14720 gene in the *sac4* mutant identified by whole-genome sequencing, cloning and Sanger sequencing of this gene was carried out. Primers (Bd5g14720 F and R) were designed for the 3' and 5' UTRs of this gene and a PCR reaction with Phusion HotStart II DNA polymerase was then performed on genomic DNA extracted from four wild-type and four *sac4* plants. The resulting 1,669 bp fragment showed as a clean band on an agarose gel. Cloning and sequencing of the PCR reaction was performed as previously described but with an internal sequencing primer (Bd5g14720_inner F) in addition to the M13 F and R primers. Five colonies were sequenced for each *sac4* and wild-type plant.

2.15 Mapping the causal mutations of the *sac1* and *sac2* lines

2.15.1 Bulked segregant analysis

The *sac1* and *sac2* mutant lines were backcrossed to wild-type and the resulting BCF_1 progeny allowed to self. The resulting BCF_2 seeds were collected, 260 from one BCF_1 plant for each mutant sown and stem material from the resulting BCF_2 plants subject to saccharification analysis. Homozygous mutant BCF_2 plants were identified by their saccharification phenotype.

2.15.2 Whole genome sequencing of homozygous mutant BCF_2 DNA

For *sac1* and *sac2*, DNA from the top 40 BCF_2 plants with the highest saccharification was obtained by grinding ≤100 mg leaf material in LN, followed by extraction using a Biosprint 15 with a Biosprint 15 DNA Plant Kit, according to the manufacturer's instructions. DNA was quantified using a Qubit Fluorometer (Life Technologies) and then pooled for each mutant, with equal amounts of DNA for each plant. Paired-end whole-genome sequencing was performed on the DNA by TGAC, Norwich, using an

Illumina HiSeq 2000 sequencer. 11.9 and 12.1 Gbp data was achieved for *sac1* and *sac2*, giving coverage of 43.8 and 44.4x, respectively. Assembly of the sequence data to the reference genome (accession Bd21) and identification of homozygous and heterozygous SNPs in the BCF_2 pools of the mutant lines that did not occur in the wild-type dataset was performed by TGAC. The rest of the analysis was performed myself. SNPs were removed that did not resemble a mutation caused by sodium azide mutagenesis (i.e. G to A, C to T or A to T and vice versa), if the phred quality score was <20, if the read depth was below five and if the strand bias had a score of >1. This left 1125 SNPs for *sac1* and 1498 SNPs for *sac2*.

Chapter 3 – Screening for and characterisation of high saccharification mutants in *Brachypodium distachyon*

3.1 Introduction

The dwindling reserves of fossil fuels coupled with the rapidly increasing world population and the environmental effect of burning these fuels mean that more sustainable alternatives are required (1). One of the most promising candidates for the replacement of fossil fuel-based liquid transportation fuels is biofuel, produced from plant biomass. Bioethanol is produced by breaking down plant sugars into their monosaccharide constituents, followed by fermentation to ethanol. First generation bioethanol is produced from parts of the plant that are rich in starch or sucrose, such as corn grain or sugar cane, which can easily be broken down into monosaccharides. Although this bioethanol can be produced efficiently (5), there are disadvantages to its production, such as competition with food supply (7, 8) and, in some cases, the little, if any saving in GHG emissions (11, 12). Second generation bioethanol is produced from lignocellulosic biomass, which is composed of plant cell walls. This biomass can be obtained from agricultural residues, forestry waste, municipal waste or low input grass crops, which have the potential to be grown on degraded land. Second generation bioethanol therefore avoids the issue of food competition and has been shown to result in greater GHG savings than first generation bioethanol (4). However, although lignocellulosic biomass contains a large amount of potential sugar, it has a very recalcitrant nature, making it difficult to convert to the monosaccharides required for fermentation (25). The production process therefore requires expensive and energy intensive pretreatments and high enzyme loadings, making it a commercially unviable product (206, 207).

A number of studies have researched into ways in which the recalcitrance of lignocellulosic biomass can be reduced, in order to decrease energy and enzyme inputs

and increase the release of sugar from the biomass (saccharification) (reviewed in 330-334). Although a number of these studies have been successful in producing biomass with increased saccharification, the majority have taken a reverse genetic approach and so have been limited to the genes already known to be involved in the synthesis of cell wall components. Furthermore, they have often resulted in reduced biomass, fitness and/or stem strength of the plants (reviewed in 28, 240, 243, 273, 335). Plant cell wall synthesis is a complex process, controlled by a large number of genes, many of which are not yet known (314). Therefore, a forward genetic approach has potential to identify novel genes that can affect cell wall saccharification. By identifying multiple factors that impact on saccharification, these have the potential to be combined to obtain even greater lignocellulose digestibility. Furthermore, a forward genetic approach has the potential to identify plants with improved saccharification potential but with normal growth and development, that would not be affected in field performance. In this study, a chemically mutagenised population of plants was screened for mutant lines with increased cell wall saccharification compared to wild-type. The cell walls of the high saccharification lines identified were then characterised to identify any alterations compared to wild-type, in order to investigate which components appear to be important for saccharification. Subsequently, the mutations responsible for the increased saccharification can be mapped, enabling the potential discovery of novel genes with an effect on cell wall saccharification.

3.2 A high-throughput screen for saccharification

An automated assay system has been developed in the McQueen-Mason group at the University of York to measure saccharification of lignocellulosic biomass in a high throughput manner (318). The system can reliably detect differences in the ease of saccharification of plant biomass and is able to process a large number of samples with

minimum human intervention. The first step of the screen is a robotic platform that grinds and accurately weighs the biomass to be tested into a 96-well plate. This 96-well plate is then transferred to a liquid-handling robotic platform which performs a mild pretreatment, followed by partial hydrolysis of polysaccharides with an industrial cellulase cocktail (336). Finally, sugar released from the material during enzymatic hydrolysis is quantified colourimetrically as reducing sugar equivalents using the same liquid-handling platform (318). The mild saccharification treatments employed enable differences in the ease of saccharification of the cell wall to be detected, rather than aiming for maximum sugar release from the material (336).

A variety of pretreatments can be carried out using the robotic platform, including a range of temperatures and solutions at a range of pHs. A number of different pretreatments were carried out on wild-type Brachypodium plants in order to determine the optimal pretreatment conditions with which to screen for variations in saccharification. A range of temperatures were tested: 50, 90 and 130°C, and a range of solutions: water, 0.5 N and 1 N sodium hydroxide, and 0.5 N and 1 N sulphuric acid. The level of saccharification resulting from enzymatic hydrolysis following these pretreatments is shown in Figure 14. It was found that, in general, saccharification increased as pretreatment temperature increased, however there was a much bigger increase between 50 and 90°C than between 90 and 130°C. For example, for 0.5 N sodium hydroxide and 0.5 N sulphuric acid respectively, between 50 and 90°C saccharification increased by 54% and 84%, whereas between 90 and 130°C saccharification only increased by 8% and 21%. Interestingly, pretreatment with sodium hydroxide resulted in much higher levels of saccharification than acid pretreatment; at 90°C the 0.5 N sulphuric acid pretreatment yielded 50% more reducing sugars compared to hot water treatment, while pretreatment with 0.5 N sodium hydroxide yielded over three times the amount compared to hot water and over twice the amount compared to acid. Also of note is the fact that increasing the concentration of the acid or

alkali used for pretreatment made no difference to sugar release at 50°C, and at 90°C and 130°C, actually reduced sugar release. This could be because monosaccharides are being released from the biomass and solubilised during the harsher pretreatment. In the saccharification assay used in this study, a wash step is carried out between the pretreatment and enzymatic hydrolysis steps. A harsher pretreatment could therefore result in less total sugar in the starting material for the enzymatic hydrolysis. Using these observations, a 0.5 N sodium hydroxide pretreatment carried out at 90°C was selected for the mutant screen. This was because sodium hydroxide resulted in the highest saccharification and, although 130°C resulted in higher saccharification than 90°C, this was only a small and non-significant increase (t-test, $p = 0.1$) (Figure 14).
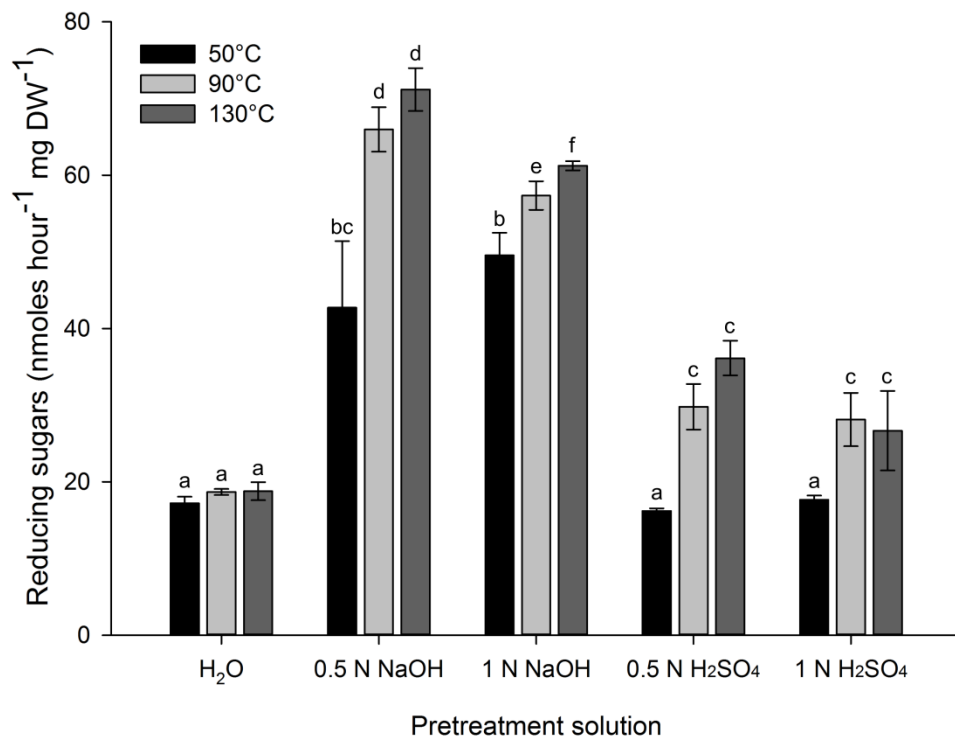
Figure 14: Quantification of sugar release after enzymatic hydrolysis of wild-type Brachypodium stems following pretreatment with a range of conditions. Pretreatment conditions were: water, 0.5 N and 1 N sodium hydroxide, and 0.5 N and 1 N sulphuric acid at 50, 90 and 130°C. Data represents mean ± SD, n = 3.

3.3 Identification of mutants with increased saccharification

A population of 2500 sodium azide mutagenised M_2 Brachypodium plants was used for the saccharification screen. This population was produced by Dr. Richard Sibout and Prof. Herman Hofte at INRA, Versailles (317). Sodium azide mutagenesis has been shown to result primarily in base substitutions and induces very few chromosome aberrations (337). It has also been shown that sodium azide mutagenesis results primarily in nucleotide transitions rather than transversions, i.e. the interchanging of A to G or C to T, although A to T mutations are also observed (338). The mutant plants were grown to maturity and straw from the primary and secondary stems was harvested and screened for saccharification using the high-throughput saccharification assay described above, with one wild-type plant per 96-well plate for comparison (each 96-well plate holds 20 samples). The screen revealed a normal distribution for variation in saccharification across the population of mutant plants (Figure 15A) (p = 0.496, Kolmogorov-Smirnov test), with a maximum saccharification 67.03% higher than wild-type and a minimum saccharification 30.31% lower than wild-type. Mutant plants were selected for further analysis if they met two criteria; the mutant saccharification value had to be more than 20% above the wild-type plant on the same 96-well plate and this difference significant (t-test, $p \leq 0.01$), and the mutant saccharification value had to lie outside the 95% confidence intervals of the wild-type population. Using these criteria, 37 mutants with high saccharification were selected. In order to examine the heritability of this high saccharification phenotype, 12 seeds from each of these 37 mutant plants were sown and the resulting plants subject to saccharification analysis. Again, a wild-type plant on each 96-well plate was used for comparison. This revealed 12 mutant lines in which the high saccharification phenotype was retained in all of the offspring (Figure 16). These mutant lines were deemed to have heritable increases in saccharification and they were named *saccharification1* (*sac1*) to *sac12*. The average increase in saccharification in these 12 mutant lines ranged from 20 to 58% above wild-type level (Figure 15B). The other 25 mutants that showed high saccharification in the initial

screen did not retain this high saccharification consistently in the next generation. The level of saccharification of three of these mutants and their offspring are shown in Figure 16 and it is clear that there does not appear to be any heritability pattern of the saccharification phenotype.
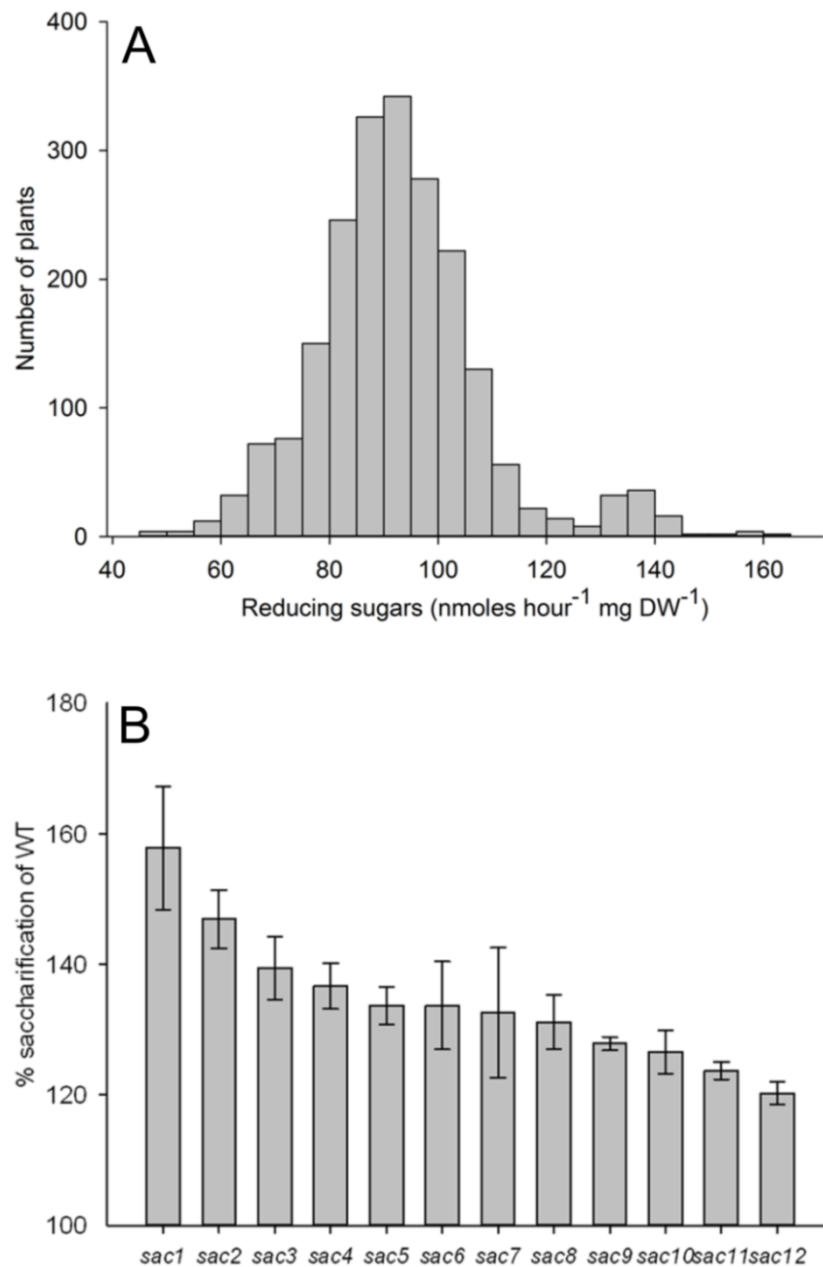
Figure 15: Sugar release from mutant Brachypodium stems after 0.5 M NaOH pretreatment at 90°C and enzymatic hydrolysis. (*A*) Histogram showing distribution in saccharification of the whole mutant population in the initial screen. (*B*) Mean saccharification of the offspring of mutants with a heritable increase in saccharification observed in the second screen. Data represents mean ± SD, n = 7-11.
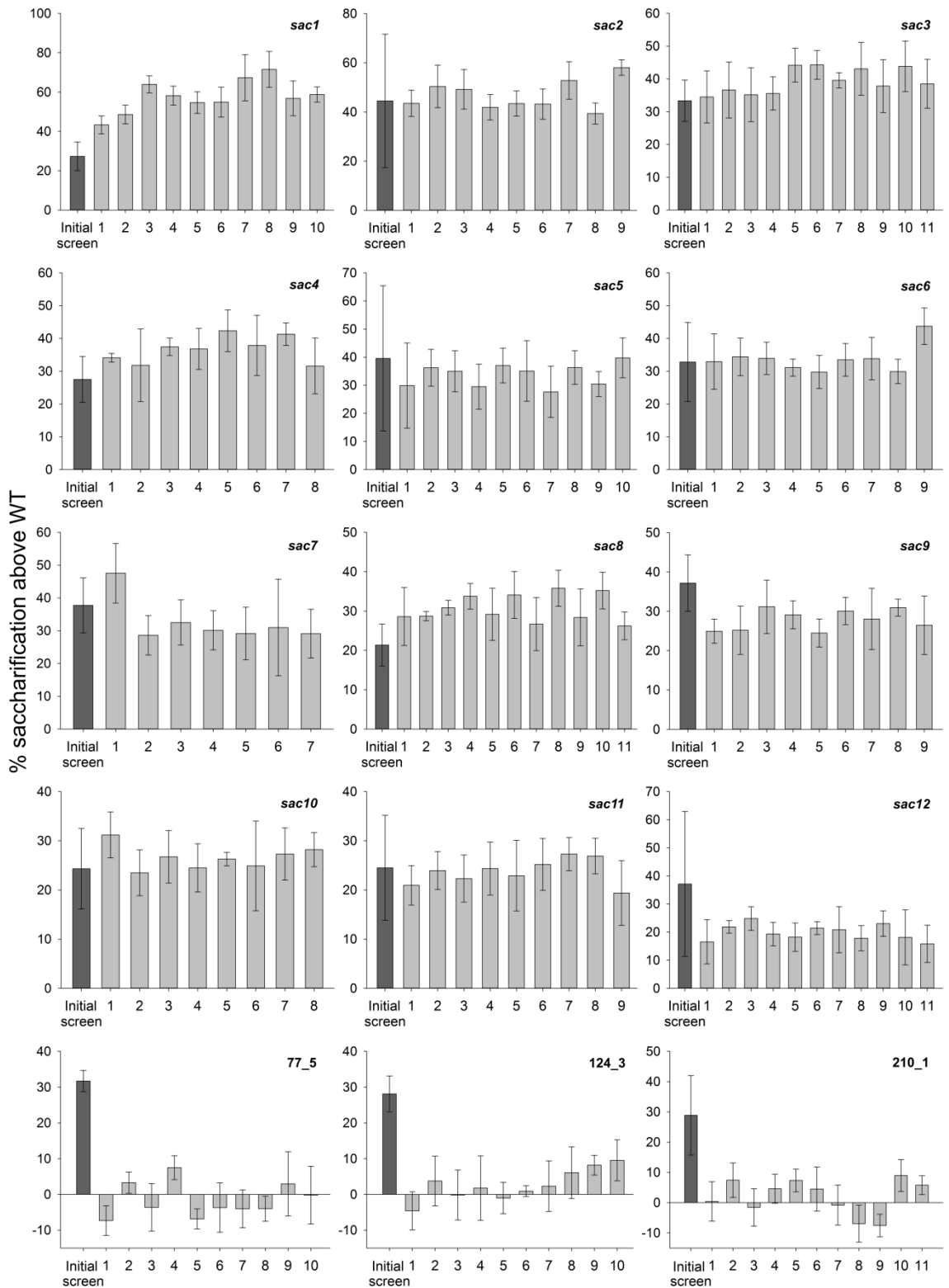
Figure 16: Saccharification of the 12 mutants which had heritable increases in saccharification and three mutants that did not show a heritable increase in saccharification. Bar graphs show the level of saccharification seen in the initial screen and in the offspring of the individual mutants. Data represents mean ± SD, n = 4 technical replicates.

3.4 Cell wall composition of the high saccharification mutants

In order to investigate the compositional changes underlying the increases in saccharification observed in the *sac* mutant lines, the cell wall composition of these mutants was analysed. Content of lignin, cellulose, hemicellulose and ferulic acid was examined, as well as lignin and hemicellulose composition. Cell wall composition analyses were carried out on stem material that had been subjected to sequential extraction with chloroform and methanol. This gave a crude cell wall residue which will be referred to as alcohol insoluble residue (AIR).

3.4.1 Lignin content and composition

Examination of lignin content using the acetyl bromide method (320) revealed that none of the mutant lines had an increase in lignin content over wild-type, whereas six mutants (*sac3, 4*, *5*, *9, 11* and *12*) had significantly lower lignin content than wild-type (Table 5). The *sac4* mutant showed the most severe alteration in lignin content, with a reduction of 35% compared to wild-type, while more moderate reductions were seen for *sac12* (-26%), *sac5* (-24%), *sac11* (-22%), *sac3* (-20%) and *sac9* (-16%).

Lignin composition of the *sac* mutants was investigated using thioacidolysis. This work was carried out by the group of Prof. Catherine Lapierre at INRA, Versailles. The method measures the amount of S, G and H lignin monomers present in the lignin that are linked via β-*O*-4 bonds. The thioacidolysis revealed that five of the *sac* mutants showed a significant reduction in total thioacidolysis yield (H + G + S) compared to wild-type; *sac1*, *sac2*, *sac4* and *sac9* all showed strong reductions (31 − 44%), while *sac5* showed a more modest reduction (19%) (Table 6). In addition, six of the mutants showed alterations in the ratio of lignin monomers compared to wild-type; *sac1*, *4, 6* and *7* all showed a significant decrease in thioacidolysis S:G ratio and an increase in the relative amount of thioacidolysis H monomers (% H units), while *sac2* and *sac9* showed only an increase in % H units (Table 6). The alterations in % H units were particularly

strong, with significant increases ranging from 55 to 76%. The significant increases in % G units ranged from 29 – 34%, while the significant reductions in % S units were more modest (-13 – -16%). There was a strong negative correlation between % S units and % G units) (r = -1.0, p = 0.00) and a strong positive correlation between total thioacidolysis yield and S:G ratio (r = 0.85, p = 0.00), which have both been previously reported (219). There was also a strong negative correlation between S:G ratio and % H units (r = -0.89, p = 0.00). These correlations suggest that there are strong levels of feedback within the lignin monomer synthesis pathway.

Table 5: Cell wall composition of the 12 *sac* mutants.

	Lignin	Cellulose	Hemicellulose	Ferulic acid
wild-type	278.1 ± 19.9	399.5 ± 56.9	386.5 ± 10.6	2.4 ± 0.3
sac1	273.3 ± 10.5	**287.3 ± 8.7***	389.6 ± 16.6	**1.7 ± 0.1***
sac2	267.6 ± 27.2	401.6 ± 51.8	386.2 ± 19.7	2.3 ± 0.5
sac3	**221.3 ± 6.8***	398.7 ± 41.0	376.7 ± 39.3	2.2 ± 0.3
sac4	**182.1 ± 16.4***	333.6 ± 57.4	**443.0 ± 14.4***	**1.7 ± 0.1***
sac5	**211.3 ± 18.9***	343.9 ± 64.7	**457.0 ± 1.2***	2.3 ± 0.6
sac6	273.2 ± 17.5	335.5 ± 30.7	400.1 ± 12.6	2.1 ± 0.4
sac7	256.2 ± 2.8	**274.3 ± 44.5***	**523.7 ± 52.1***	2.4 ± 0.1
sac8	246.1 ± 4.2	402.0 ± 38.2	358.6 ± 13.0	2.0 ± 0.1
sac9	**231.8 ± 17.7***	425.9 ± 1.5	396.8 ± 11.3	2.3 ± 0.5
sac10	250.9 ± 13.9	401.8 ± 61.5	391.3 ± 35.8	2.4 ± 0.0
sac11	**216.8 ± 28.6***	353.4 ± 61.6	394.2 ± 23.3	2.2 ± 0.4
sac12	**205.8 ± 27.5***	332.0 ± 29.7	375.3 ± 25.5	2.4 ± 0.0

Data expressed in mg g^{-1} AIR and shows mean ± SD (n = 4). Asterisks and bold lettering represent significant difference (p ≤ 0.05) when compared to wild-type.

Ferulic acid content was measured in the *sac* mutants by hydrolysis of the ester bonds between ferulic acid and hemicellulose with sodium hydroxide, followed by analysis of the products by high performance liquid chromatography (HLPC). This revealed that two of the mutants, *sac1* and *sac4,* had significantly lower ferulic acid content than wild-type (-28% and -27% respectively) (Table 5).

Table 6: Total thioacidolysis yield and lignin monomer relative mol % of wild-type and *sac* mutant stem AIR.

	Thioacidolysis yield (μmol g^{-1} AIR)	% H	% G	% S	S:G
wild-type	142.0 ± 5.4	3.2 ± 0.1	26.7 ± 0.7	70.1 ± 0.7	2.63
sac1	**96.3 ± 3.6***	**5.2 ± 0.1***	34.3 ± 1.3	**60.6 ± 1.2***	**1.78***
sac2	**97.8 ± 1.8***	**4.9 ± 0.2***	28.3 ± 0.4	66.8 ± 0.5	2.36
sac3	116.5 ± 8.3	4.17 ± 0.2	27.5 ± 0.3	68.3 ± 0.4	2.49
sac4	**79.4 ± 3.4***	**5.1 ± 0.1***	**34.4 ± 0.1***	**60.5 ± 0.0***	**1.76***
sac5	**114.7 ± 0.9***	4.1 ± 0.3	27.3 ± 0.2	68.6 ± 0.3	2.51
sac6	97.9 ± 7.6	**5.1 ± 0.2***	**34.8 ± 0.6***	**60.1 ± 0.8***	**1.73***
sac7	78.4 ± 9.7	**5.6 ± 0.2***	**35.8 ± 0.6***	**58.6 ± 0.4***	**1.64***
sac8	146.2 ± 3.9	3.4 ± 0.2	25.8 ± 0.3	70.9 ± 0.5	2.75
sac9	**93.0 ± 2.2***	**5.4 ± 0.1***	33.3 ± 1.2	**61.2 ± 1.2***	1.84
sac10	155.1 ± 9.1	3.1 ± 0.2	28.0 ± 0.9	68.9 ± 1.1	2.47
sac11	119.8 ± 6.9	4.0 ± 0.1	28.7 ± 0.5	67.3 ± 0.6	2.34
sac12	139.2 ± 2.2	3.7 ± 0.1	26.8 ± 0.7	69.5 ± 0.6	2.59

Data shows mean ± SD, n = 3. Asterisks and bold lettering represent significant difference (p ≤ 0.05) when compared to wild-type.

3.4.2 Polysaccharide content and composition

Hemicellulose content and composition were measured by hydrolysis into component monosaccharides using trifluoroacetic acid (TFA), followed by analysis and quantification by high performance anion exchange chromatography (HPAEC). Three mutant lines, *sac4*, *sac5* and *sac7*, exhibited significantly higher total hemicellulose content than wild-type, whereas no mutants showed a significant reduction (Table 5). The *sac7* mutant showed quite a strong increase in total hemicellulose (29%), while *sac4* and *sac5* showed more moderate increases (17% and 15% respectively). Table 7 shows the composition of hemicellulose in the mutant lines. For the *sac7* mutant, which had the highest hemicellulose content, the main monosaccharide responsible for this increase was xylose, with arabinose, glucose and mannose also significantly higher than in wild-type. Similarly, the *sac4* and *sac5* mutants exhibited significant increases in xylose, arabinose and mannose compared to wild-type, with fucose also being

significantly increased in *sac4*. The *sac8* and *sac9* mutants, despite not showing a significant alteration in total hemicellulose content, did show significant changes in the content of some of the individual monosaccharides. In the *sac8* mutant, levels of xylose and arabinose were significantly lower than in wild-type and mannose was significantly higher, while the *sac9* mutant showed significant increases in arabinose, galactose and mannose.

Crystalline cellulose content was measured by hydrolysis of AIR with Updegraff reagent to remove hemicellulosic sugars and amorphous cellulose, followed by further hydrolysis of the resulting pellet with sulphuric acid to convert the crystalline cellulose to glucose for spectrophotometric measurement. This revealed that most of the mutants had similar crystalline cellulose content to wild-type, while *sac1* and *sac7* both had significantly reduced levels (-28% and -31% respectively) (Table 5).

Table 7: Monosaccharide composition of hemicelluloses in wild-type and *sac* mutants (µg mg^{-1}).

	Fucose	Arabinose	Galactose	Glucose	Xylose	Mannose
wild-type	2.09 ± 0.99	82.98 ± 3.95	49.72 ± 3.76	53.33 ± 2.76	118.22 ± 5.80	18.98 ± 2.18
sac1	3.15 ± 0.45	86.56 ± 4.04	51.17 ± 3.57	51.43 ± 2.46	123.32 ± 7.99	17.23 ± 2.20
sac2	2.40 ± 0.78	87.99 ± 7.44	50.18 ± 2.64	54.25 ± 4.09	123.32 ± 7.47	16.83 ± 2.82
sac3	1.47 ± 0.80	79.32 ± 4.26	44.42 ± 3.52	61.89 ± 20.57	107.85 ± 5.91	**27.48 ± 3.83***
sac4	**4.53 ± 0.35***	**95.74 ± 5.83***	57.54 ± 5.14	55.98 ± 4.39	**133.30 ± 7.28***	**28.73 ± 0.31***
sac5	3.03 ± 1.32	**94.70 ± 4.04***	55.67 ± 1.41	58.40 ± 0.31	**135.18 ± 2.45***	**28.49 ± 6.46***
sac6	2.66 ± 0.58	87.20 ± 6.72	51.71 ± 4.56	58.85 ± 2.97	123.30 ± 8.17	17.34 ± 0.92
sac7	2.77 ± 0.86	**99.33 ± 6.90***	58.96 ± 8.37	**68.05 ± 2.75***	**159.02 ± 14.40***	**38.64 ± 4.76***
sac8	1.76 ± 0.38	**71.46 ± 4.90***	45.72 ± 4.88	50.10 ± 1.94	**100.54 ± 4.09***	**30.73 ± 4.57***
sac9	3.45 ± 0.18	**92.25 ± 3.19***	**56.36 ± 1.40***	52.73 ± 1.08	125.42 ± 7.14	**22.65 ± 1.17***
sac10	1.69 ± 0.83	80.49 ± 3.26	49.06 ± 2.38	60.79 ± 6.90	118.94 ± 1.84	18.56 ± 1.24
sac11	2.12 ± 0.51	86.67 ± 10.64	47.69 ± 4.08	63.47 ± 14.75	124.06 ± 12.41	16.31 ± 1.99
sac12	2.26 ± 1.23	82.26 ± 12.97	46.86 ± 6.40	53.55 ± 3.34	116.63 ± 14.39	**26.03 ± 1.43***

Data shows mean ± SD, n = 3. Asterisks and bold lettering represent significant difference ($p \leq 0.05$) when compared to wild-type.

3.4.3 Mutants with no major alterations in cell wall composition

The *sac10* mutant showed no major differences in any of the compositional analyses employed in this study. However, stem cross sections of this mutant revealed a much less uniform structure of the metaxylem vessels within the vascular bundles in this mutant compared to wild-type, suggesting that this mutant could have a developmental mutation (Figure 17). Cross-sections of the other *sac* mutants were similar to wild-type.
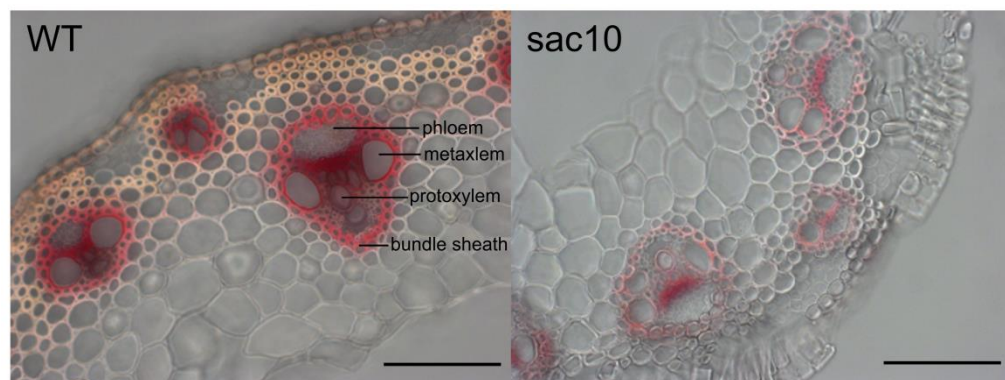
Figure 17: Transverse stem sections of wild-type and *sac10* plants stained using the Weisner method. Objective = 20x, scale bar = 100 μM. Sections of 10 plants were examined to ensure the difference in metaxylem vessel structure was observed in multiple biological replicates. The difference in staining intensity between wild-type and *sac10* in the image was not observed across all biological replicates; these images were chosen because they demonstrate the observed difference in metaxylem vessel structure most clearly.

3.5 Stem mechanical properties of the *sac* mutants

Three-point bending tests were performed to test the strength (maximum force the stem can withstand) and stiffness (initial force required to bend the stem) of stem material from wild-type and mutant plants. This revealed that none of the mutants had a reduction in either stem strength or stiffness compared to wild-type. Indeed, one of the mutants, *sac9*, showed a significant increase in both of these characteristics (Figure 18).
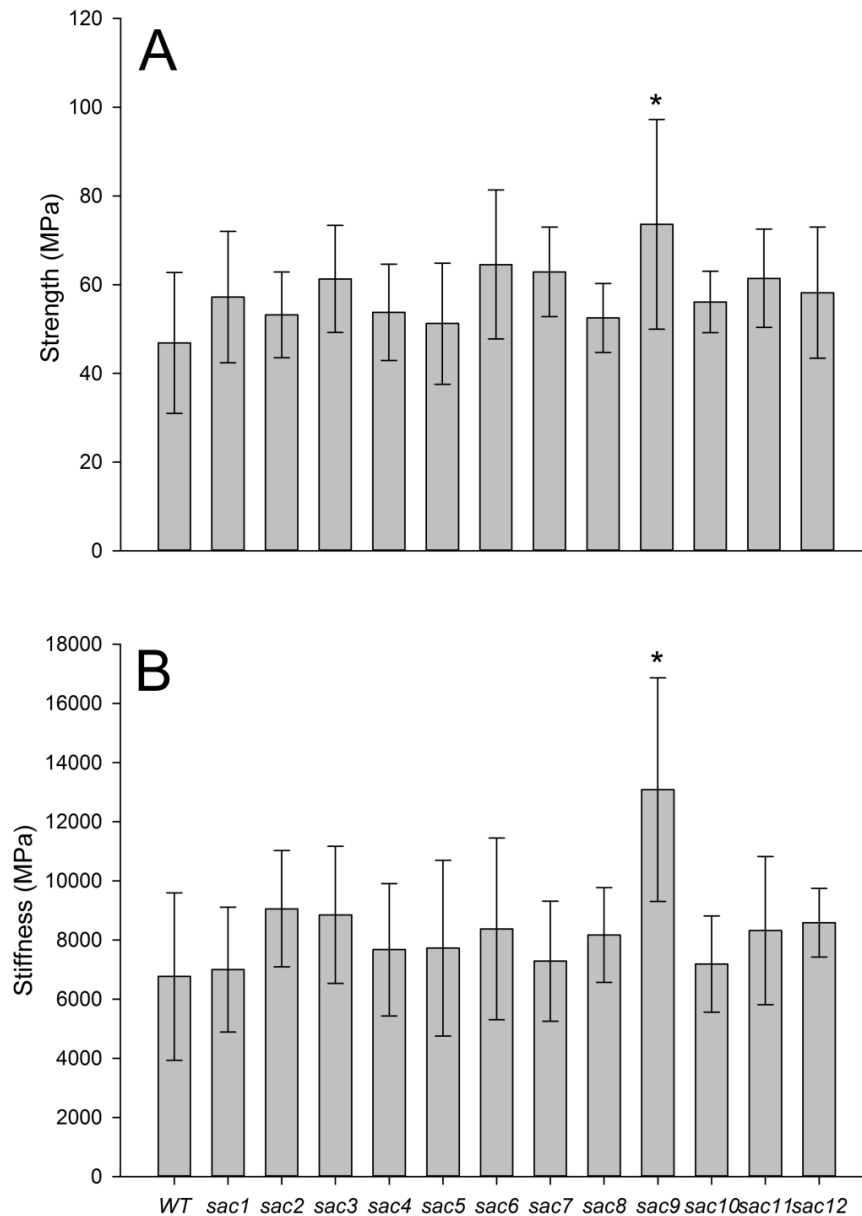
Figure 18: Stem physical properties of *sac* mutant and wild-type plants. (A) Stem strength measured by the maximum bending stress. (B) Stem stiffness measured by the initial force required to bend the stem. Data represents mean ± SD and n = 15. Asterisks represent significant difference ($p \leq 0.05$) when compared to wild-type.

3.6 Development, fitness and morphology of the *sac* mutants

We deliberately excluded severely dwarfed plants from the initial screen in order to avoid complications arising from plants exhibiting large developmental changes, and because such mutants may not be so useful from a biotechnological perspective. Growth

and development phenotyping of the *sac* mutant plants revealed that the majority of the *sac* mutants were similar to wild-type and none of the mutants had any severe alterations (Figure 19). Furthermore, because these mutant plants still contained all of the background mutations caused by the chemical mutagenesis, any change in phenotype may not necessarily reflect how the specific mutation causing the increase in saccharification is affecting phenotype. Most of the mutants had similar percentage germination to wild-type (91%), the exceptions being *sac8* and *sac12* which showed reduced germination (75% and 71% respectively). The majority of mutants also had a similar germination time to wild-type (2.9 days), although *sac1* and *sac6* took significantly longer to germinate (4.5 and 3.6 days respectively), while *sac*5 took significantly shorter (2.1 days). Time taken to flower after germination follows a very similar pattern: *sac1* and *sac6* plants flower significantly later than wild-type, and *sac5* flowers significantly earlier. For the remainder of the phenotypic measurements of *sac6*, this mutant was similar to wild-type. However, *sac1* and *sac5* continue to be affected in other characteristics. The height of *sac1,* which had late germination and flowering, was reduced compared to wild-type, the average internode length was shorter and less seed was produced. The *sac5* mutant, which had early germination and flowering, was the opposite with an increase in height, stem biomass, internode length and seed production compared to wild-type. Other mutants with an altered phenotype were *sac7* which was shorter than wild-type, had shorter internodes and reduced stem biomass. The *sac11* mutant was the opposite with an increase in height, stem biomass and seed production. The *sac2* and *sac3* mutants also had a significant increase in seed production compared to wild-type. Number of internodes and number of tillers was very similar across all lines. In summary, the mutants which seem to be affected negatively are *sac1* and *sac7* which are shorter and produce less seed than wild-type (as well as a reduction in stem biomass for *sac7*), and *sac8* and *sac12* which have lower percentage germination.
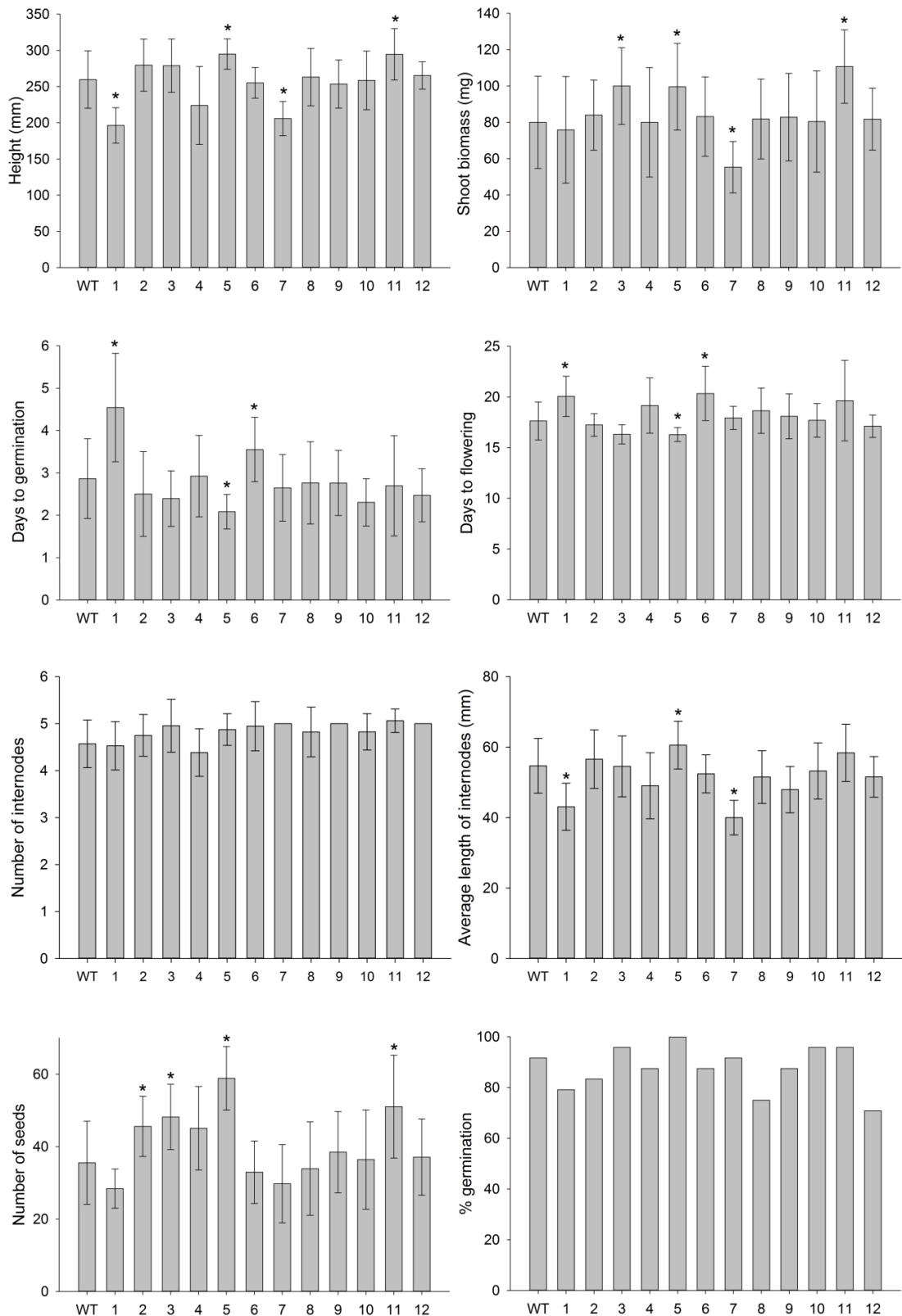
Figure 19: Phenotyping of the growth and development of wild-type and the 12 *sac* **mutants.** Plants were grown to maturity in the glasshouse and allowed to dry before harvesting to measure height, biomass, number/length of internodes and number of seeds. Data represents mean ± SD, n = 24. Asterisks represent significant difference (p ≤ 0.05) when compared to wild-type.

3.7 Relationships between cell wall components and saccharification

In order to investigate whether any associations were apparent between the various cell wall components and level of saccharification among the *sac* mutants, Pearson correlations were calculated (Figure 20). This revealed that there was a significant correlation between ferulic acid and saccharification, with a reduction in ferulic acid content coupled to an increase in saccharification ($r = -0.56$, $p = 0.05$). No correlation was found between lignin content, cellulose content, hemicellulose content or lignin composition and saccharification. Interestingly, no correlation was found between lignin content and the other cell wall components. A reduction in lignin content has previously been thought to be compensated for by an increase in either hemicellulose content (219, 339, 340) or crystalline cellulose (339, 341, 342), but this does not seem to be the case for the *sac* mutants. There was a significant correlation however between cellulose and hemicellulose content ($r = -0.59$, $p = 0.04$), particularly with xylose ($r = -0.63$, $p = 0.03$). This suggests that there could be some sort of compensatory effect between the two components.

3.8 Effect of different pretreatments on the *sac* mutants

Four of the *sac* mutants, *sac1*, *2*, *4* and *11*, were tested with a range of pretreatments before being subject to enzymatic hydrolysis to see if the increased saccharification was maintained. These four mutants were chosen so that a range of cell wall composition alterations were covered. Pretreatments tested were water, 1 N sodium hydroxide, and 0.5 N and 1 N sulphuric acid, in addition to the 0.5 N sodium hydroxide used for the screen. All pretreatments were carried out at 90°C for 20 min. The increase in saccharification of the four mutants compared to wild-type observed with 0.5 N sodium hydroxide in the initial mutant screen is maintained with 1 N sodium hydroxide although, except for the *sac2* mutant, these increases are not as large as at the lower

concentration (Figure 21). Interestingly, the increase in saccharification is not seen when an acid pretreatment is used and the four *sac* mutants show very similar levels of saccharification to wild-type. After hot water pretreatment, two of the mutants, *sac2* and *sac11*, maintain their high saccharification (Figure 21).
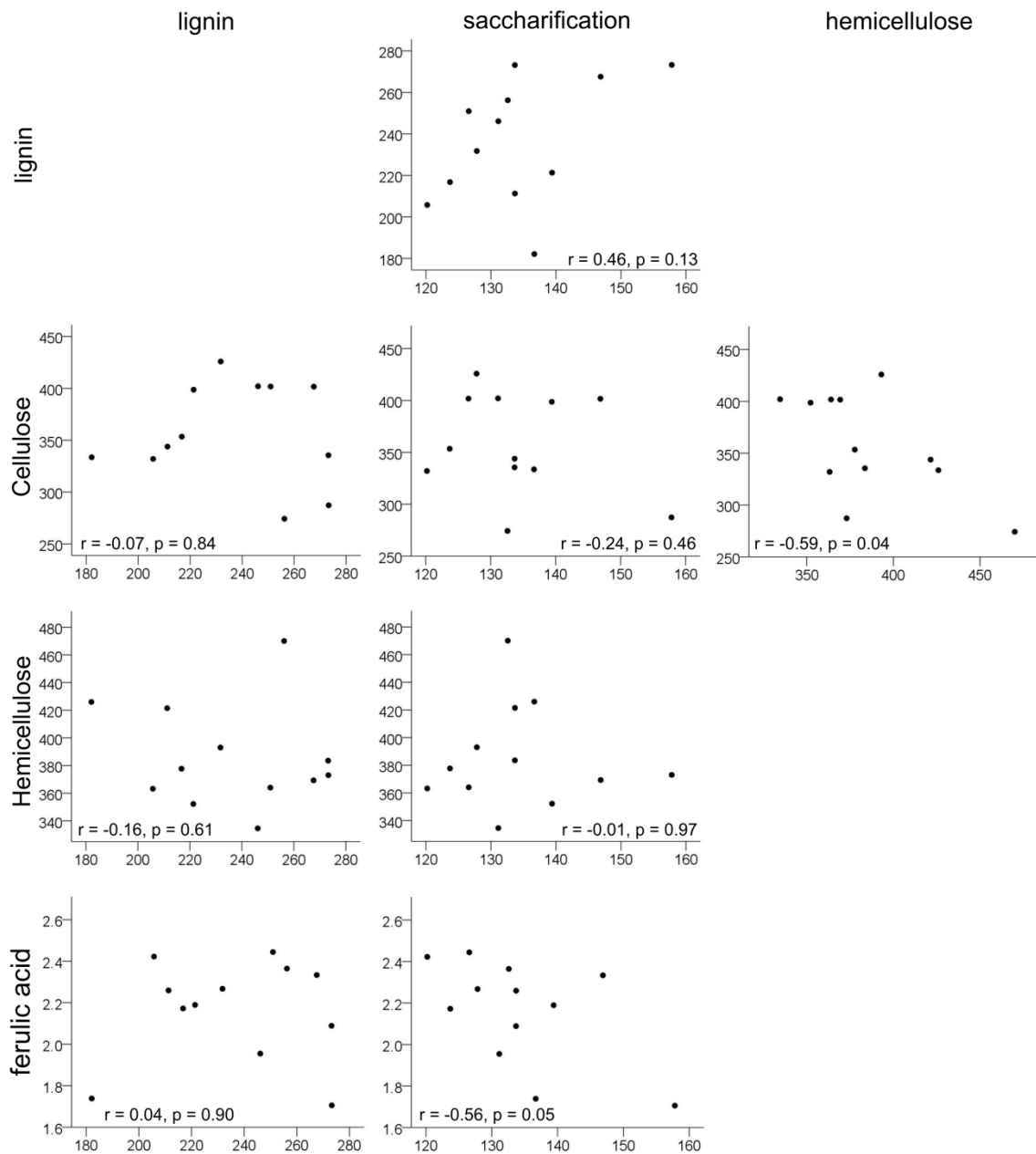
Figure 20: Associations between cell wall components and level of saccharification among the high saccharification mutants. Scatterplots contain data from the 12 *sac* mutants, with the Pearson correlation coefficient and its corresponding p-value shown at the bottom of each plot.
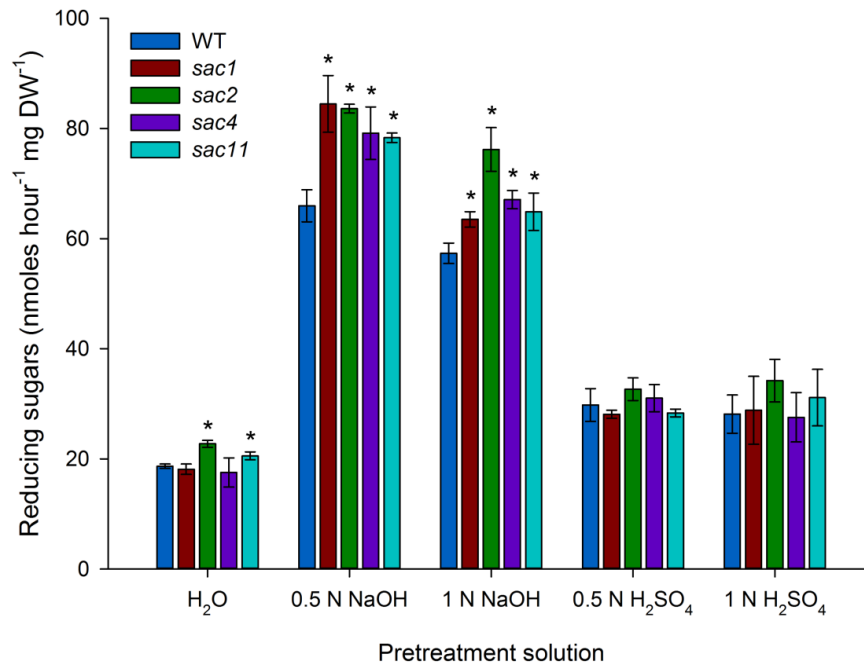
Figure 21: Sugar release after enzymatic hydrolysis of biomass from wild-type and *sac* mutant plants, following a range of different pretreatments. Stem material from four of the *sac* mutants, *sac1*, *sac2*, *sac4* and *sac11*, and wild-type were subject to a range of pretreatments, followed by enzymatic hydrolysis using the high-throughput saccharification assay. Data represents mean ± SD and n = 3. Asterisks represent significant difference ($p \leq 0.05$) when compared to wild-type.

3.9 Discussion

A population of chemically mutagenised Brachypodium plants were screened for saccharification with an industrial cellulase cocktail, resulting in the identification of 12 mutant lines that exhibited significant and heritable improvements in saccharification compared to wild-type. A range of cell wall components and growth characteristics of these mutants were then analysed in order to get a better understanding of the mutant phenotypes. Figure 22 provides a summary of the cell wall composition analysis and it is apparent from this that the high saccharification phenotype is accompanied by a range of cell wall modifications among the *sac* mutants. However, it is also apparent that there appears to be a recurring theme for each of the cell wall components across the *sac* mutants. Lignin content, crystalline cellulose content, ferulic acid content and S:G ratio only show significant decreases among the *sac* mutants, they do not show any

significant increases. Hemicellulose content and % H monomers however only show significant increases. When Van Acker et al. (219) analysed cell wall composition and saccharification in a set of 20 Arabidopsis lignin mutants, they found a similar result; increasing lignin content, S:G ratio and ferulic acid content all had a negative impact on saccharification, whereas increasing hemicellulose content had a positive impact. The authors did not find an impact of cellulose content but this may be because the study only encompassed lignin mutants and thus may not reflect a whole population trend.

3.9.1 Mutants with altered lignin content

It is widely accepted that lignin plays a major role in the recalcitrance of lignocellulose to digestion and many studies have shown that a reduction in lignin content results in an increase in saccharification (reviewed in Li et al. (334)). In line with this, six of the *sac* mutants (*sac3*, *4*, *5*, *9*, *11* and *12*) showed significantly lower levels of lignin than wild-type. Among the six low lignin mutants, in three of them (*sac3*, *sac11* and *sac12*) lignin content was the only cell wall characteristic to be altered significantly from wild-type. Previous studies have shown negative correlations between lignin content and saccharification, but these studies used populations of plants selected for their lignin content or composition (218, 219, 234). Interestingly, when saccharification and lignin content are plotted against each other for the Brachypodium *sac* mutants, no significant correlation was found but the trend indicates a positive relationship, with the lines containing the highest lignin content having the highest saccharification (Figure 20) (r = 0.46, p = 0.13). This perhaps suggests that, although this work does confirm a major role for lignin content in the recalcitrance of the cell wall as none of the mutants had an increase in lignin content compared to wild-type, when lignin is below wild-type level, other cell wall components become important in determining saccharification. Indeed, it is notable that a number of the *sac* mutants, including the two with highest saccharification (*sac1* and *sac2*), show no significant alteration in lignin content compared to wild-type.
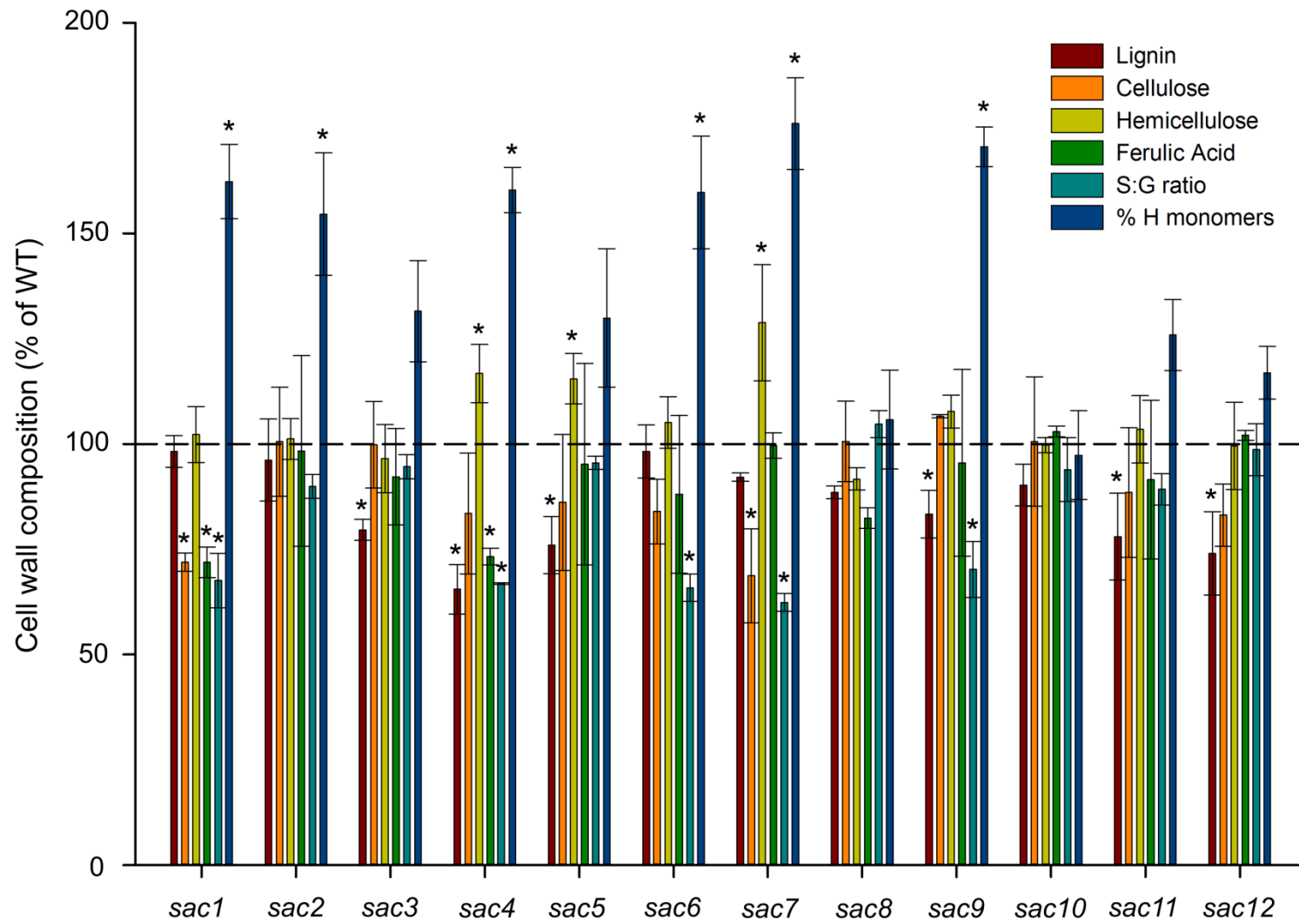
Figure 22: Cell wall composition of stem AIR from the 12 *sac* mutants, expressed as percentage of wild-type. Data represents mean ± SD and n = 4. Asterisks represent significant difference (p ≤ 0.05) when compared to

3.9.2 Mutants with reduced lignin and increased hemicellulose content

In the *sac4* and *5* mutant lines, the reduction in lignin content was accompanied by significant increases in hemicellulose content. Such increases in these polysaccharides has been reported in other plants with altered lignin content (219, 339, 340), and could, in part, be a compensation mechanism, preventing any loss in stem strength and pest/pathogen resistance that could be caused by a reduction in lignin content. Indeed, it has recently been shown through modelling that hemicellulose, and the interactions between hemicellulose and cellulose, make a large contribution to the stiffness of the cell wall (343). Other studies have also suggested a role for hemicellulose in compensating for reduced lignin content (219, 339, 340). The high throughput assay used in this study measures saccharification by quantifying reducing sugars. It therefore does not distinguish between glucose and the monosaccharides released from hemicellulosic polysaccharides. The observed increase in saccharification of *sac4* and *sac5* could therefore be due to the combined action of increased accessibility of the cell wall sugars to cellulolytic enzymes due to the reduction in lignin, and an increase in the total pool of available sugars due to the increase in hemicellulose content. Improvements in enzyme technology have made it possible to ferment mixtures of C5 and C6 sugars for the production of bioethanol (344, 345) and so the release of hemicellulosic sugars is becoming as important as cellulosic sugars. Indeed, fermentation of both cellulosic and hemicellulosic sugars from wheat straw resulted in a significant increase in total ethanol yield (346).

3.9.3 Mutants with alterations in cell wall components other than lignin

3.9.3.1 Cellulose

Our studies revealed a number of high saccharification mutants with little or no change in lignin content, but exhibiting alterations in other wall components. Amongst these was *sac1* which showed the largest increase in saccharification. One of the features of

sac1 cell walls was a significant (28%) decrease in crystalline cellulose content compared to wild-type, which is also seen in the *sac7* mutant (31% lower than wild-type). Crystalline cellulose is highly recalcitrant to digestion, whereas the amorphous regions of cellulose are much easier to digest. Consequently, the amorphous regions are released by the TFA treatment used to release hemicellulosic sugars for quantification, whereas the crystalline cellulose is not. The *sac7* mutant showed an increase in the amount of glucose released by the TFA hydrolysis in comparison to wild-type, which could be attributed to amorphous portions of the cellulose, xyloglucan or MLG. Xyloglucan and MLG are only present in minor amounts in grass secondary cell walls (16, 347), suggesting that the majority of glucose released by TFA treatment will be derived from amorphous cellulose. This, together with the observed reduction in crystalline cellulose, suggests a reduction in the degree of crystallinity of cellulose in the *sac7* mutant. This may well be contributing to the increase in saccharification as the tight packing of glucan chains within highly crystalline cellulose prevents access of digestive enzymes. It has been demonstrated that a reduction in the degree of cellulose crystallinity results in a linear increase in the rate of hydrolysis of Avicel cellulose with cellulases (256). Furthermore, plants with reductions in cellulose content and/or crystallinity have previously been shown to have an increase in saccharification (202, 257, 258). The *sac1* mutant did not show any alteration in glucose release by TFA hydrolysis compared to wild-type, suggesting that the overall amount of cellulose is reduced in this mutant. Therefore, the increase in saccharification is likely due to one of the other cell wall alterations in this mutant, i.e. lignin composition or ferulic acid content.

The majority of studies on cellulose synthesis mutants have reported a severe growth and development effect, including embryonic lethality (265), dwarfing (266), severe morphogenesis and fertility effects (34, 267), thin and incomplete cell walls (268), and reductions in stem strength (269, 270). The examples above are all for alterations in

primary cell wall cellulose synthesis but alterations in secondary cell wall cellulose synthesis has also been shown to have a strong effect on phenotype, causing weak stems, collapsed xylem, thinner cell walls and dwarfing (35, 271, 272, 348, 349). The *sac1* and *sac7* mutants, which had reductions in crystalline cellulose content, exhibited some of these growth defects. For example they both showed a reduction in plant height compared to wild-type (24 and 21% reductions respectively), the *sac7* mutant showed a reduction in shoot biomass, and the *sac1* mutant showed reduced seed production and the seeds took longer to germinate. However, these alterations seem to be more modest than reported for most cellulose mutants; the plants were not severely dwarfed, seeds had a similar germination percentage to wild-type, they did not show any reduction in stem strength or stiffness in comparison to wild-type and whole-plant morphology was not altered. Similar alterations in height and germination time, with no alterations in general morphology, have been reported in Brachypodium with reduced cellulose crystallinity in another study (350).

3.9.3.2 Ferulic acid

Another feature of the *sac1* cell walls, in addition to the reduction in crystalline cellulose content, was a significant reduction in the amount of ester-bound ferulic acid, a feature also observed for the *sac4* mutant. In cell walls of grasses, ferulic acid esters are abundant on arabinosyl side chains of GAX (103). These feruloyl esters may be oxidatively dimerised to form crosslinks between or within arabinoxylan chains and it has also been suggested that they may form the major link between hemicellulose and lignin in grass cell walls (103). It is therefore thought that ferulic esters could play a role in the recalcitrance of plant cell walls to digestion and a reduction in these residues has been shown to result in increased saccharification in a number of plant species (109, 300, 301). Consequently, the high saccharification of *sac1* and *sac4* could, in part, be due to a lower content of ferulic esters.

3.9.4 Mutants with alterations in lignin composition

Lignin composition was measured by thioacidolysis which only measures lignin monomers that are linked by β-O-4 bonds. Ratios and percentages given for lignin composition in this thesis therefore refer to lignin monomers that are β-O-4 linked. However, the monomers released by thioacidolysis have been shown to make up the majority of the total lignin content in Brachypodium (370) and therefore the ratios measured by thioacidolysis should quite closely reflect the ratios in total lignin. A major feature in a number of the *sac* mutants was an alteration in lignin composition. In most cases the change in composition consisted of a reduction in S:G ratio accompanied by an increase in % H monomers *(sac1*, *4*, *6*, *7)*, while *sac2* and *sac9* had just an increase in % H monomers. This is an intriguing result as a number of studies have shown that an *increase* in S:G ratio results in improved biomass digestibility/saccharification (234, 235, 351, 352). This is thought to be due to the number of methoxy groups on the lignin monomers (S>G>H), resulting in lignin rich in S units having a simpler structure that is more easily depolymerised due to the lower number of reactive sites (236). Furthermore, S units are more likely to form the more easily digestible β-*O*-4 bonds during lignin polymerisation than the more recalcitrant C-C bonds (241). However, there have also been a number of studies showing that both transgenic and natural accessions in a number of plant species with a decrease in S:G ratio but no reduction in lignin content have an increase in saccharification/digestibility (switchgrass: (353, 354), ryegrass: (355), maize: (356, 357), tall fescue: (358)). This could be due to the increased branching of G-rich lignin, resulting in the formation of lignin clusters which are less able to penetrate the polysaccharide matrix, while the less branched S-rich lignin forms more linear structures with increased penetration to the polysaccharides (233, 359). The plants that show a positive relationship between saccharification and S:G ratio are all dicots, whereas the plants that show a negative relationship are grasses. This may be due to the ferulic acid linkages that occur between lignin and hemicellulose in grasses but not in dicots (16). Therefore, the linear structures of S-rich lignin that are able to

penetrate the polysaccharide structure may only affect saccharification when the lignin has the ability to cross-link to the polysaccharide structure through ferulic acid linkages. Studies have also shown that an increase in H units can increase saccharification, which is thought to be due to the lack of methoxy groups on H units which help to stabilize the free radical and aid oxidation (218, 239). This could result in a lower number of reactive species and so smaller chain lengths and a less complex molecule (241).

Alternatively, the frequency of high saccharification mutants in this study with high G and H units could be due to the alkaline pretreatment used for the saccharification assay. Studies have shown that about 50% of native grass lignin can be solubilized in alkali, due to their high frequency in terminal units (360, 361). These terminal units are predominantly H and G monomers, whereas S monomers are essentially internal units. Accordingly, increasing the frequency of H and G monomers in grass lignin may make the lignin polymers more susceptible to alkaline pretreatment and, subsequently, more easily digested.

3.9.5 Mutants with no significant changes in major cell wall components

Among the 12 *sac* mutants were two, *sac8* and *sac10*, that showed no major differences in total amount of any of the cell wall components that we measured. Sarath et al. (362) reported that 50% of the variation they saw in ethanol yields from enzymatic hydrolysis of divergently bred switchgrass was due to changes in tissue and cell wall architecture rather than composition. Consonantly, the *sac10* mutant showed a much less uniform structure of the metaxylem vessels. This may suggest that the cell walls in this mutant are not constructed properly, making them easier to break down. Alternatively, although *sac8* did not show a change in the amount of cell wall components, this mutant did show an alteration in hemicellulose composition, with a significant reduction in arabinose and xylose content compared to wild-type. These reductions indicate a decrease in arabinoxylan content, the main matrix polysaccharide in grasses (16).

Arabinoxylan molecules can physically protect cellulose from enzymatic digestion due to their association with cellulose by extensive hydrogen bonding (331, 363, 364), and the cross-linking of hemicelluloses and lignin through ferulic acid on the arabinose side chains (103). It may, therefore, be possible that a reduction in arabinoxylan content in the *sac8* mutant allows easier enzyme access to the cellulose, thereby promoting saccharification.

3.9.6 Effect of pretreatment on saccharification

Pretreatment of wild-type and *sac* mutant stem material with either acid or alkali revealed that these two pretreatment solutions had different effects on saccharification. Alkaline pretreatment consistently resulted in much higher levels of sugar release by enzyme hydrolysis than acid pretreatment. This observation has been reported previously, both using the same high-throughput saccharification assay as used in this study (365) and using low-throughput methods of pretreatment and enzyme hydrolysis (366-368). This could be because alkaline pretreatment has been shown to primarily affect the lignin fraction within the lignocellulosic structure, while acid pretreatment primarily affects the hemicellulosic fraction (193, 198). Lignin has been shown to be an important factor in the recalcitrance of lignocellulosic biomass, both by reducing the accessibility of cellulases to cellulose and due to the adsorption of cellulases to lignin (334).

An additional reason for the higher saccharification seen after alkaline pretreatment could be because, in our saccharification assay, the pretreatment solution is removed before enzymatic hydrolysis and the method for quantification of sugar release does not distinguish between cellulosic sugars and hemicellulosic sugars. Therefore, because the acid pretreatment primarily solubilises hemicelluloses, these solubilised sugars would be removed after pretreatment and so would not be included in the quantification of saccharification. In addition, acid pretreatment has been shown to generate more

degradation products that are inhibitory to enzyme hydrolysis than alkaline treatment (369).

The high saccharification of the 12 *sac* mutants observed with a pretreatment of sodium hydroxide was not maintained when an acidic pretreatment with sulphuric acid was used. This is most likely because the acid and alkali pretreatments have such different effects on lignocellulose structure and composition, and the *sac* mutants were selected on the level of saccharification using an alkali pretreatment.

3.9.7 Effect of the mutations on growth and development

Importantly for the application of information gained in this screen for the improvement of plant material as a feedstock for bioethanol, none of the high saccharification mutants identified in this screen had any severe growth defects or reductions in stem strength. This is the opposite of many of the reverse genetic approaches taken to increase saccharification, which often report severe stunting or reductions in stem strength. For example, downregulation of HCT and C3H enzymes in alfalfa led to an increase in saccharification of 53 – 84% above wild-type level but also a reduction in biomass of up to 40% (218). Similarly, downregulation of CCR in poplar led to increases in saccharification of 14 - 78%, but reductions in biomass of around 50% (222). Mutations in two of the primary cell wall cellulose synthase genes in Arabidopsis (CESA1 and CESA3) led to a reduction in cellulose crystallinity, an increase in saccharification of 32 – 42% but also a reduction in height and biomass (257). Similarly, overexpression of a mutated Arabidopsis CESA3 gene in tobacco resulted in a 40 – 51% increase in saccharification efficiency but reduced height and biomass, as well as a decumbent growth form (264). Arabidopsis *irx* mutants with reduced xylan synthesis have increases in saccharification of up to 50% but exhibit a dwarf phenotype, severely reduced stem strength and collapsed xylem vessels (273, 274). The increase in saccharification observed in the *sac* mutants identified in this study are similar to the

increases observed in these reverse genetic studies, but mostly without negative impacts on growth or strength. The advantage of a forward genetic screen is that any severely dwarfed plants, or plants with severe developmental abnormalities, could be excluded from the initial screen. Our screen therefore automatically selected for plants with an increase in saccharification but where these mutations did not affect overall plant growth and development. However growth and development of the *sac* mutants was only investigated under greenhouse conditions. It is possible that the mutant lines may not perform as well as wild-type in field conditions, for example they may be more susceptible to pathogen attack or abiotic stress. It will be important to examine this in the future, especially if the causal genes are investigated in biofuel crops.

Chapter 4 – Role of hydroxycinnamoyl-CoA shikimate/ quinate transferase (HCT) in lignocellulose digestibility in Brachypodium

4.1 Introduction

Lignin is currently thought to be one of the major contributors to the recalcitrance of lignocellulosic biomass to enzymatic digestion. The role of lignin in the cell wall is to provide strength and rigidity to the cell and to act as a barrier to pests and pathogens (130, 131). This function is fulfilled by lignin being deposited into the wall after cell growth has terminated to form a protective layer within and around the polysaccharide structure. Lignin has a complex structure due to the monomers, or monolignols, being polymerised by free radical chemistry rather than enzymatic addition of repeat structures. This means that a number of different bonds can form between the monolignols, making it difficult for enzymes to break down (133). Therefore, lignin acts as an efficient barrier against the polysaccharide-degrading enzymes used for bioethanol production, preventing them from accessing the cell wall sugars. Furthermore, breakdown products of lignin produced during pretreatment for bioethanol production can inhibit enzyme activity (210) and cellulases can adsorb to lignin, reducing the availability of these enzymes for the breakdown of cellulose (212).

Lignin is formed from three main monolignols: *p*-hydroxyphenyl (H), guaiacyl (G) and sinapyl (S) units. They are synthesised from phenylalanine through the phenylpropanoid pathway via a number of different enzymes (Figure 23) (133). The three monolignols are incorporated into lignin in different amounts. In Brachypodium, S and G units make up the majority of lignin, with an S:G ratio of approximately 2.0, while the H units only make up approximately 4.0% of the lignin (317, 370, 371). There are two main linkages that occur between monolignols within the lignin polymer, β-*O*-4 and C-C. The β-*O*-4 linkages are the most abundant, making up about 80% of the linkages in angiosperms,

and are more easily cleaved than the C-C bonds (156). Abundance of the different linkages depends partly on the ratio of the different monolignols. For example lignin rich in S units has lower concentrations of C-C bonds because the C-5 position is not available for coupling due to the extra methoxy group (156).

Altering the activity or expression of many of the enzymes within the monolignol synthesis pathway has been shown to increase saccharification of the resulting plant material. This can be a result of either lowering the lignin content of the cell wall or altering lignin composition. For example, knockdown of C4H, 4CL, CCoAOMT, CCR, HCT, C3H and CAD have all been shown to reduce lignin content and increase saccharification (218, 219, 221-223, 239, 370). Alternatively, altering expression of F5H and COMT, which convert coniferaldehyde to sinapaldehyde and so divert synthesis of G units to S units (Figure 23), has been shown to cause a change in S:G ratio and level of saccharification, with no effect on lignin content (219, 226, 235). Furthermore, knockdown of C3H and HCT, which are required for the synthesis of S and G units but not H units (Figure 23), has been shown to cause an increase in the percentage of H monomers and an increase in saccharification (218, 236, 238-240). HCT is particularly interesting because knockdown of this enzyme was shown to result in the highest increase in saccharification among six mutants of the lignin biosynthetic pathway in alfalfa (218). However, application of this knowledge to improve saccharification of biofuel or cereal crops for bioethanol production is hampered by the severe dwarfing of plants in which HCT has been down-regulated (218, 238, 239), making them unattractive as a feedstock.

In order to utilise the information obtained from the high-throughput screen for saccharification, described in chapter 3, it is imperative to identify the causal mutations of the increased saccharification in the *sac* mutants. Without this information, the increases in saccharification cannot be transferred to the relevant crops for biofuel production. Traditional mapping is a long process and was not feasible in the timescale

Figure 23: Lignin monomer synthesis pathway. The conversion of *p*-coumaroyl CoA to caffeoyl CoA by HCT and C3H is highlighted by a red box. HCALDH: hydroxylcinnamaldehyde dehydrogenase. For other abbreviations see Figure 6.
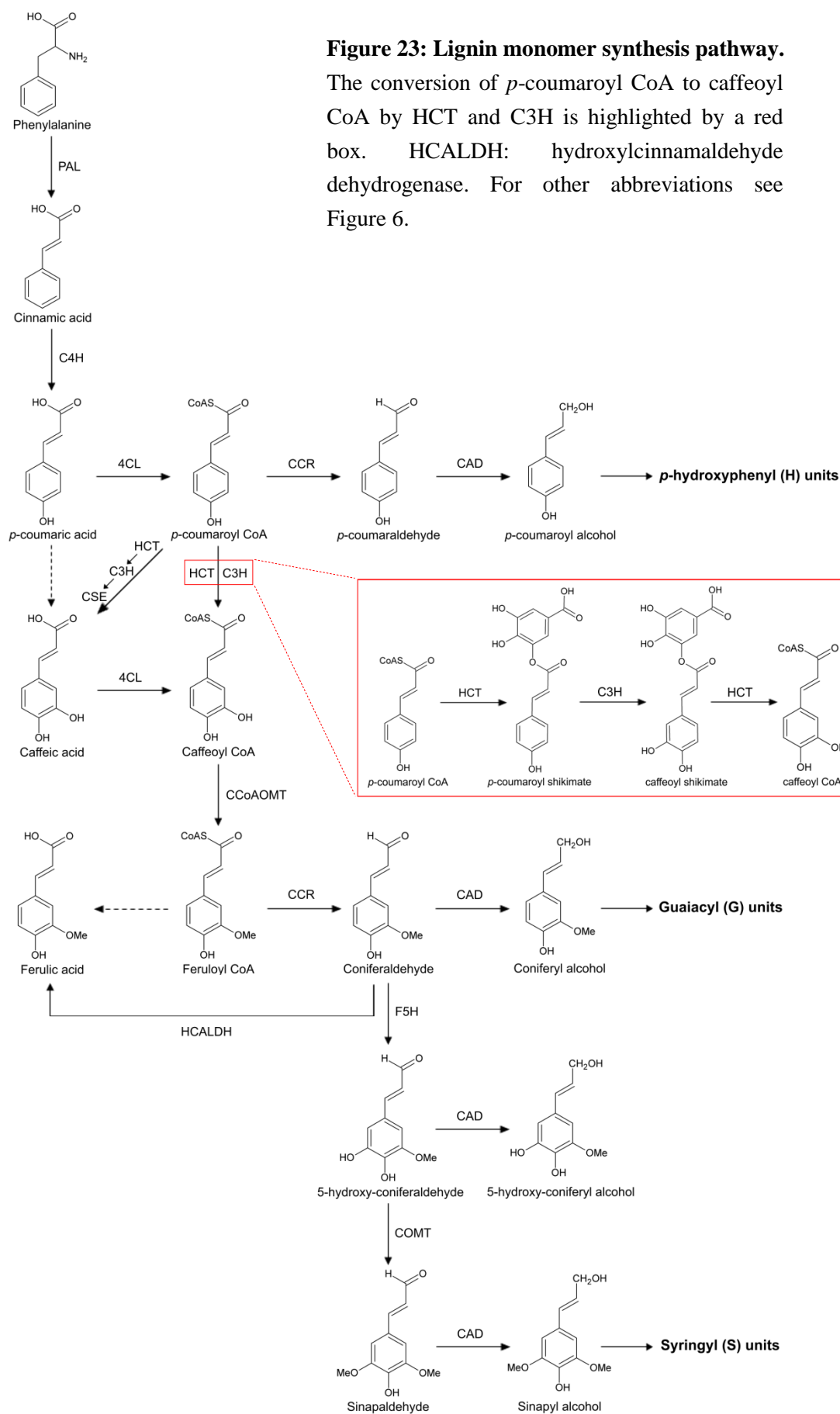
of this PhD, once the high saccharification mutants had been identified. Mapping was also hampered initially by difficulty in crossing Brachypodium. In order to obtain early indications of potential causal mutations whilst backcrossing was in progress for mapping, three of the *sac* mutants (*sac3*, *sac4* and *sac9*) were whole genome sequenced along with wild-type. The sequence information was analysed to identify all of the SNPs caused by the chemical mutagenesis which were in the homozygous state. The SNPs identified were then searched for any that occurred in genes known to be involved in cell wall synthesis or saccharification and then further refined to eliminate those that would cause no change in amino acid sequence. Retaining just the SNPs that would cause a change in amino acid was a pragmatic approach to allow any identified mutations to be more easily rationalised and validated as having an effect in the plant. However, mutations in non-coding regions could also be the causal mutation, for example by altering expression level or splicing. Finally, SIFT (<u>s</u>orts <u>i</u>ntolerant <u>f</u>rom <u>t</u>olerant) analysis was used to identify those SNPs most likely to impact on protein structure and function. The mutant lines *sac3*, *sac4* and *sac9* were chosen for this analysis due to their low lignin content and/or change in lignin composition. Lignin synthesis is the most well studied cell wall component and a large number of genes involved in its synthesis have been identified. This knowledge could help in screening the SNPs for potential mutated genes responsible for the phenotypes.

4.2 Whole-genome sequencing of the *sac* mutants to identify potential causal mutations

Genomic DNA from *sac3*, *sac4* and *sac9* was extracted from M_4 generation plants and whole-genome sequencing was performed by TGAC, Norwich using an Illumina HiSeq 2000 sequencer. A minimum of 4.7 Gbp of high quality paired-end sequencing data was achieved for each sample, giving a minimum coverage of 15.6x.

4.2.1 Identifying and probing homozygous SNPs

The whole-genome sequencing data of the wild-type and three *sac* mutants was aligned to the Brachypodium reference genome sequence (Bd21 accession) and homozygous SNPs identified. All SNPs that were found in both the wild-type and mutant datasets were then removed from the mutant datasets. This analysis was performed by Yi Li and Zhesi He, in the CNAP Bioinformatics group at the University of York. SNPs that did not resemble a mutation caused by sodium azide were then also removed. A Microsoft Excel macro was then created to search the datasets for SNPs in genes related to cell wall synthesis and saccharification by key-word searching of the annotation (Figure 24). This left a small number of SNPs (35-50) for each dataset. A computer program was then created that, given an un-spliced DNA sequence of a gene and the position of a SNP in that gene, will return the codon that has been changed and the position of that codon (Figure 25). This program enables one to see whether a SNP is present in a coding or non-coding region of a gene and, if it is in a coding region, the position of and change in amino acid in the protein sequence. SNPs were only retained if they caused a change in amino acid. Finally, SNPs were analysed using SIFT analysis to give an idea of how well these amino acid changes could be tolerated (372). SIFT analysis works on the basis that a highly conserved amino acid will be important for the structure or function of the protein. It aligns the sequence of the protein of interest with a large number of proteins with similar sequences (identified by BLAST analysis) to see how well conserved each amino acid is. Then, based on the properties of the different amino acids, predicts how well a change to each amino acid at each position will be tolerated. Substitutions with a SIFT score below 0.05 are predicted to not be tolerated.

```
Sub CopyRows()

# search terms, ie cell wall genes/key words, are in excel
sheet 'search terms' in row 'D' which is where search terms
related to ferulates are
For i = 2 To Sheets("search_terms").Cells(Rows.Count,
"D").End(xlUp).Row

# looks in sheet 'dataset', rows 'L:V' which is where raw data
from sequencing is and searches for search terms as above
Set vFound =
Sheets("dataset").Columns("L:V").Cells.Find(What:=Sheets("sear
ch_terms").Range("D" & i).Value, _
MatchCase:=False)

If Not vFound Is Nothing Then
vStart = vFound.Address
vMatch = True

# if find a search term, copy cells 'A' and 'AL' from that row
and paste them to  sheet 'ferulates'
Do
Sheets("dataset").Range("A" & vFound.Row & ":" & "AL" &
vFound.Row).Copy Destination:= _
Sheets("ferulates").Range("A" &
Sheets("ferulates").Cells(Rows.Count, "A").End(xlUp).Row + 1 _
& ":" & "AL" & Sheets("ferulates").Cells(Rows.Count,
"A").End(xlUp).Row + 1)

# keep on searching in 'dataset' sheet until reach the end
Set vFound =
Sheets("dataset").Columns("L:V").Cells.FindNext(vFound)
Loop Until vFound.Address = vStart
End If

Next i

If vMatch <> True Then
MsgBox ("NO MATCHES FOUND")
End If

End Sub
```

Figure 24: Microsoft Excel macro created to search datasets of homozygous SNPs acquired from whole-genome sequencing for SNPs that occurred in genes involved in cell wall synthesis. The macro will search the dataset, which is in one worksheet of the Excel workbook, for keywords that are in another worksheet within the same workbook. The keywords are separated into different cell wall components by column. If a keyword is found in the dataset, it will copy the row the keyword occurred in from the dataset and paste it into another worksheet within the same workbook. Text highlighted in yellow are terms that can be changed in order to search an alternative list of key words, which are in a different column of the 'keywords' worksheet, and/or add the results to a new worksheet within the workbook.

```python
#opens file with unspliced nucleotide sequence in it
sequence_file = open('Bradi2g50967.1.txt')

# Reads file and takes off new lines → makes it into a
continuous sequence, puts it into variable called 'seq'. Closes
the file
sequence_old = sequence_file.read()
seq = sequence_old.replace('\n','')
sequence_file.close()

# Makes list of numbers going from 0 to length of sequence above
numbers = range(0, len(seq))

# Creates new list 'letters' and adds each nucleotide in 'seq'
as an element in this list
letters = []
for i in seq:
    letters.append(i)

# Creates 2 lists 'codon' and 'useless'. If nucleotide is a
capital --> appends it to 'codon' list (exon) and also appends
the position in seq +1 from 'numbers' list. If nucleotide is a
lowercase --> appends it to 'useless' list (exon) and also
appends the position in seq +1 from 'numbers' list
codon = []
useless = []
import re
pattern = re.compile('[a-z]')
for i, j in zip(letters, numbers):
    if i.isupper():
        codon.append([i,j+1])
    else:
        useless.append([i, j+1])

# Makes a list of codons (ie 3 nucleotides at a time) followed
by the position of these 3 nucleotides in the original sequence
for 'exons'
letters2 = []
numbers2 = []
for x in codon:
    letters2.append(x[0])
    numbers2.append(x[1])

exon_list = []
for i, j in zip(range(0, len(letters2), 3), range(0,
len(numbers2), 3)):
    codon = letters2[i:i+3]
    position = numbers2[j:j+3]
    exon_list.append([codon, position])

# Writes to file
for x in exon_list:
    output_file = open('Bradi2g50967.1_exon.txt', 'at')
    output_file.write('{0}'.format(x[0]) + '\t' +
'{0}'.format(x[1]) + '\n')
```

```
        output_file.close()

letters3 = []
numbers3 = []
for x in useless:
    letters3.append(x[0])
    numbers3.append(x[1])

#does the same thing for 'introns'
intron_list = []
for i, j in zip(range(0, len(letters3), 3), range(0,
len(numbers3), 3)):
    codon = letters3[i:i+3]
    position = numbers3[j:j+3]
    intron_list.append([codon, position])

for x in intron_list:
    output_file = open('Bradi2g50967.1_intron.txt', 'at')
    output_file.write('{0}'.format(x[0]) + '\t' +
'{0}'.format(x[1]) + '\n')
    output_file.close()
```

Figure 25: Python script to identify whether an amino acid has been changed as a result of a SNP and, if it has, the position of that amino acid within the encoded protein. The program takes an un-spliced DNA sequence of a gene, numbers the nucleotides from one to the length of the sequence and creates two Excel files: 'exons' and 'introns'. These files contain a list of codons in the order they appear in the protein sequence in the first column and the position of these codons in the DNA sequence in the second column. The 'exons' file contains all nucleotides that had a capital letter in the original sequence (i.e. the coding sequence) and the row number refers to the position of that codon in the protein. The 'introns' file contains all nucleotides that had a lower case letter in the original sequence (i.e. introns and 3' and 5' UTRs).

4.2.2 Potential causal mutations

At the end of this selection process, there were two possible candidates among the three *sac* mutants for the SNPs causing the increased saccharification and changes in cell wall composition. For the *sac3* mutant no SNPs were identified in known cell wall/ saccharification related genes for which the change in amino acid was predicted to not be tolerated by SIFT analysis.

4.2.2.1 Mutation in a cellulose synthase (CESA) gene

For the *sac9* mutant a SNP was identified in Bradi1g29060, encoding a CESA enzyme (CESA6), which would cause a change in amino acid from alanine to threonine. The alanine is highly conserved across similar proteins and this change in amino acid was predicted to not be tolerated (SIFT score of 0.01, score of ≤0.05 is predicted to not be tolerated). The mutation is two amino acids from a highly conserved aspartic acid residue which is thought to be involved in catalytic activity or the attachment of substrate to enzyme (Figure 26) (373). However, the alteration in cell wall composition of *sac9* does not particularly support a reduction in functionality of a CESA gene, with just a small, non-significant increase in the cellulose content of this mutant compared to wild-type. The only significant alterations in cell wall composition observed in the *sac9* mutant in this study was a reduction in lignin content and an alteration in lignin composition (Table 5 and Table 6). Therefore, this SNP was not pursued any further.
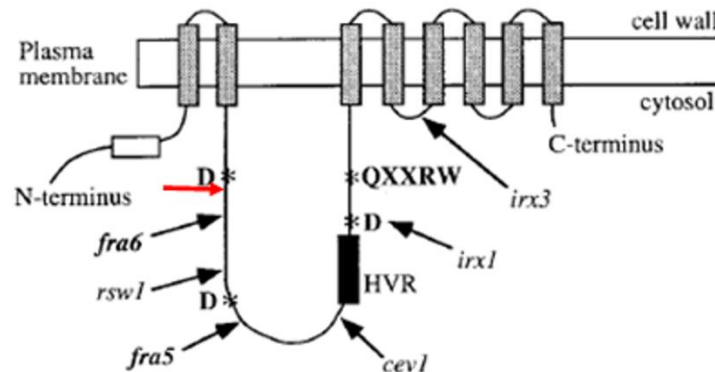
Figure 26: Diagram of the predicted topology of plant CESA proteins. Conserved aspartic acid residues and QXXRW motif within the protein are indicated with an asterisk, the position of the SNP identified in this study is shown with a red arrow and the positions of mutations in known CESA mutants are indicated with black arrows. HVR: hypervariable region. The original diagram was taken from Zhong et al. (374).

4.2.2.2 Mutation in a lignin monomer biosynthetic gene, HCT

In the *sac4* mutant a SNP was identified in the gene Bradi5g14720. BLAST analysis of the protein sequence showed that it had sequence similarity to hydroxycinnamoyl-CoA shikimate/quinate transferases (HCT) in other plant species (92% identity to rice HCT1, E-value = 0.0; 65% identity to Arabidopsis HCT, E-value = 0.0). HCT is a member of the monolignol biosynthetic pathway. The mutation identified in the Bradi5g14720 gene would cause an amino acid change from proline to leucine and attained a low SIFT score (0.00), predicting that this substitution would not be tolerated in the structure/function of the encoded protein. As prolines are known to introduce sharp turns in polypeptide chains, this substitution is likely to alter the structure of this protein. The alterations in cell wall composition of the *sac4* mutant are similar to those reported for *hct* mutants in other plants (218, 238, 239, 375), with a large reduction in lignin content (-35%), a large increase in the proportion of H units in the lignin (60%) and a reduction in ferulic acid content (-27%) compared to wild-type (Table 5 and Table 6). These alterations in cell wall composition fit with a reduction in HCT activity as the enzyme acts early in the lignin monomer synthesis pathway, along with C3H, to convert *p*-coumaroyl-CoA to caffeoyl CoA (Figure 23). This conversion is required for the production of G and S lignin monomers but not for H monomers, and is required for the production of feruloyl-CoA and coniferaldehyde from which ferulic acid is thought to be biosynthesised.

4.3 Investigation of the HCT mutation in the *sac4* mutant line

As the SNP identified in the Bradi5g14720 gene of the *sac4* mutant appeared to be a good candidate for the causal mutation, due to the similarity between the altered cell wall composition of the *sac4* mutant and *hct* mutants reported in other plants, we decided to investigate this mutation further.

4.3.1 Confirming the presence of the SNP in *sac4* plants

In order to confirm that the SNP identified in the Bradi5g14720 gene by whole-genome sequencing was present in *sac4* plants but not in wild-type plants, the Bradi5g14720 gene was cloned and sequenced from four *sac4* and wild-type plants (five plasmids sequenced for each plant). This confirmed that the *sac4* plants were homozygous for the alternative allele at the SNP position, while the wild-type plants were homozygous for the reference allele.

4.3.2 HCT activity in wild-type and *sac4* plants

The specific activity of the HCT enzyme is to take the substrates *p*-coumaroyl-CoA and shikimic acid and to produce *p*-coumaroyl shikimate by exchanging the CoA for the shikimic acid (Figure 27). HCT activity in wild-type and *sac4* mutant plants was assessed by measuring the formation of the product *p*-coumaroyl shikimate by ultra performance liquid chromatography mass spectrometry (UPLC-MS) when crude protein extracted from plant stem material was incubated with the substrates *p*-coumaroyl-CoA and shikimic acid, following the method of Vanholme et al. (375). Initially, a time course was carried out to see how enzyme activity changed over time (Figure 28). This showed that HCT activity was reduced in the mutant stem compared to wild-type across the whole time course. The activity plateaued at approximately 60 minutes for both extracts; however the reaction with the *sac4* extract plateaued at a level more than 50% lower than the reaction with the wild-type extract. A reaction with boiled protein extract and a reaction with just the protein extract and no substrates were used as negative controls, neither of which showed any activity (Figure 28).

Based on the time course, a time point of 30 minutes was chosen to investigate the difference between HCT activity in wild-type and *sac4* stem material further, using multiple biological replicates. This time point was before the activity of either extract plateaued. Activity assays were carried out in the same way but using protein extracted

from five separate wild-type and *sac4* plants in separate reactions. This revealed that the reduction in activity of the *sac4* extract compared to wild-type seen in the time course was maintained (Figure 29). On average, activity in the mutant plant was reduced by 27.88% compared to wild-type and this difference was significant (t-test, p = 0.022).
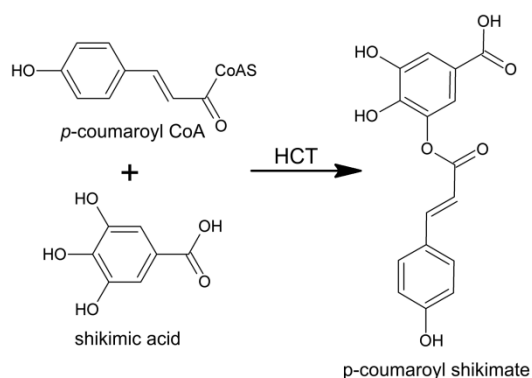
Figure 27: Action of the HCT enzyme within the lignin monomer synthesis pathway. HCT: hydroxycinnamoyl-CoA shikimate/quinate transferase.
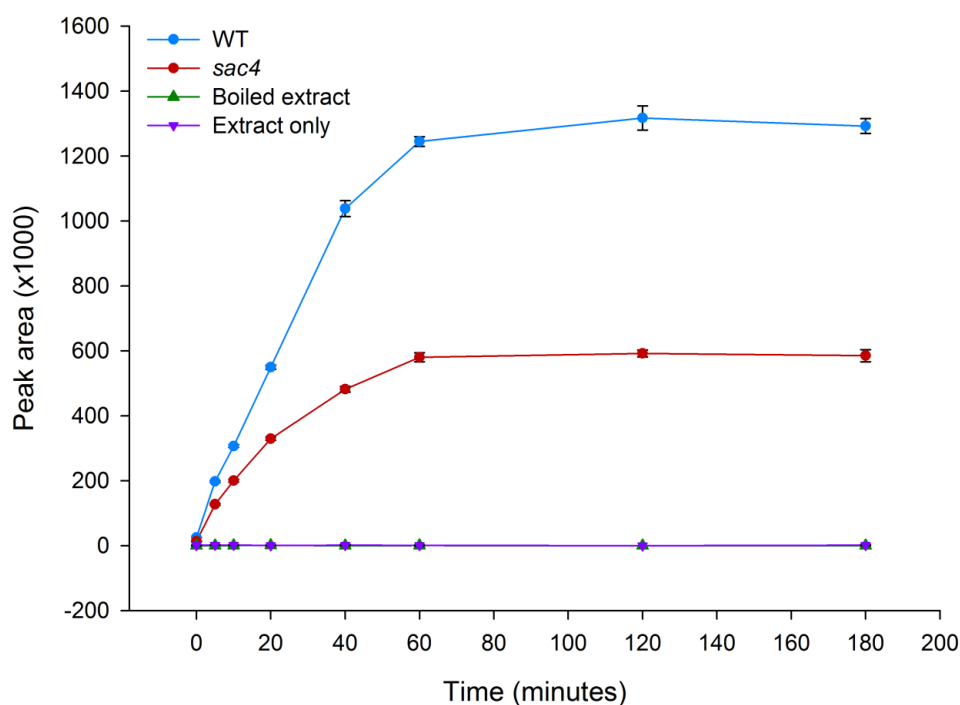
Figure 28: Time course showing HCT activity of crude protein extracted from wild-type and *sac4* plants. Graph shows accumulation of *p*-coumaroyl shikimate when the protein extract was incubated with substrates shikimic acid and *p*-coumaroyl-CoA, measured by UPLC-MS. A boiled enzyme reaction and a reaction with just the protein extract and no substrates were used as negative controls. Data represents mean ± SD, n = 3 technical replicates.
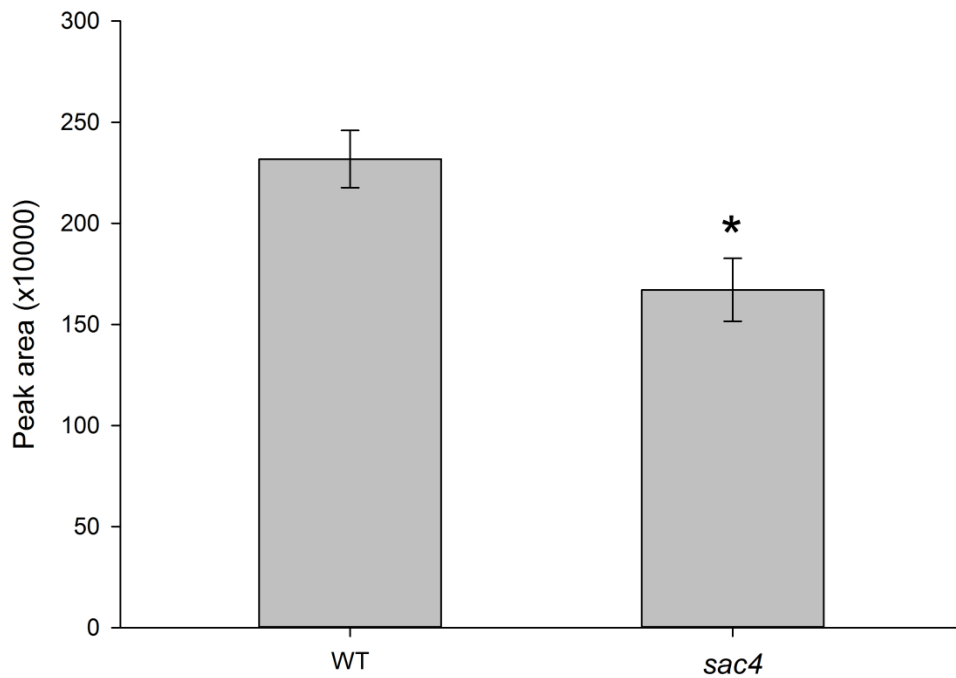
Figure 29: HCT activity of crude protein extracted from wild-type and *sac4* plants. Graph shows accumulation of *p*-coumaroyl shikimate when the protein extract was incubated with substrates shikimic acid and *p*-coumaroyl-CoA for 30 min, measured by UPLC-MS. Data represents mean ± SE, n = 5.

4.3.3 HCT activity of recombinant expressed wild-type and mutant versions of the Bradi5g14720 protein

In order to investigate whether the reduction in HCT activity seen in the *sac4* stem material could result from the SNP identified in the Bradi5g14720 gene, the wild-type and mutant versions of the protein were expressed in *E. coli* and HCT activity assays carried out using the resulting purified protein.

4.3.3.1 Expression and purification of the proteins in *E. coli*

E. coli codon-optimised versions of the mutant and wild-type Bradi5g14720 gene were synthesised by GeneArt® (Life technologies) and cloned into a pGEX vector in order to produce the protein as a fusion with glutathione-S-transferase (GST) for ease of purification. The pGEX vector with synthesised gene inserted was then transformed into *E. coli* and expression of the Bradi5g14720 protein was optimised by growing the

131

cells under a number of different conditions. Expression was tested at three different temperatures (20, 30 and 37°C), in the presence of three different IPTG concentrations (0.1, 0.5 and 1 mM), and for two different incubation periods (3 and 14 hours). Protein produced under each condition was examined by SDS-PAGE and Western blot using an anti-GST antibody. Figure 30 shows an example of the levels of expression of soluble protein when cells were cultured with 0.1 mM IPTG at the different temperatures and incubation periods (SDS-PAGE and Western blots of incubation with 0.5 and 1 mM IPTG are not shown). It was found that level of expression was highest when cultures were incubated for 14 hours at 20°C, with 0.1 mM IPTG. Therefore, these conditions were used for all subsequent expression.

The wild-type and mutant proteins were expressed and purified by affinity chromatography using glutathione-linked sepharose media for activity assays. Figure 31 shows analysis of the soluble purified protein by SDS-PAGE and Western blot, with the expected product at ~70 kDa. Some products with lower molecular weights were also observed. Protein bands were excised and analysed by peptide mass spectrometry following trypsinolysis, which confirmed that the band at ~70 kDa contained the GST tag and expressed protein. The band between the 55 and 70 kDa markers was shown to be an *E. coli* protein, molecular chaperone GroEL, and the band at ~25 kDa was the GST tag alone.
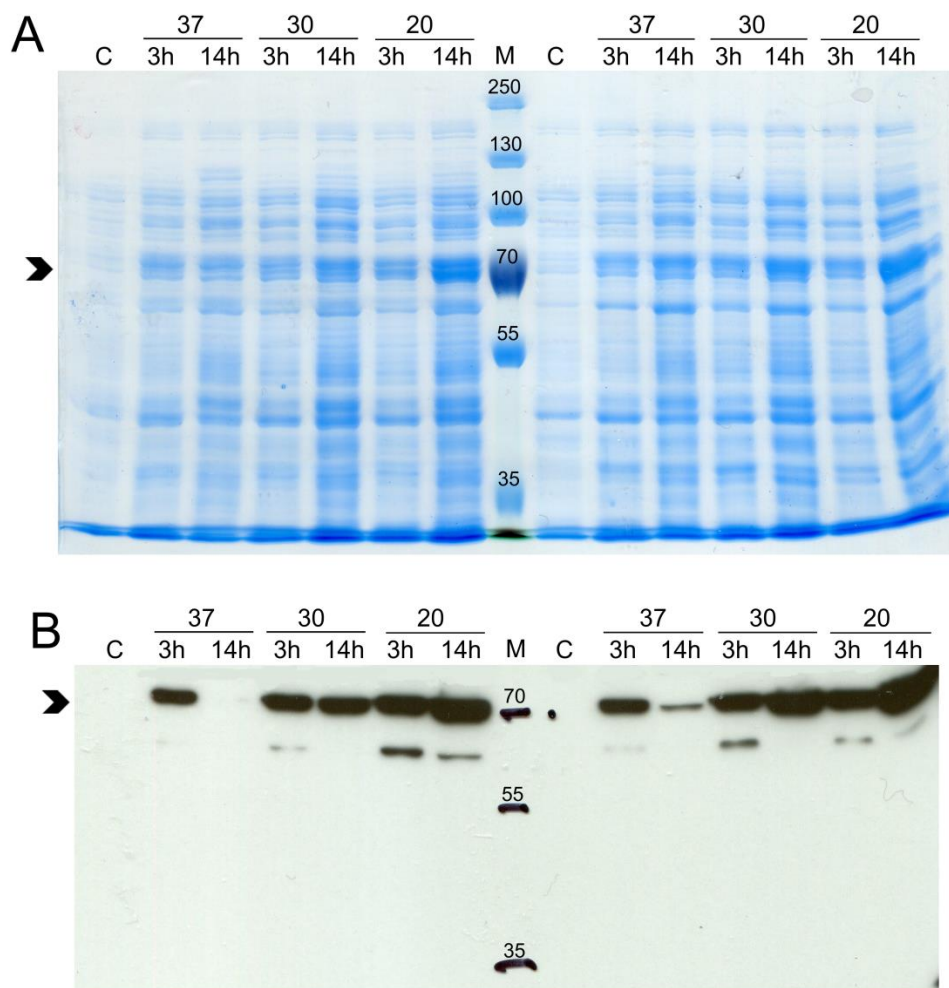
Figure 30: Optimisation of expression of recombinant expressed wild-type and mutant Bradi5g14720 protein in *E. coli*. Recombinant expressed wild-type (left side of the marker) and mutant (right side of the marker) versions of the Bradi5g14720 protein were optimised for expression. *E. coli* cultures were induced to express the protein with 0.1 mM IPTG and incubated at 37, 30 or 20°C for three or 14 hours before running the soluble fraction on an SDS-PAGE gel (A) and then carrying out a Western blot using an anti-GST antibody (B). Molecular weight of Bradi5g14720: 48 kDa, and molecular weight of Bradi5g14720 plus GST tag: 74 kDa. The arrow heads show the expected molecular weight of Bradi5g14720 plus GST tag. C: control, M: protein marker. Molecular weights (kDa) of protein marker are shown.
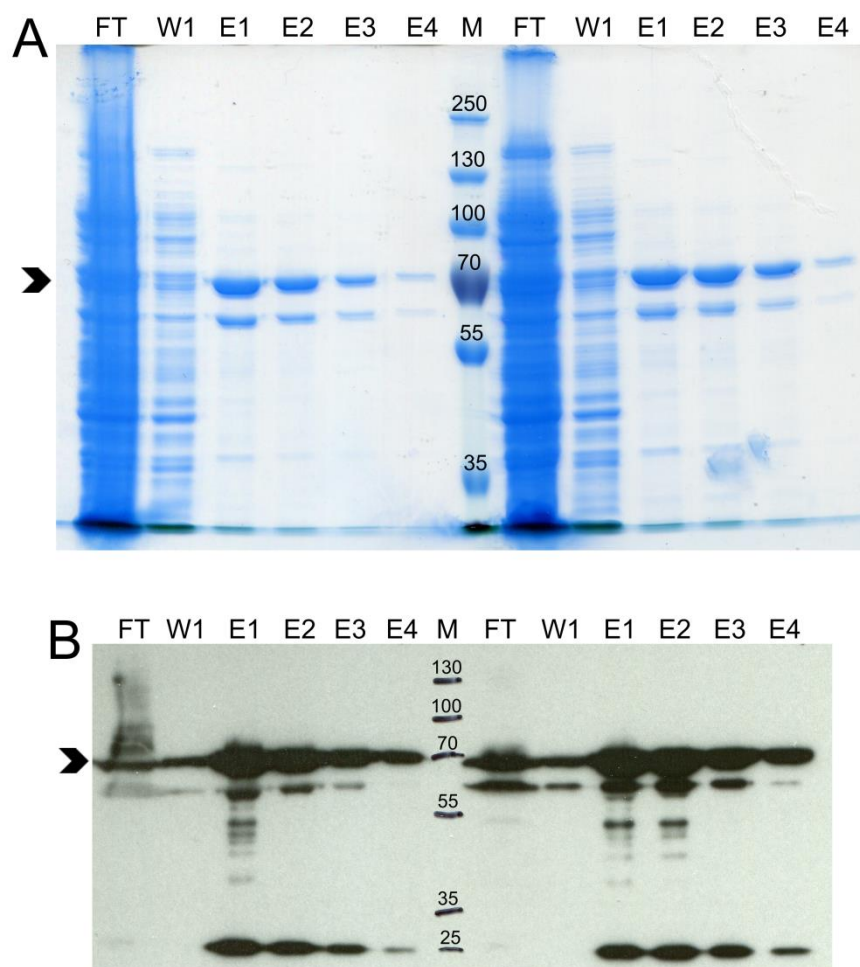
Figure 31: Purification of recombinant expressed wild-type and mutant Bradi5g14720 protein. Recombinant expressed wild-type (left side of the marker) and mutant (right side of the marker) versions of the Bradi5g14720 protein were purified using Glutathione Sepharose and analysed by SDS-PAGE (A) and Western blot using an anti-GST antibody (B). FT: flow-through, W: wash, E: elution, M: marker. Molecular weights (kDa) of protein marker are shown. The arrow heads show the expected molecular weight of Bradi5g14720 plus GST tag (74 kDa).

4.3.3.2 HCT activity assays

HCT activity assays were carried out by measuring formation of the product *p*-coumaroyl shikimate by UPLC-MS when the recombinant protein was incubated with the substrates *p*-coumaroyl-CoA and shikimic acid, in the same way as with the plant extracts. This revealed that the wild-type recombinant protein had 85-fold higher activity than the mutant protein, and the mutant recombinant protein showed just very

slightly higher activity than observed with the empty vector control (Figure 32). This suggests that the mutation identified in the Bradi5g14720 gene almost completely knocks out the activity of the enzyme.
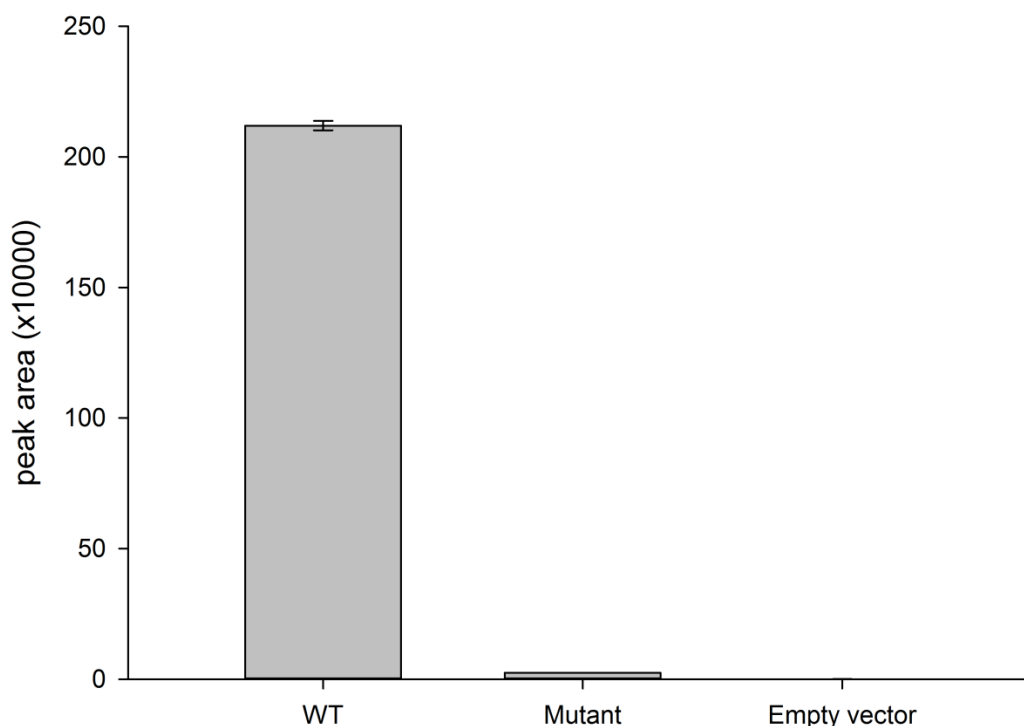
Figure 32: HCT activity of recombinant expressed wild-type and mutant versions of the Bradi5g14720 protein. Graph shows accumulation of *p*-coumaroyl shikimate when the recombinant protein was incubated with substrates shikimic acid and *p*-coumaroyl-CoA for 30 min, measured by UPLC-MS. Protein expressed and purified from an empty pGEX-4T3 vector was used as a negative control. Data represents mean ± SD, n = 3.

4.4 Characterisation of the HCT gene family in Brachypodium

The relatively mild 30% reduction in HCT activity of *sac4* plant material despite the almost complete knockout in HCT activity of the recombinant expressed mutant Bradi5g14720 protein suggests there are additional HCT enzymes in Brachypodium which act in a redundant manner. In order to further investigate the role of Bradi5g14720 and other putative HCT proteins in Brachypodium, phylogenetic, gene expression and enzyme kinetic analyses were carried out.

4.4.1 Phylogenetic analysis of putative HCT enzymes in Brachypodium

BLAST analysis using the Arabidopsis HCT protein against the Brachypodium protein database revealed two other putative HCT proteins in Brachypodium (Bradi3g48530 (identity = 62%, E-value = 0.0) and Bradi4g36830 (identity = 32%, E-value = $2e^{-52}$)). Multiple alignment, followed by construction of a neighbour joining phylogenetic tree was carried out with the three putative Brachypodium HCT proteins, and HCT proteins in Arabidopsis, *Nicotiana tabacum*, *Coffea canephora*, *Populus trichocarpa*, *Oryza sativa*, *Zea mays, Panicum virgatum, Hordeum vulgare* and *Sorghum bicolor* (Figure 33). The HCT proteins from Arabidopsis, *N. tabacum*, *C. canephora*, *P. trichocarpa* and *O. sativa* have previously been published (238, 375-378), whereas the *Z. mays, P. virgatum, H. vulgare* and *S. bicolor* HCT proteins were selected by sequence similarity to the Arabidopsis HCT protein. Only the Arabidopsis and *N. tabacum* HCT enzymes and *O. sativa* HCT4 enzyme have been characterised in terms of their biochemical function (238, 377). The phylogenetic analysis revealed that the Bradi5g14720 protein, in which the SNP was identified in this study, groups with the *O. sativa* HCT1 protein, as well as a putative HCT protein from *H. vulgare*. Neither of these proteins have been characterised in terms of their biochemical function. The other Brachypodium putative HCTs also clustered with other monocot HCTs. Bradi3g48530 grouped closely to the HCT2 proteins in *O. sativa*, *Z. mays* and *P. virgatum*, as well as putative HCT proteins from *H. vulgare* and *S. bicolor*. None of these proteins have been functionally characterised. Bradi4g36830 grouped with *O. sativa* HCT3 and HCT4 and, interestingly, these three proteins grouped more closely to dicot HCT proteins than other monocot HCT proteins. *O. sativa* HCT4, expressed in *E.coli*, has been shown to have activity with shikimic and quinic acid as acyl group acceptors and *p*-coumaroyl-CoA, caffeoyl-CoA and feruloyl-CoA as acyl group donors (377), as observed with recombinant HCT enzymes from *N. tabacum* and *P. trichocarpa* (379, 380). The phylogenetic analysis indicated there to be three homologous HCT clades in *O. sativa* and Brachypodium (Figure 33). Therefore, the three Brachypodium genes,

Bradi5g14720, Bradi3g48530 and Bradi4g36830, will hereafter be referred to as BdHCT1, 2 and 3, respectively, after the rice HCT proteins that they cluster with. This is indicated on the phylogenetic tree (Figure 33).
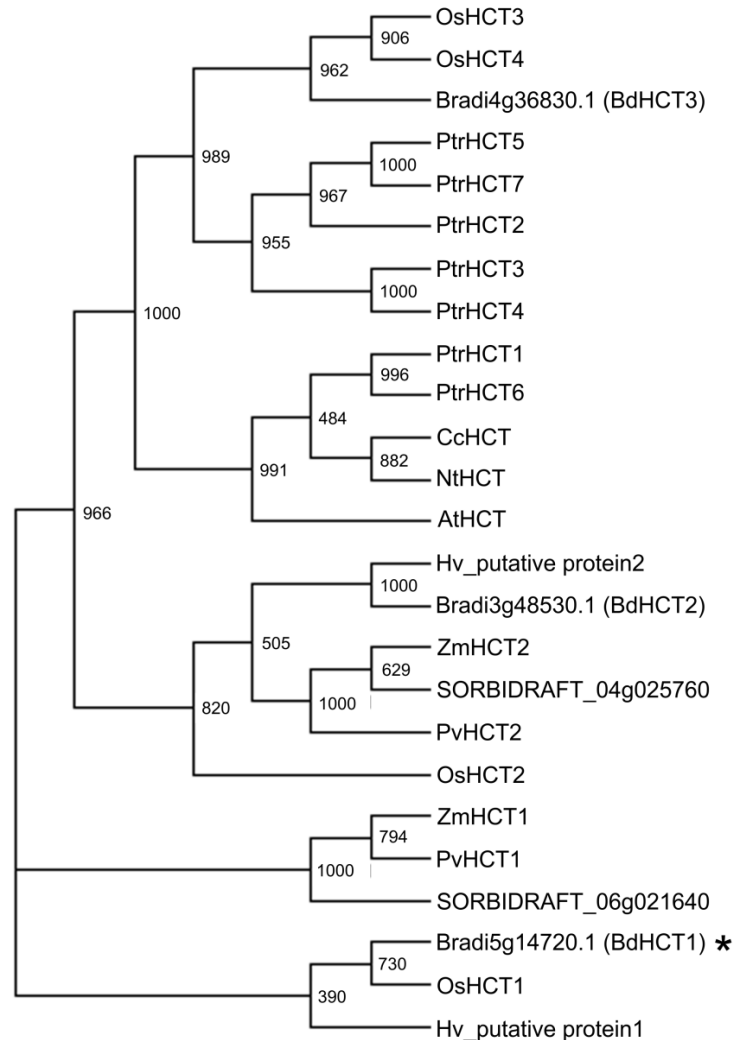
Figure 33: Phylogenetic analysis of Bradi5g14720 and two other putative HCT proteins in Brachypodium, with HCT proteins in Arabidopsis, *Nicotiana tabacum, Coffea canephora***,** *Populus trichocarpa***,** *Oryza sativa***,** *Zea mays***,** *Panicum virgatum***,** *Hordeum vulgare* **and** *Sorghum bicolor***.** The proteins are: AtHCT (At5g48930) from Arabidopsis; NtHCT (Q8GSM7) from *N. tabacum*; CcHCT (ABO47805.1) from *C. canephora*; OsHCT1 (115459249), OsHCT2 (115447256), OsHCT3 (115466797) and OsHCT4 (115466807) from *O. sativa*; ZmHCT1 (LOC100279386) and ZmHCT2 (LOC100194386) from *Z. mays*; PtrHCT1 (jgi|554899), PtrHCT2 (jgi|835948), PtrHCT3 (jgi|825948), PtrHCT4 (jgi|578723), PtrHCT5 (jgi|586045), PtrHCT6 (jgi|587193) and PtrHCT7 (jgi|784746) from *P. trichocarpa*; PvHCT1 (BAO20883.1) and PvHCT2 (AGM90558.1) from *P. virgatum*; Hv_predicted protein1 (BAJ95750.1) and Hv_predicted protein2 (BAK06288.1) from *H. vulgare*; and SORBIDRAFT_06g021640 and SORBIDRAFT_04g025760 from *S. bicolor*. The Brachypodium protein in which the SNP was identified in this study is indicated with an asterisk.

4.4.2 Gene expression of the Brachypodium HCT genes in wild-type and *sac4* plants

Real-time PCR (RT-PCR) was used to examine the level of expression of the BdHCT1 gene in wild-type plants in a variety of tissues (leaf, nodes, spikelet, basal internodes, middle internodes and top internode). In order to investigate the relative expression levels of this gene in the different tissues, expression in each tissue was calculated relative to expression in leaf tissue (Figure 34A). This showed that level of expression was relatively uniform across the different tissues, with a maximum difference of 2.8 fold in the level of expression (Figure 34A). Expression was slightly higher in the nodes and internodes than the spikelet and leaves.

Gene expression of the two other putative HCT genes identified in Brachypodium (BdHCT2 and BdHCT3) was also investigated in wild-type plants. Comparison of the expression of the BdHCT2 gene between different tissues (Figure 34B) showed that level of expression was relatively uniform across the different tissues, with a maximum difference of 1.8 fold in the level of expression. For the BdHCT3 gene, expression varied more widely between the tissues, with a difference of 11-fold in the level of expression between some tissues. Expression was highest in the spikelet and nodes, at an intermediate level in the internodes and lowest in the leaf (Figure 34C).

Gene expression of the three HCT genes was then investigated in the internodes of *sac4* plants compared to wild-type plants. Internodes were used because lignin biosynthesis genes should be highly expressed in this material, where the majority of secondary cell wall thickening and lignification occurs. Furthermore, this is the material on which HCT activity assays were carried out. The relative expression of the three HCT genes in mutant plants compared to wild-type is presented in Figure 35 and shows that level of expression in the mutant plants was not greatly altered compared to wild-type for any of the three genes; any difference observed was not significant. The level of expression in the internodes was much lower for BdHCT3 compared to BdHCT1 and BdHCT2.

138
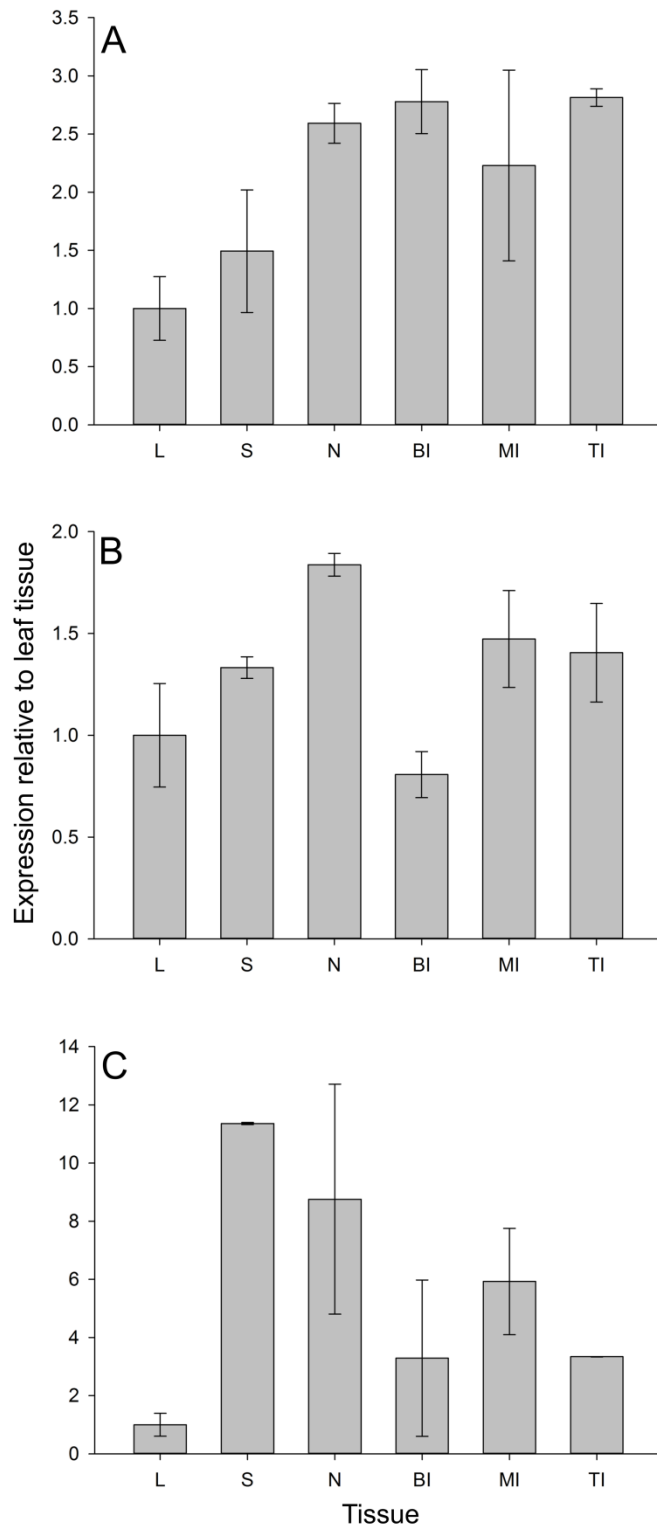
Figure 34: Gene expression of putative HCT genes in Brachypodium wild-type plants.
Transcript levels of A: BdHCT1, in which the SNP was identified; B: BdHCT2; and C: BdHCT3 were quantified in various tissues by RT-PCR. Transcript levels were normalised against the house-keeping gene, Ubiquitin 10, and expression calculated relative to that in leaf tissue for each gene. Plants were harvested at 8 weeks old. L: leaf, S: spikelet, N: node, BI: basal internode, MI: middle internode, TI: top internode. Data represents mean ± SD, n = 3.
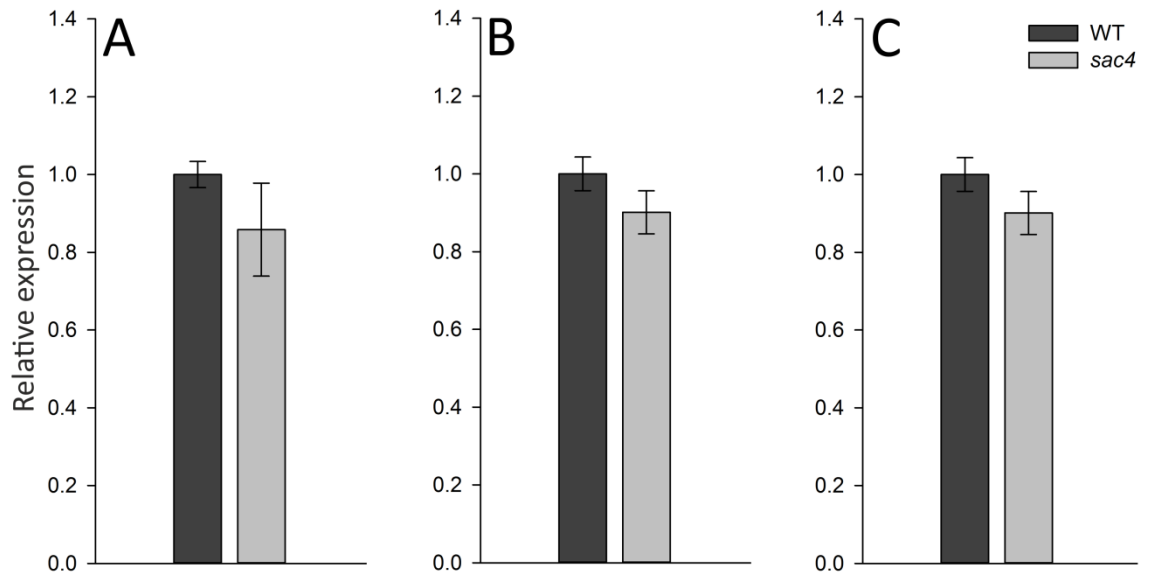
Figure 35: Gene expression of putative HCT genes in wild-type and *sac4* internodes. Transcript levels of A: BdHCT1, in which the SNP was identified; B: BdHCT2; and C: BdHCT3 were quantified in wild-type and *sac4* internode tissue by RT-PCR. Transcript levels were normalised against the house-keeping gene, Ubiquitin 10, and expression in the mutant plants calculated relative to that in wild-type plants for each gene. Plants were harvested at 8 weeks old. Data represents mean ± SD, n = 3.

4.4.3 Characterisation of enzyme activity of the HCT family in Brachypodium

In order to further understand the role of the three putative HCT enzymes in Brachypodium, the three proteins were expressed in *E. coli* and their kinetic properties with a variety of substrates examined.

4.4.3.1 Expression and purification of the HCT proteins

The expression and purification of the BdHCT1 protein has already been described above (see section 4.3.3.1). Expression of BdHCT2 and BdHCT3 was tested under different conditions to optimise for the greatest level of expression. As with BdHCT1, the highest level of expression was observed when cultures were incubated for 14 hours at 20°C, with 0.1 mM IPTG. Therefore, all three proteins were expressed using these conditions for the kinetic analyses. The BdHCT2 and BdHCT3 proteins were expressed and purified as described for BdHCT1. Figure 36 shows analysis of the soluble purified

proteins by SDS-PAGE and Western blot, with the expected product at ~70 kDa. Some breakdown products at lower molecular weights were also observed. Despite the high signal of the bands at approximately 50 kDa and 20-25 kDa on the Western blot, the SDS-PAGE shows that the proteins in these bands only made up a small proportion of the total soluble protein. Protein bands were excised and analysed by peptide mass spectrometry following trypsinolysis, which confirmed that the bands at ~70 kDa contained the GST tag and the expressed proteins. The band between the 55 and 70 kDa markers was shown to be an *E. coli* protein, molecular chaperone GroEL, and the band at ~25 kDa was the GST tag alone.
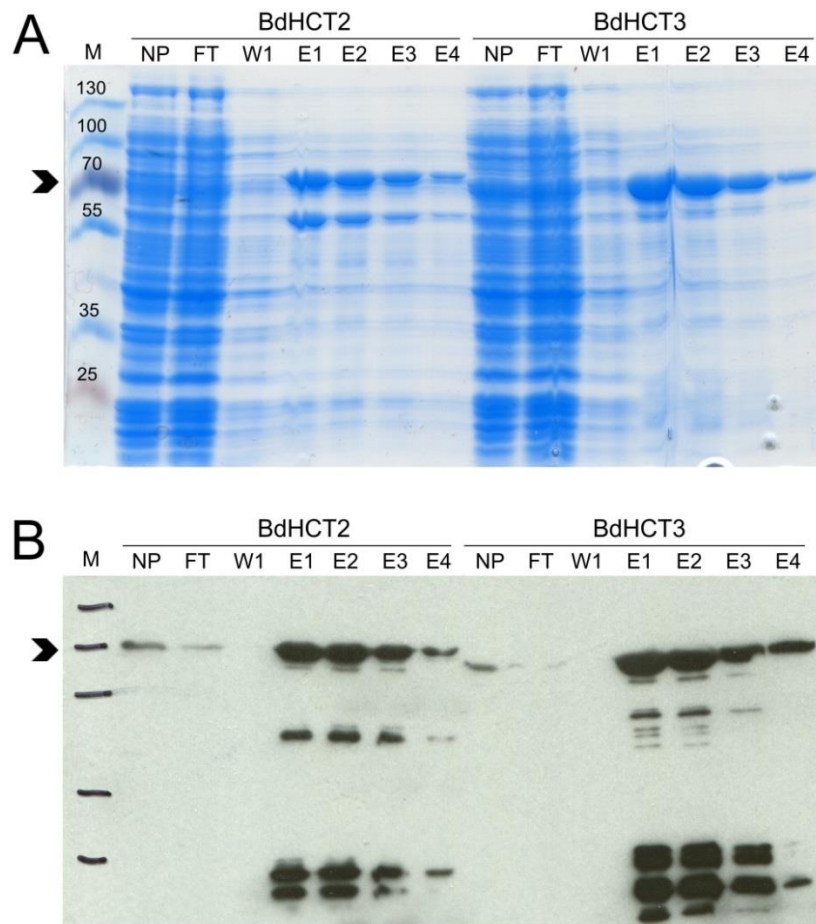
Figure 36: Purification of recombinant expressed BdHCT2 and BdHCT3 protein. The BdHCT2 and BdHCT3 proteins were expressed in *E. coli*, purified using Glutathione Sepharose and analysed by SDS-PAGE (A) and Western blot using an anti-GST antibody (B). NP: non-purified; FT: flow-through, W: wash, E: elution, M: marker. Molecular weights (kDa) of protein marker are shown. The arrow heads show the expected molecular weight of Bradi5g14720 plus GST tag (74 kDa).

4.4.3.2 Enzyme kinetics

Recombinant HCT enzymes from *N. tabacum*, *P. trichocarpa* and *O. sativa* have been shown to have activity with a number of substrates, namely: *p*-coumaroyl-CoA, caffeoyl-CoA and feruloyl-CoA as acyl group donors and shikimic acid and quinic acid as acyl group acceptors (377, 379, 380). In addition, recombinant *O. sativa* HCT4, which groups with BdHCT3 in the phylogenetic analysis, has activity with glycerol as an acyl acceptor (377). Therefore, the activity of the three putative HCT enzymes in Brachypodium was tested to see if the different enzymes had different activities and substrate preferences. Activity with *p*-coumaroyl-CoA, caffeoyl-CoA and feruloyl-CoA as acyl donors and with shikimic acid, quinic acid and glycerol as acyl acceptors was tested and the results are presented in Table 8. BdHCT1 and BdHCT2 both showed activity with all three acyl group donors when shikimic acid was the acceptor, but only with *p*-coumaroyl-CoA when quinic acid was the acceptor. BdHCT3 only showed activity when *p*-coumaroyl-CoA and shikimic acid were used as substrates. None of the enzymes had any activity when glycerol was used as the acceptor. The chromatograms and MS spectra from UPLC-MS of the products formed with the various substrates are shown in Figure 37: *p*-coumaroyl-shikimate (A and B), caffeoyl shikimate (C and D), feruloyl shikimate (E and F), and *p*-coumaroyl quinate (G and H).

The kinetic analyses showed that, out of the three enzymes, BdHCT3 appeared to have the highest efficiency when *p*-coumaroyl-CoA and shikimic acid were used as substrates, although Vmax was quite low for this enzyme (Table 8). For BdHCT1 and BdHCT2, affinity and efficiency was higher with shikimic acid as the acceptor compared to quinic acid. For BdHCT1, when shikimic acid was used as acyl acceptor, efficiency was similar with all three acyl donors. Efficiency was slightly higher with feruloyl-CoA, but Vmax was quite low when this donor was used (Table 8). For BdHCT2, when shikimic acid was used as acyl acceptor, caffeoyl-CoA was the preferred substrate, followed by feruloyl-CoA and then *p*-coumaroyl-CoA (Table 8).

142
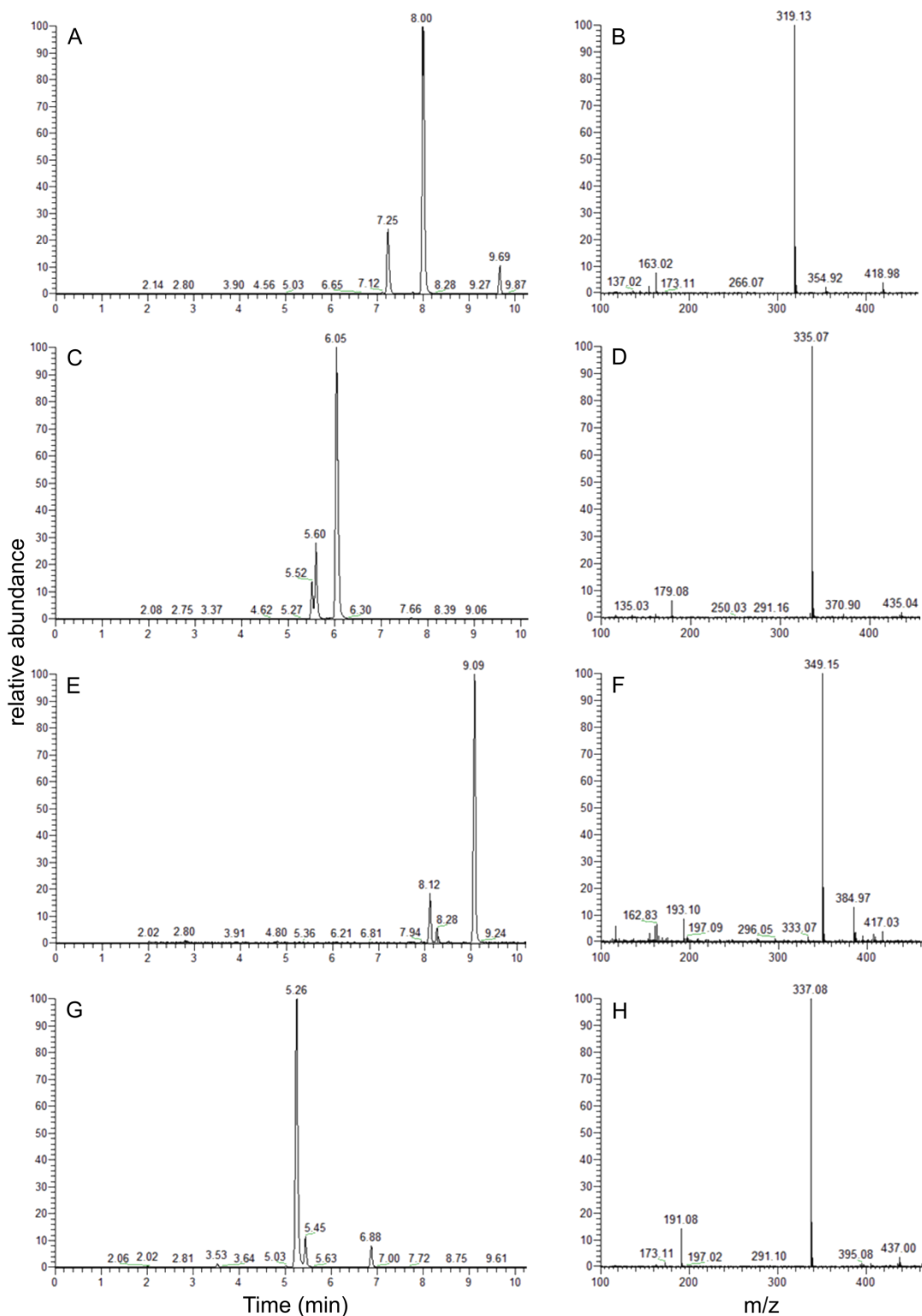
Figure 37: Detection of HCT enzyme products by UPLC-MS. A, C, E and G show chromatograms and B, D, F and H show MS spectra of the various products. Products shown are: *p*-coumaroyl-shikimate (A and B), the peaks at retention times 8.00 and 9.69 min are the trans- and cis-isomers, m/z = 319; caffeoyl shikimate (C and D), peak at retention time 6.05 min, m/z = 335; feruloyl shikimate (E and F), peak at 9.09 min, m/z = 349; and *p*-coumaroyl quinate (G and H), peak at 5.26 min, m/z = 337. The retention times and MS spectra were confirmed with authentic reference compounds.

Table 8: Kinetic parameters of recombinant Brachypodium putative HCT enzymes

Varying substrate	Saturating substrate	Km (μM)	Vmax (μM min^{-1})	Vmax/Km
BdHCT1				
p-coumaroyl-CoA	Shikimic acid	53.2	3.99	7.51×10^{-2}
Caffeoyl-CoA	Shikimic acid	55.3	4.67	8.45×10^{-2}
Feruloyl-CoA	Shikimic acid	12.5	1.35	10.77×10^{-2}
Shikimic acid	*p*-coumaroyl-CoA	36.2	1.89	5.23×10^{-2}
Quinic acid	*p*-coumaroyl-CoA	114	0.39	0.34×10^{-2}
BdHCT2				
p-coumaroyl-CoA	Shikimic acid	102.8	1.64	1.60×10^{-2}
Caffeoyl-CoA	Shikimic acid	53.6	4.30	8.03×10^{-2}
Feruloyl-CoA	Shikimic acid	8.6	0.58	6.80×10^{-2}
Shikimic acid	*p*-coumaroyl-CoA	74.6	2.70	3.61×10^{-2}
Quinic acid	*p*-coumaroyl-CoA	127.8	1.43	1.12×10^{-2}
BdHCT3				
p-coumaroyl-CoA	Shikimic acid	8.7	0.12	11.77×10^{-2}

Km and Vmax values were calculated using the Michaelis-Menton method. See Materials and Methods for concentrations of substrates and assay conditions.

4.5 Discussion

Using whole-genome sequencing, followed by key word searching to identify homozygous SNPs in genes related to cell wall synthesis and saccharification, a SNP in the *sac4* mutant was identified in an HCT gene (Bradi5g14720, BdHCT1). This SNP caused a change in amino acid and was predicted to affect the structure or function of the encoded protein. The alterations in cell wall composition and increased saccharification of the *sac4* mutant appeared to match those observed for *hct* mutants in other plant species. Therefore, this mutation was further investigated to explore whether this could be the cause of the high saccharification of this mutant.

Stem material from *sac4* mutant plants was shown to have a 30% reduction in HCT activity compared to wild-type. Furthermore, recombinant expressed mutant and wild-

type versions of BdHCT1 revealed that the SNP identified in this gene almost completely knocked out activity of the enzyme. It therefore seemed likely that the reduction in HCT activity observed for the *sac4* plant material was a result of the inactivity of this HCT enzyme. The reason for the relatively small reduction in HCT activity of the mutant plant material, despite the mutant recombinant protein almost completely lacking HCT activity, could be explained by redundancy among HCT enzymes in Brachypodium. Arabidopsis, *N. tabacum* and *Medicago sativa* are thought to contain just a single gene encoding HCT (238, 239, 379), whereas studies have predicted there to be four copies in *O. sativa* (377), seven in *Populus* (378, 380), and two in *Z. mays* (381), through searching the proteomes for sequence similarity to known HCT enzymes. Therefore, the Brachypodium genome was searched for further genes with sequence similarity to HCT enzymes, and phylogenetic analysis carried out with HCT proteins in other plants. This led to the identification of two additional putative HCT enzymes in Brachypodium, named BdHCT2 and BdHCT3. All three enzymes were further characterised in terms of their function and expression.

RT-PCR confirmed that all three of the HCT genes are expressed in the stem of Brachypodium plants, although this expression was quite low for BdHCT3. The stem is the material on which HCT activity assays were carried out and where you would expect HCT activity to be high due to lignification of secondary cell walls in this tissue. Indeed, HCT has been shown to accumulate in lignified tissues of stems, and HCT activity has been shown to be highest in stems compared to other tissues in *N. tabacum* (238). It is therefore feasible that the retention of HCT activity in the stems of the mutant *sac4* plants is due to activity of these homologues of the mutated protein. In order to explore this further, the kinetic parameters of the three putative HCT enzymes in Brachypodium were characterised. This showed that when *p*-coumaroyl-CoA and shikimic acid were used as substrates, the substrates required for the HCT step of lignin monomer synthesis, all three showed activity. Furthermore, all three enzymes showed a

preference for shikimic acid over quinic acid as acyl acceptor, as observed for HCT enzymes in other plant species (377, 379, 380). This supports the hypothesis that the BdHCT2 and BdHCT3 enzymes could be contributing to the HCT activity in the *sac4* mutant plants. BdHCT1 and BdHCT2 showed higher efficiency with caffeoyl-CoA and feruloyl-CoA as acyl donors than with *p*-coumaroyl-CoA. This preference for caffeoyl-CoA over *p*-coumaroyl-CoA has been observed previously for HCT enzymes in other plants (377, 379). Furthermore, despite the high efficiency of BdHCT1 and BdHCT2 with feruloyl-CoA as the acyl donor, the Vmax was low in both cases.

The function of the HCT enzyme was first characterised by Hoffmann et al. (379) who showed that a recombinant version of the enzyme could synthesise shikimate and quinate esters from CoA esters. This reaction is required in the lignin monomer synthesis pathway because the enzyme C3H can only use the shikimate and quinate esters of *p*-coumarate as substrates, and not the free acid form or the CoA ester (382, 383). It was also shown that HCT transfers the acyl group more efficiently to shikimate than quinate (379) and that C3H has higher affinity to the shikimate ester than the quinate ester (382). This preference for shikimate over quinate as the acyl acceptor was observed for all three putative Brachypodium HCT enzymes in this study.

The cell wall composition of the *sac4* mutant resembles that of *hct* mutants reported in other plants; it is also compatible with the position of HCT in the monolignol biosynthesis pathway. In all plants in which HCT downregulation has been studied (Arabidopsis, *N. tabacum*, *M. sativa* and *P. nigra*, a reduction in HCT activity has led to a large reduction in lignin content and an increase in the proportion of H units in the lignin to various extents (238, 239, 375). The same is true for *c3h* and *cse* mutants, which act synergistically with HCT (135, 383). These alterations in lignin are a result of the relatively early step in the lignin monomer synthesis pathway that these enzymes catalyse, meaning that a knockdown in enzyme activity results in reduction of the metabolic flux through the monolignol synthesis pathway. Furthermore, HCT is

required for the synthesis of G and S lignin units but not for H units. The range in the level of increased H units in different plants could reflect varying amounts of plasticity in the lignin monomer synthesis pathway in different plants. Alternatively, it could suggest different numbers of copies of the enzyme or simply different methods of gene downregulation. For example, in Arabidopsis, in which HCT expression was knocked down by RNAi, H units were increased to 85% of total lignin monomers (238). In *N. tabacum* however, where HCT was downregulated by virus-induced gene silencing, H units only increased to 8% of total lignin monomers (238).

HCT downregulation in Arabidopsis, *N. tabacum* and *M. sativa* has been shown to result in a reduction in S:G ratio, as observed for the *sac4* mutant in this study (238, 239). It was suggested that this change in S:G ratio could be due to the decline in flux through the monolignol synthesis pathway coupled with a higher affinity of the CAD enzyme than the F5H enzyme for coniferaldehyde, resulting in preference for the pathway to G units over S units (Figure 23) (238). Interestingly, the opposite was observed in HCT downregulated *P. nigra* trees, which exhibited an increase in S:G ratio (375). This suggests that HCT downregulation has different effects on flux through the monolignol biosynthesis pathway in different plants.

HCT downregulation has been shown to result in a reduction in ester-bound ferulic acid content in *M. sativa* (384), as observed in the *sac4* mutant in this study. Ferulic acid content was not quantified in the Arabidopsis or *N. tabacum hct* mutants. It is thought that the donor substrate for the addition of ferulic acid esters on to GAX is feruloyl-CoA (105, 106, 312). If this is the case then this could explain the reduction in ester-bound ferulic acid content in *hct* mutants, because HCT is required for the synthesis of feruloyl-CoA (Figure 23). Alternatively, ferulic acid may be directly added to the arabinose residue. In Arabidopsis, ferulic acid is thought to be biosynthesised from coniferaldehyde in the monolignol biosynthesis pathway, via the enzyme HCALDH (Figure 23). This was hypothesised because an Arabidopsis mutant of HCALDH

displayed an approximately 50% reduction in cell wall ferulic acid content and a recombinant version of the enzyme showed coniferaldehyde dehydrogenase activity (385). More recently however it has been postulated that ferulic acid can also be synthesised from feruloyl-CoA, due to the observation that CCR knockouts in Arabidopsis accumulated both ferulic acid and ferulic acid derivatives (386, 387). This has only been demonstrated in Arabidopsis. Both feruloyl-CoA and coniferaldehyde require HCT for their synthesis.

Interestingly, when HCT activity has been downregulated in other species (Arabidopsis, *N. tabacum* and *M. sativa*) it has resulted in severe dwarfing of the plants (238, 239), whereas phenotyping of *sac4* mutant plants revealed that there was no change in growth, development or stem strength compared to wild-type (Figure 18 and Figure 19). This could be due to the hypothesised functional redundancy between Brachypodium HCT enzymes discussed above, whereas Arabidopsis, *N. tabacum* and *M. sativa* have just a single HCT gene. In fact, a recent study has described a *P. nigra* mutant with a truncated HCT protein and reduced HCT activity that has no noticeable change in growth phenotype; *P. nigra* has seven putative copies of the HCT gene and the authors hypothesise that the lack of stunting is due to redundancy within this group (375). Expression of the other HCT genes was not increased compared to wild-type in the *P. nigra* mutant plants, as observed for the *sac4* mutant. This suggests that there is no feedback mechanism among the multiple copies to make up for loss of activity in one. Another reason for the lack of altered growth phenotype in the *sac4* mutant could be the increased hemicellulose content in this mutant, acting to alleviate any strength reductions of the cell wall. Increases in hemicellulose content have been observed previously in plants with reduced lignin content (219, 340).

An increase in saccharification or digestibility has also been observed in *hct* mutants in other plants (218, 238, 239). In fact, a study examining six mutants of individual enzymes of the lignin monomer synthesis pathway in *M. sativa* found that the *hct*

mutant attained the highest level of saccharification, both with and without pretreatment (218). The most obvious reason for this increase in saccharification is the reduction in lignin content exhibited in *hct* mutants. However, rescuing the lignin deficiency phenotype of a C3H mutant while maintaining the alterations in lignin composition revealed that the high saccharification was retained, suggesting that the change in lignin composition is in fact responsible for the high saccharification of this mutant (240). The same could well be true for *hct* mutants as C3H acts synergistically with HCT. The reduction in ferulic acid could also be contributing to the increased saccharification. Ferulic acid can bind to the arabinose on the side chains of arabinoxylans (103) and can also be incorporated into the lignin polymer through oxidative coupling (160). The ferulic acid can then dimerise and so can form a cross-link between hemicellulose molecules and between lignin and hemicellulose (158, 159). These cross-links are thought to strengthen the cell wall and a reduction in ferulic acid has been shown to increase saccharification and digestibility in a number of plant species (109, 300, 301).

The increase in saccharification of the *sac4* mutant (37%) was lower than that observed in HCT mutants in *M. sativa* (~100% increase for pretreated biomass) and *N. tabacum* (60% increase), although the methods for measuring saccharification are somewhat different in detail in each of these studies. However, the large reduction in biomass of the *M. sativa* and *N. tabacum hct* mutants makes these high saccharification transgenics unattractive for biotechnological or crop breeding applications (218, 238, 239). An ideal gene knockout for the improvement of lignocellulosic biomass for the production of bioethanol would provide a large increase in saccharification with no effect on plant biomass or fitness. The mutation identified in the BdHCT1 gene in this study appears to fit these requirements. Further work is however required in order to fully link the mutation identified in the HCT gene to the increase in saccharification of the *sac4* mutant.

It has been presumed for a number of years that the HCT enzyme also performs the reverse reaction, after C3H, to convert caffeoyl shikimate to caffeoyl-CoA and shikimic acid. However, HCT has been shown to have very low activity for this reaction in *N. tabacum* (379). Recently, it has been shown that an additional enzyme, caffeoyl shikimate esterase (CSE), is in fact likely to play this role in Arabidopsis (135). Vanholme et al. (135) showed that Arabidopsis *cse* T-DNA insertion lines exhibited a reduction in lignin content and an increase in the proportion of H monomers, as with *hct* and *c3h* mutants, and accumulated caffeoyl shikimate. Furthermore, recombinant Arabidopsis CSE can convert caffeoyl shikimate to caffeic acid, but only has very low affinity for *p*-coumaroyl shikimate and cannot convert caffeoyl shikimate to caffeoyl-CoA (135). The authors therefore suggested that CSE converts the caffeoyl shikimate, produced by the activity of HCT, to caffeic acid, although this has not been investigated in other plant species. An additional enzyme, 4CL, would then convert the caffeic acid to caffeoyl-CoA (135). 4CL has been shown to have similar efficiencies with *p*-coumaric acid and caffeic acid (388), another known substrate of 4CL in the monolignol synthesis pathway. It has also been suggested that a protein complex of the C3H and C4H enzymes may bypass the HCT/CSE/4CL reactions by converting *p*-coumaric acid directly to caffeic acid (136). However, this is only achieved at low efficiency (136) and the highly altered lignin composition of *hct* and *cse* mutants suggest that these enzymes are required for normal lignin synthesis. BLAST analysis of the Arabidopsis CSE protein and a putative *O. sativa* CSE protein (135) against the Brachypodium proteome revealed that there are no proteins with high sequence similarity (highest similarity = 39% identity, compared to a 62% identity between the Arabidopsis and *O. sativa* CSE proteins). However, the closest Brachypodium hit by BLAST analysis did achieve a low E-value (E = $8e^{-66}$) and contains the same hydrolase-4 superfamily conserved domain. It would be interesting to test the activity of the putative CSE, HCT and C3H/C4H enzymes in Brachypodium to see how efficiently they perform the caffeoyl shikimate to

caffeic acid, caffeoyl shikimate to caffeoyl-CoA and *p*-coumaric acid to caffeic acid reactions, respectively.

Phylogenetic analysis revealed that the BdHCT3 protein grouped most closely with the *O. sativa* HCT3 and HCT4 proteins (Figure 33). The *O. sativa* HCT4 has previously been shown to have higher efficiency with glycerol as the acyl group acceptor compared to shikimic acid and quinic acid (377). However, no activity was observed with glycerol as the acyl acceptor with any of the Brachypodium HCT enzymes. Furthermore, although the *O. sativa* recombinant HCT4 showed activity with glycerol, hydroxycinnamoyl-glycerides were not identified in the plant material, leading the authors to hypothesise that glycerol may not be an *in vivo* substrate of this enzyme (377). The BdHCT3 enzyme did show activity with *p*-coumaroyl-CoA and shikimic acid as substrates but the Vmax was quite low. This may suggest that BdHCT3 uses an alternative substrate as its primary acyl group acceptor, other than shikimic acid, quinic acid or glycerol. Furthermore, RT-PCR showed that the BdHCT3 gene appears to be relatively highly expressed in the spikelet and have quite low expression in the stems, further supporting a role for this enzyme in something other than lignin synthesis. It would be interesting to test HCT activity in the spikelet of Brachypodium and also to test the activity of the recombinant BdHCT3 protein with alternative substrates. For example, anthranilate, malate, tyramine, spermidine, spermine, putrescine, agmatine, and benzyl alcohol are acceptors of other plant acyl transferases and have been tested as acceptors with HCT enzymes from other plants, although without positive results (377, 379).

Chapter 5 – Mapping the causal mutations of the high saccharification phenotype

5.1 Introduction

Hemicellulose is a major component of cell walls, making up around 40% of the dry weight (16). The main hemicelluloses in both dicots and grasses are xylans. However in dicots the cell wall xylans are decorated with (methyl) glucuronic acid ([Me]GlcA) and are known as glucuronoxylans (GX) (54). In grasses the xylans are known as glucuronoarabinoxylans (GAX) and are decorated abundantly with arabinose, and with smaller amounts of [Me]GlcA, galactosyl, xylosyl and ferulic acid units (27, 54). Both GX and GAX have acetyl group substitutions on some of the backbone xylosyl residues (54) and are synthesised in the Golgi by a suite of enzymes which are responsible for elongation of the backbone and addition of the side chains (67-76, 79-83, 94-97, 106, 108, 109). Hemicelluloses form long chains that associate with cellulose by extensive hydrogen bonding (51-53). It is thought that the hemicellulose chains form a monolayer coating of the cellulose microfibrils but that they can also span between and cross-link the microfibrils (52). Their believed role in the cell wall is to form a protective and strengthening network with cellulose by coating the microfibrils and holding them together; however hemicellulose can also act as a plasticiser to allow extensibility of the cell wall by keeping the fibrils apart from each other (54-56). Despite the major role that hemicellulose appears to play in the cell wall, its influence on saccharification has not been as extensively studied as lignin and is less well understood. Plants with mutations in glycosyltransferases responsible for synthesis of the xylan backbone have been shown to have a reduction in hemicellulose content and an increase in saccharification (273-275). Similar results have been shown for plants expressing hemicellulose degrading enzymes (276). However, increases in hemicellulose have also been associated with an increase in saccharification (219, 283).

The substitutions of hemicellulose have also been shown to affect saccharification but, again, the affect is not clear cut. For example, reduction in the acetylation of hemicelluloses, by either chemical pretreatment or knock-out of the enzymes responsible for adding the acetyl groups, has been shown to increase saccharification in a number of plants (77, 285-287). However, other studies have reported that a reduction in acetylation has no effect on saccharification efficiency (80, 291). Xylans in the cell wall are also commonly substituted with [Me]GlcA. Triple mutants of the enzymes responsible for addition of glucuronic acid to the xylosyl backbone residues led to a complete loss of [Me]GlcA substitution but only a small increase in saccharification (95). The hemicellulose substitution with the most experimentally verified effect on saccharification is ferulic acid. Natural accessions, transgenic plants mutated in ferulic acid substitution enzymes or expressing ferulic acid esterases, and plants with chemically or synthetically altered hemicellulose all show a link between ferulic acid and saccharification, with a reduction in ferulic acid resulting in an increase in saccharification (109, 117, 297-301). Ferulic acid attaches to the arabinosyl side chains of xylans in grasses and can also form a covalent bond with monolignols (103, 160). The ferulic acid molecules can oxidatively dimerise to form a cross link between/within hemicellulose molecules and between hemicellulose and lignin (158, 159). It is also thought that ferulic acid may act as a nucleation site for lignin (161).

The identity of enzymes responsible for the addition of ferulic acid to GAX is still not certain. Enzymes in the BAHD acyltransferase family have been implicated in ferulic acid transfer, but it has been hypothesised that they are involved in the addition of ferulic acid to UDP-arabinose and that an additional enzyme is required for addition of feruloyl arabinose to the xylan backbone (110). A bioinformatics approach has implicated members of the glycosyltransferase GT61 family in this process (106) and a recent study has implicated a GT61 member in the addition of xylosyl residues to GAX (117). Interestingly, knockouts of the gene encoding this enzyme led to a large

reduction in ferulic acid content, leading the authors to propose that presence of the xylose residue aids transfer of ferulic acid to the arabinosyl side chain.

5.2 Bulked segregant analysis by whole genome sequencing

In order to map the causal mutations that led to increased saccharification of the *sac* mutants, bulked segregant analysis of the mutant lines was undertaken to distinguish the causal mutation from the background mutations. This mapping method has recently been developed to use whole genome sequencing and can yield results more rapidly than traditional mapping methods. The method works by firstly backcrossing the mutated plant line to wild-type of the same accession used for the original mutagenesis (Figure 38A). The resulting BCF_1 plant will be heterozygous for all homozygous mutations carried in the F_0 mutant plant (Figure 38B). This BCF_1 plant is then allowed to self in order to produce a BCF_2 progeny with a 1:2:1 ratio of wild-type:heterozygous:homozygous mutant for these mutations (Figure 38C). Due to crossing over of the wild-type and mutant chromosome that occurs during meiosis of the heterozygous BCF_1 plant, each BCF_2 plant with a homozygous mutant genotype for the mutation causing the selected phenotype should have a different set of background mutations. Therefore, if you pool genomic DNA from a large enough number of homozygous mutant BCF_2 plants for the selected phenotype, perform whole genome sequencing on this DNA and then call the SNPs in comparison to the wild-type plant, this will enable you to distinguish between causal and background mutations. The causal mutation for the selected phenotype, as well as any closely linked mutations, should be found in 100% of the sequencing reads because these mutations will be present in all of the plants that were sequenced. Less closely- and un-linked background mutations will occur in less than 100% of the sequencing reads because they will not be present in all of the plants that were sequenced (Figure 38E). Therefore, a plot of allele

frequency (i.e. the proportion of sequencing reads that a mutation occurs in) of each SNP relative to position across the genome should produce a single peak in allele frequencies, where a cluster of SNPs all have allele frequencies close to one. Less closely linked mutations should have allele frequencies of around 0.5 ($0.3 - 0.8$) as they will approximate to heterozygous across the population of BCF_2 homozygous mutant plants.
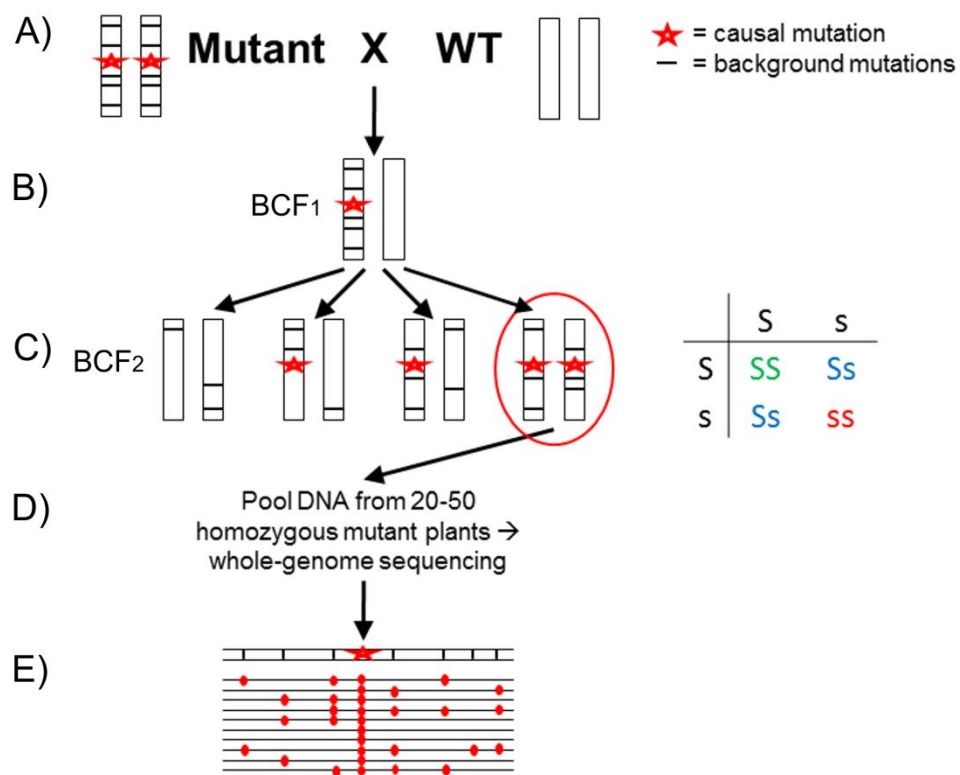
Figure 38: Flow diagram showing the method of bulked segregant analysis by whole genome sequencing used to map the causal mutations of the *sac* mutants. A) The mutant plant is backcrossed to wild-type of the same accession. B) The resulting BCF_1 plant is allowed to self. C) BCF_2 progeny from self of BCF_1 has a 1:2:1 ratio of wild-type:heterozygous:homozygous mutant for the mutation causing the trait of interest. D) 20-50 BCF_2 generation plants with a homozygous mutant phenotype are selected, genomic DNA extracted and pooled, and whole-genome sequencing carried out on this DNA. E) The causal mutation will occur in 100% of the sequencing reads, whereas background mutations will occur in less than 100% of the sequencing reads.

The principle of bulked segregant analysis was first conceived in 1991 by Michelmore et al. (389) in order to identify random fragment length polymorphism (RFLP) and random amplified polymorphic DNA (RAPD) markers linked to a particular trait. Since then the method has been developed to use whole-genome sequencing data rather than markers and has been carried out successfully by a number of groups, working on a variety of organisms (390-394). Using this method of mapping can potentially be quicker and easier than traditional fine mapping methods. Firstly, it reduces the number of BCF_2 plants that need to be screened from thousands to hundreds and does not require phenotyping of the BCF_3 generation (395). Secondly, it does not require outcrossing of the mutant plant to a polymorphic line. Outcrossing can be problematic because a trait may be altered when crossed to a different ecotype, making screening of that phenotype in the BCF_2 pool difficult and inaccurate (392, 393). This could be particularly problematic with a quantitative trait. Because saccharification is a quantitative trait and differences in saccharification have been observed between different accessions of Brachypodium (Sylvain Legay, personal communication), bulked segregant analysis seemed a sensible way to map the causal mutations of the *sac* mutants identified in this study.

Two *sac* mutant lines were selected for mapping, *sac1* and *sac2*, because they showed the largest increase in saccharification and because they did not have any change in their lignin content in comparison to wild-type (Figure 39). As lignin is the best studied cell wall component in terms of both its synthesis and its effect on cell wall saccharification, it seemed less likely that a mutant with a large decrease in lignin content would reveal a novel gene involved in cell wall saccharification.

	sac1	sac2	sac3	sac4	sac5	sac6	sac7	sac8	sac9	sac10	sac11	sac12
lignin			↓↓↓*	↓↓↓↓*	↓↓↓*		↓	↓↓	↓↓*	↓	↓↓↓*	↓↓↓*
hemicellulose				↑↑*	↑↑*		↑↑↑*	↓				
crys cellulose	↓↓↓*			↓↓	↓↓	↓↓	↓↓↓↓*		↑		↓↓	↓↓
ferulic acid	↓↓↓*		↓	↓↓↓*		↓		↓↓			↓	
S:G ratio	↓↓↓↓*	↓		↓↓↓*		↓↓↓*	↓↓↓*	↑	↓↓↓*	↓	↓	
H monolignols	↑↑↑↑*	↑↑↑↑*	↑↑	↑↑↑↑*	↑↑	↑↑↑↑*	↑↑↑↑↑*	↓	↑↑↑↑↑*	↓↓	↑	

Figure 39: Summary of alterations in cell wall composition of the 12 *sac* mutants. Arrows represent percentage increased or decreased compared to wild-type. No arrow: 0-5% increase/decrease, one arrow: 5-10% increase/decrease, two arrows: 10-20% increase/decrease, three arrows: 20-30% increase/decrease, four arrows: 30-40% increase/decrease, five arrows: 40-50% increase/decrease. Asterisks represent a significant difference when compared to wild-type.

5.3 Identifying homozygous mutant plants for the saccharification phenotype among the BCF$_2$ populations

Plants from the *sac1* and *sac2* mutant lines were backcrossed to wild-type and the resulting BCF$_1$ plants allowed to self to produce BCF$_2$ seeds. With the aim of selecting 50 homozygous mutant BCF$_2$ plants for whole-genome sequencing, 260 BCF$_2$ seeds were sown from the self of one BCF$_1$ plant for each of the *sac* mutants (*sac1* and *2*). Of these 260, 241 and 216 germinated for *sac1* and *sac2* respectively. In order to identify plants among this population that were homozygous mutant for the causal mutation, saccharification of the BCF$_2$ population, as well as wild-type, mutant and BCF$_1$ plants, was measured using the automated assay system. The data from these saccharification assays are in Figure 40 and Figure 41 for *sac1* and *sac2* respectively. For both mutants there is clear and consistent higher saccharification in the mutant plants compared to wild-type. Furthermore, the level of saccharification of BCF$_1$ plants fell in between the values for the wild-type and mutant plants. This suggests that there is incomplete dominance of the wild-type allele at the position of the causal mutation. A single copy of the wild-type allele would therefore result in an intermediate phenotype between mutant and wild-type rather than the wild-type phenotype normally expected with a recessive mutation. For both *sac1* and *sac2*, the level of saccharification of the BCF$_2$ population ranges from mutant levels down to wild-type level.
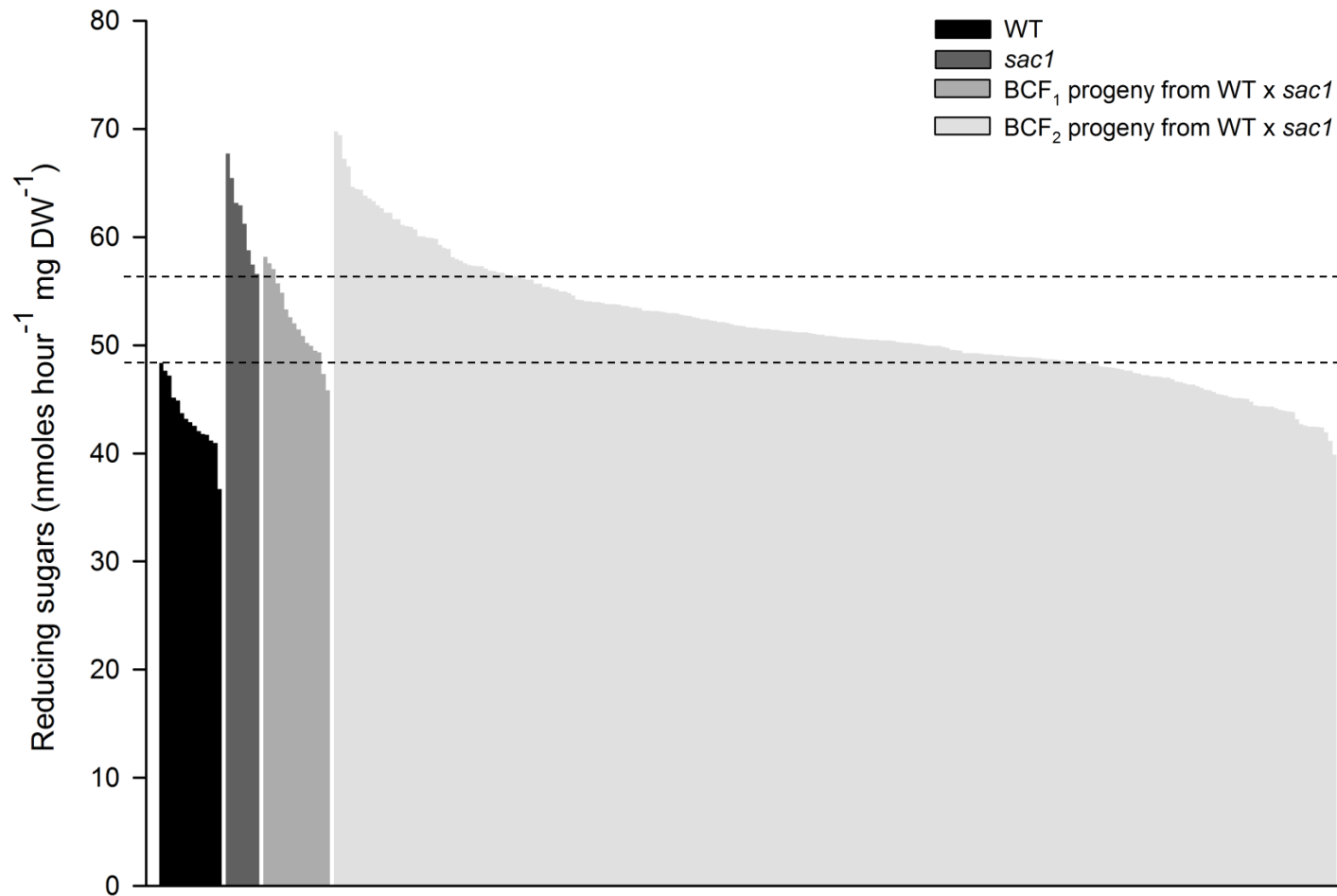
Figure 40: Saccharification analysis to determine homozygous mutants of a BCF$_2$ population from a backcross between wild-type and *sac1*. Level of saccharification of wild-type and *sac1* plants, BCF$_1$ plants and the BCF$_2$ population resulting from a self of one of the BCF$_1$ plants is presented. Dotted lines show thresholds for the lowest *sac1* saccharification and highest wild-type saccharification.
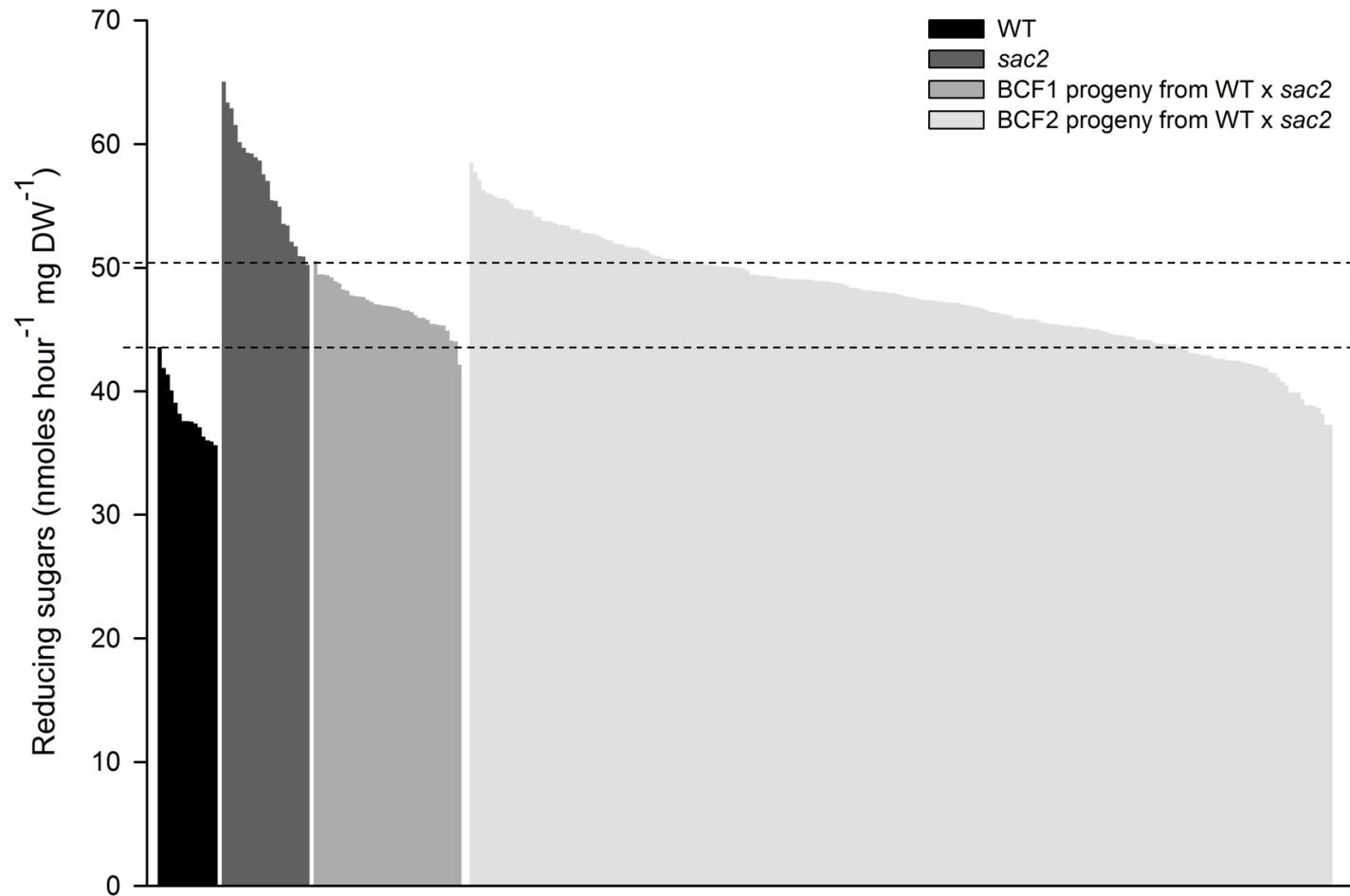
Figure 41: Saccharification analysis to determine homozygous mutants of a BCF$_2$ population from a backcross between wild-type and *sac2*. Level of saccharification of wild-type and *sac2* plants, BCF$_1$ plants and the BCF$_2$ population resulting from a self of one of the BCF$_1$ plants is presented. Dotted lines show thresholds for the lowest *sac2* saccharification and highest wild-type saccharification.

In order to work out the observed percentages of the BCF$_2$ population which were homozygous mutant, heterozygous and homozygous wild-type for the mutation causing increased saccharification, the lowest saccharification value of the mutant plants and the highest saccharification value of the wild-type plants were used as threshold values for each *sac* mutant line individually. BCF$_2$ plants with saccharification values above the mutant threshold were deemed as being homozygous mutant for the causal mutation and BCF$_2$ plants with saccharification values below the wild-type threshold were deemed as being homozygous wild-type for the causal mutation. BCF$_2$ plants with saccharification values between the two thresholds were deemed as being heterozygous for the causal mutation. This uses the assumption that a plant with a similar saccharification phenotype to a plant that is homozygous wild-type/heterozygous/homozygous mutant for the causal mutation will have a matching genotype, which may not be correct in all cases. Table 9 shows the observed percentages of plants with homozygous mutant, heterozygous and homozygous wild-type phenotypes for *sac1* and *sac2*. It also shows the resulting p-value of a chi^2 test used to assess whether these observed percentages differed significantly from the 1:2:1 ratio of wild-type:heterozygous:homozygous mutant phenotype expected with a recessive causal mutation with incomplete dominance of the wild-type allele. The BCF$_2$ populations of the *sac1* and *sac2* mutants both show percentages that do not significantly differ from the expected ratios.

Table 9: Percentages of the BCF$_2$ populations of the *sac1* and *sac2* mutants that had saccharification phenotypes resembling homozygous wild-type, heterozygous and homozygous mutant genotypes for the causal mutation.

	Homozygous wild-type BCF$_2$s (%)	Heterozygous BCF$_2$s (%)	Homozygous mutant BCF$_2$s (%)	Chi^2 (1:2:1) p-value
sac1	25.31	54.77	19.92	0.47
sac2	18.06	54.17	27.78	0.19

The p-value is for a chi^2 test comparing these percentages to the expected 1:2:1 ratio.

5.4 Level of saccharification in the BCF₂ homozygous mutant plants compared to the first and second screens

Figure 42 shows the average increase in saccharification compared to wild-type of the BCF₂ plants deemed to be homozygous mutant for the causal mutation. The figure compares this increase in saccharification to that observed for plants of the same mutant line in the first and second screens used to originally select the mutants, and for the mutant plants grown at the same time as the BCF₂ populations. As you can see, the level of saccharification has remained reasonably consistent, although a larger increase in saccharification was seen for the *sac1* mutant in the second screen compared to the other generations.
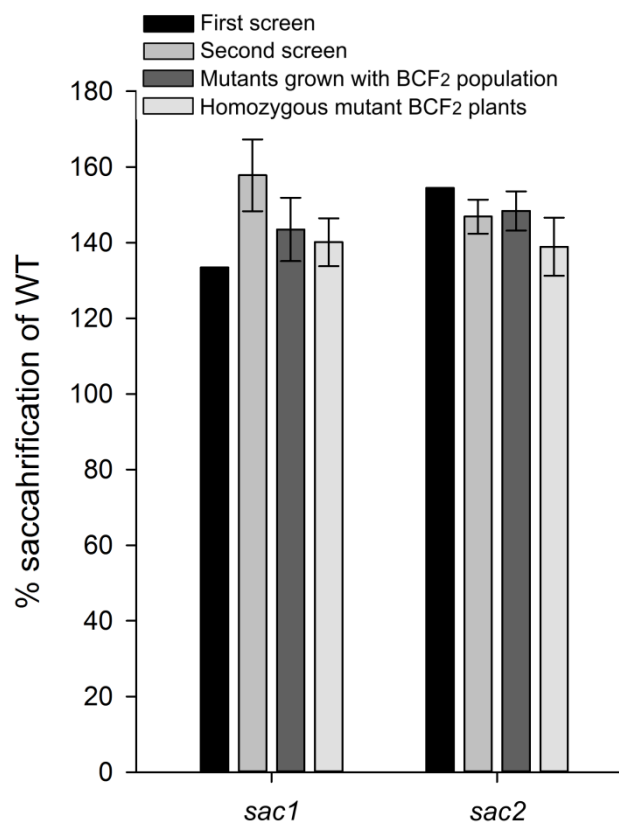
Figure 42: Level of saccharification at different stages of the selection and mapping process for *sac1* and *sac2*. Increase in saccharification compared to wild-type is shown for the BCF₂ plants deemed to be homozygous mutant for the causal mutation, the mutant plants grown at the same time as the BCF₂ populations, and the same mutants in the first and second screens used to originally select the mutants.

5.5 Whole-genome sequencing and SNP calling

For *sac1* and *sac2*, the top 40 BCF_2 plants with the highest saccharification were selected for whole-genome sequencing for mapping. This ensured that none of the BCF_2 plants selected for mapping had saccharification lower than the lowest saccharification value of the mutant plants. Furthermore, the lowest saccharification value of these top 40 plants was within one standard deviation below the average mutant plant saccharification. This would help to ensure as few as possible non-homozygous mutant plants were included in the analysis. DNA from these 40 BCF_2 plants was extracted, pooled and whole-genome sequencing was performed by TGAC, Norwich using an Illumina HiSeq 2000 sequencer. For *sac1* and *sac2*, 11.9 and 12.1 Gbp paired-end sequencing data was achieved, giving coverage of 43.8 and 44.4x, respectively. Assembly of the sequence data to the reference genome (accession Bd21) and identification of homozygous and heterozygous SNPs in the BCF_2 datasets that did not occur in the wild-type dataset was performed by TGAC, Norwich. The rest of the analysis was performed myself. SNPS were removed that did not resemble a mutation caused by sodium azide mutagenesis (i.e. G to A, C to T or A to T, and vice versa) and if the quality of the data for that base was low.

5.6 Plotting the allele frequencies

Once the homozygous and heterozygous SNPs had been called and filtered, the allele frequency of each SNP, i.e. the percentage of sequencing reads for the alternative allele, was plotted against the position of the SNP in the chromosome. This was done separately for each chromosome. As discussed earlier, this should result in a single peak in allele frequency, with a cluster of SNPs with high allele frequency (close to 1.0). This identifies the location of the causal SNP to a relatively narrow region in the genome (392).

5.6.1 Allele frequencies from bulked segregant analysis of *sac2*

The allele frequencies from the bulked segregant analysis of the *sac2* mutant line are shown in Figure 43. The analysis revealed that the majority of SNPs had an allele frequency below 0.9. Although there were a number of SNPs with an allele frequency of 1.0, these generally did not appear in clusters and so are unlikely to be the causal mutation. The lines on the graphs in Figure 43 show a five-point moving average of the SNP allele frequencies and this shows a single peak of high SNP frequencies in chromosome 2, with a cluster of ten SNPs with allele frequencies ≥0.9. The cluster with the second highest number of SNPs with allele frequencies ≥0.9 contained just two SNPs. Within the ten-SNP cluster, eight SNPs were in non-coding regions and two were in exons (Table 10). Only one of these SNPs located in an exon would cause a change in the encoded amino acid sequence, from a serine to an argenine (Table 10). This SNP was in Bradi2g40800 which is annotated as a DEAD-box ATP-dependent RNA helicase. SIFT analysis suggested that this SNP would be tolerated (SIFT score = 0.08, score of ≤0.05 is predicted not be tolerated).
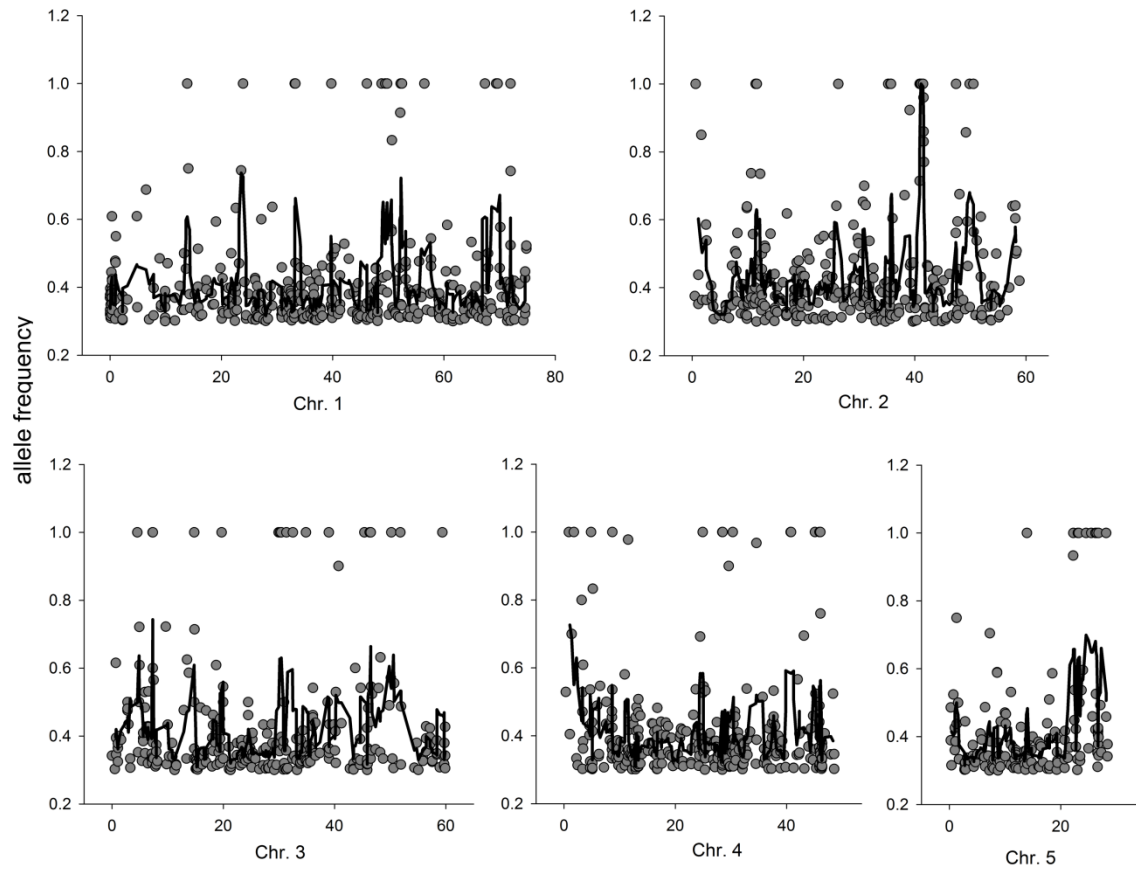
Figure 43: Allele frequency plots for the five chromosomes of the Brachypodium *sac2* mutant. A *sac2* plant was backcrossed to wild-type and DNA from the homozygous mutant plants of the resulting BCF_2 population pooled and whole-genome sequenced. SNPs with an allele frequency above 0.3 and resembling sodium azide-induced mutations were plotted against the position of the SNP in the chromosome. Grey circles show individual SNPs and the black line shows a five point moving average of the SNP allele frequencies, moving one SNP at a time.

Table 10: SNPs in the 10-SNP cluster with allele frequencies ≥0.9 in chromosome 2 of the *sac2* mutant.

SNP position	ref base	alt base	region	gene	annotation	ref aa	alt aa
41029142	A	T	non-coding				
41029143	C	T	non-coding				
41105355	T	A	**exon**	2g40800	Hydrogen exporting ATPase activity	**Serine**	**Argenine**
41133698	T	C	non-coding				
41137445	T	C	non-coding				
41137447	A	T	non-coding				
41172536	T	C	**exon**	2g40870	Putative protein	Leucine	Leucine
41499503	G	A	non-coding				
41501829	T	C	non-coding				
41602195	T	C	non-coding				

5.6.2 Allele frequencies from bulked segregant analysis of *sac1*

The allele frequencies from the bulked segregant analysis of the *sac1* mutant line are shown in Figure 44. The analysis shows a similar result to that of *sac2*, with the majority of SNPs having an allele frequency below 0.9 and those SNPs that do have an allele frequency of >0.9 occurring as single entities rather than in clusters. The five-point moving average reveals a single peak of high SNP frequencies at the start of chromosome 2, with a cluster of 14 SNPs with allele frequencies ≥0.9. The cluster with the second highest number of SNPs with allele frequencies ≥0.9 contained just four SNPs. Within the 14-SNP cluster, ten SNPs were in non-coding regions and, out of the four remaining, three were in exons and only two of these would cause a change in the encoded amino acid sequence (Table 11). One SNP caused a glycine to serine substitution in Bradi2g01480, a glycosyltransferase family GT61. The other SNP caused a change from proline to leucine in Bradi2g02520, which encodes a protein of unknown function (Table 11). BLAST analysis of this gene showed that it contained a conserved kinase domain. SIFT analysis suggested that the SNP in the GT61 protein would not be tolerated, i.e. it would disrupt structure and function drastically (SIFT score = 0.0), whereas the SNP in Bradi2g02520 would be tolerated (SIFT score = 0.17).
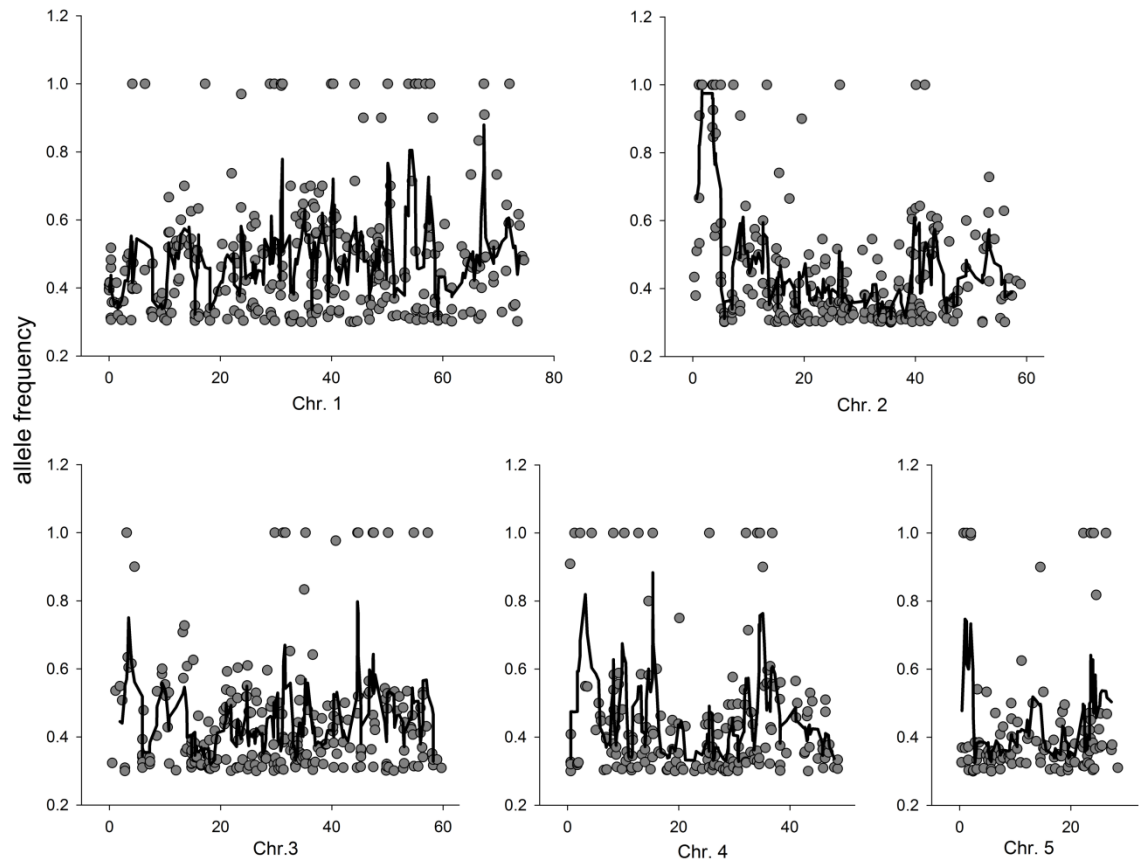
Figure 44: Allele frequency plots for the five chromosomes of the Brachypodium *sac1* mutant. A *sac1* plant was backcrossed to wild-type and DNA from the homozygous mutant plants of the resulting BCF$_2$ population pooled and whole-genome sequenced. SNPs with an allele frequency above 0.3 and resembling sodium azide-induced mutations were plotted against the position of the SNP in the chromosome. Grey circles show individual SNPs and the black line shows a five point moving average of the SNP allele frequencies, moving one SNP at a time.

Table 11: SNPs in the 14-SNP cluster with allele frequencies ≥0.9 in chromosome 2 of the *sac1* mutant.

SNP position	ref base	alt base	region	gene	annotation	ref aa	alt aa
1022576	G	A	**exon**	2g01480	glycosyltransferase family GT61	**glycine**	**serine**
1116925	C	T	non-coding				
1116938	A	G	non-coding				
1206457	T	C	non-coding				
1577928	G	A	non-coding				
1666374	G	A	**exon**	2g02510	Protein kinase family	threonine	threonine
1675680	C	T	**exon**	2g02520	Putative protein	**proline**	**leucine**
1692493	T	C	intron	2g02537	Protein kinase family	-	-
3523570	T	C	non-coding				
3523571	G	A	non-coding				
3527190	T	C	non-coding				
3644990	C	T	non-coding				
3648056	T	C	non-coding				
3663412	T	C	non-coding				

5.6.2.1 Support for the SNP in the GT61 gene as the causal mutation

Examination of the SNPs within the high allele frequency cluster of the *sac1* mutant suggested the GT61 mutation to be the most likely candidate for the causal mutation. This gene was therefore explored further. Phylogenetic analysis of the GT61 family in various monocot and dicot plants showed the Bradi2g01480 protein to be part of grass-specific clade C.III (Figure 45) (117). Clade C is highly expanded in grasses and contains rice and wheat xylan arabinosyltransferases (XATs) and XYLOSYL ARABINOSYL SUBSTITUTION OF XYLAN1 (XAX1) in rice (Figure 45) (97, 117), both involved in GAX substitution. Chiniquy et al. (117) recently showed that a *xax1* mutant in rice exhibited a large reduction in ferulic acid content and an increase in saccharification compared to wild-type. Interestingly the *sac1* mutant presents a very close phenotype, making the SNP in Bradi2g01480 a good candidate for the causal mutation. The rice *xax1* mutant is reported to show a minor 2.6% reduction in matrix polysaccharide xylose content when assessed by 2 M TFA hydrolysis of AIR. In our analysis using this method there was no significant difference in xylose content in *sac1* AIR (Table 5). To assess this further we performed serial extractions of AIR from the *sac1* mutant and wild-type to remove pectinaceous fractions, followed by extraction with 4 M potassium hydroxide to provide a xylan-enriched fraction. Monosaccharide composition analysis of this fraction by hydrolysis and HPAEC revealed a significantly lower level of xylose for *sac1* compared to wild-type ($p = 0.03$, t-test) (Figure 46). This lends further support to the hypothesis that the *sac1* phenotype results from the mutation in a XAX1 homologue in Brachypodium.
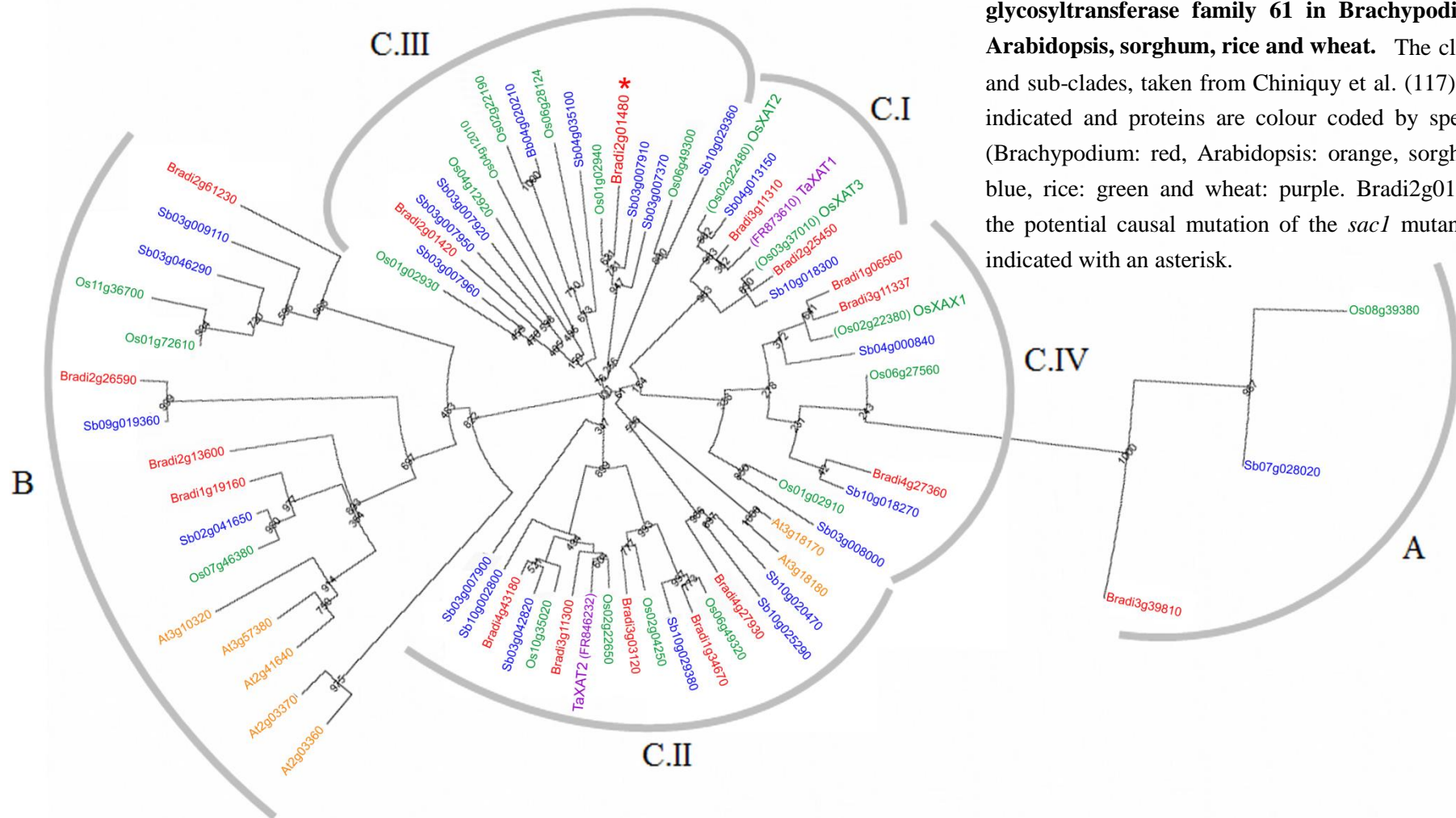
Figure 45: Phylogenetic analysis of the glycosyltransferase family 61 in Brachypodium, Arabidopsis, sorghum, rice and wheat. The clades and sub-clades, taken from Chiniquy et al. (117), are indicated and proteins are colour coded by species (Brachypodium: red, Arabidopsis: orange, sorghum: blue, rice: green and wheat: purple. Bradi2g01480, the potential causal mutation of the *sac1* mutant, is indicated with an asterisk.
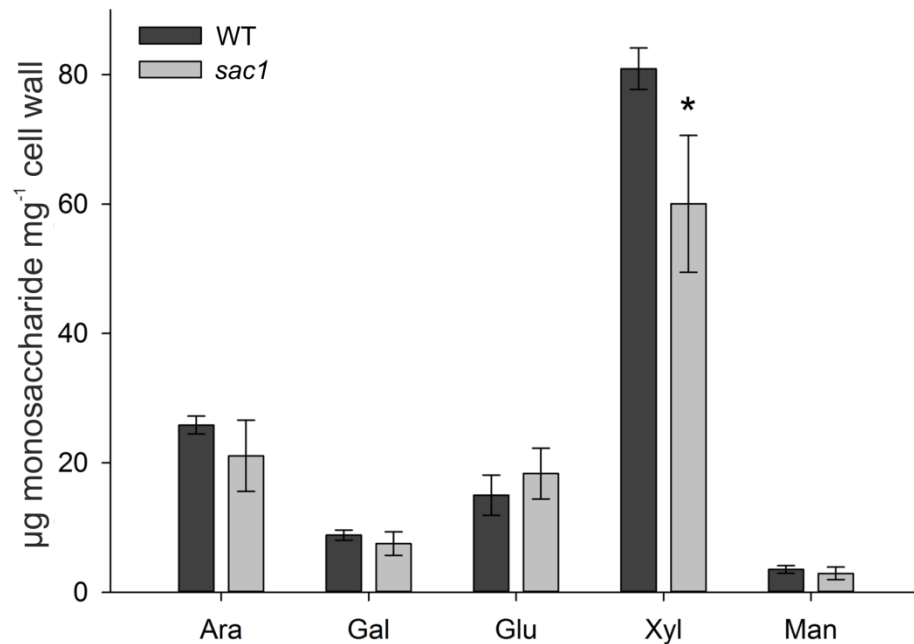
Figure 46: Monosaccharide composition of wild-type and *sac1* xylan-enriched fractions. A xylan-enriched fraction was extracted from AIR using 4 M KOH following pectin removal. Monosaccharides were quantified by hydrolysing this fraction with 2 M TFA, followed by separation and quantification by HPAEC. Data represents mean $\pm$ SD and n = 3. Asterisk indicates a significant difference (p $\leq$ 0.05) when compared to wild-type.

5.7 Discussion

Using a bulked segregant analysis approach to distinguish the causal mutations from background mutations, the mutated genes responsible for the increased saccharification of the *sac1* and *sac2* mutants have been mapped to narrow candidate regions. This analysis suggested that a single locus is responsible for the increased saccharification in each of these mutants, due to a single peak in allele frequencies across the genome. Analysis of the saccharification efficiency of the BCF_1 and BCF_2 generation plants from backcrosses of *sac1* and *sac2* to wild-type suggested that there is incomplete dominance of the wild-type allele at the position of the causal mutations.

5.7.1 Candidate causal mutations of the *sac2* mutant line

Examination of the SNPs within the high allele frequency cluster of the *sac2* mutant revealed that only one of the SNPs caused a change in amino acid. The SNP was located in a DEAD-box ATP-dependent RNA helicase. DEAD-box helicases are involved in the unwinding of RNA and feature in a number of RNA metabolism functions, including transcription, translation, splicing, ribosome assembly and RNA degradation (396). They have been implicated in embryogenesis, cell growth and division, and organ differentiation/maturation (397, 398). The *sac2* mutant does not show a very obvious cell wall phenotype, with just an increase in the percentage of H monomers in the lignin. It could therefore be possible that this line has some sort of RNA metabolism mutation, maybe resulting in an alteration in cell wall deposition. The change in amino acid caused by this SNP was predicted to be tolerated by SIFT analysis. However, the substitution did obtain a relatively low SIFT score (SIFT score = 0.08, score of ≤0.05 is predicted not be tolerated) and this method of analysis is only designed to give an indication of the predicted effect on structure/function of the encoded protein. It would be interesting to look at the transcript profile of this gene in Brachypodium plants to see if the spatial and temporal expression is consistent with a role in cell wall synthesis. Transcript data for this gene is not currently available in a public database.

Alternatively, it is possible that a different gene within this region is responsible for the high saccharification, but the whole-genome sequencing has not identified the actual causal SNP. Examination and BLAST analysis of the genes within the region of high allele frequency SNPs did not reveal anything obviously related to the cell wall. However, just outside the region, but before another SNP with low allele frequency, there is a gene (Bradi2g40590) encoding a peroxidase that BLAST analysis revealed is similar to peroxidase 72 in goatgrass (identity = 89%, E-value = 0.0) and wheat (identity = 88%, E-value = 0.0). Peroxidase 72 has been shown to be involved in lignin synthesis in Arabidopsis and *Zinnia elegans* through oxidation of monolignols for their

polymerisation (399-401). Purified peroxidase 72 protein has been shown to have highest activity for the oxidation of sinapyl alcohol, followed by coniferyl and then *p*-hydroxyl alcohol (400). Supporting this, knockouts of peroxidase 72 in Arabidopsis have been shown to have an increase in the percentage of H monomers measured by thioacidolysis (399), which was also observed in the *sac2* mutant. However, the Arabidopsis knockouts also showed a reduction in total lignin (399) content which was not observed in the *sac2* mutant line. It would be interesting to clone and sequence this gene in wild-type and *sac2* plants to examine whether any SNPs are present which were missed by the whole-genome sequencing. Alternatively it is possible that the causal mutation is one of the SNPs in a non-coding part of the DNA within this region.

5.7.2 Candidate causal mutation of the *sac1* mutant line

Examination of the SNPs within the high allele frequency cluster of the *sac1* mutant revealed that two of the SNPs caused a change in amino acid, but just one of these was predicted to affect protein structure/function by SIFT analysis. This SNP was in a glycosyltransferase gene of family GT61 and, interestingly, the cell wall alterations of the *sac1* mutant appear to be consistent with a GT61 enzyme mutation. GT61 proteins have been implicated previously in arabinoxylan substitution, due to genes encoding this family of proteins being much more highly expressed and numerous in grasses than dicots (106). Furthermore, members of GT61 clade C have been shown to be involved in addition of arabinosyl side chains to grass arabinoxylans (97). The *sac1* mutant shows a reduction in ferulic acid content and reduced xylose in its xylan-enriched cell wall fraction, characteristics also reported for the *xax1* mutant in rice (117). The two mutants also both show reductions in plant height (Figure 19). XAX1 is a GT61 family clade C member that has been proposed to catalyse the transfer of xylosyl residues on to the arabinose side chains of GAX. The observed reduction in ferulic acid is thought to be due to xylosylated arabinoxylan acting as the acceptor for ferulic acid addition (117). We propose that the Brachypodium Bradi2g01480 gene encodes an enzyme with a

similar function to XAX1 in rice. Alternatively, as Bradi2g01480 and XAX1 group in to different sub-clades of clade C, which also contains another sub-clade with proteins involved in arabinosylation of GAX (97), Bradi2g01480 may encode an enzyme involved in a different GAX substitution. The large expansion of grass genes and absence of dicot genes in clade C.III, of which Bradi2g01480 is a member (Figure 45), suggests that this sub-clade is involved in enzymatic activity that is not required in dicots. This could be the addition of xylosyl or ferulic acid residues on to the arabinose side chains of GAX, both of which are unique to grasses. The large reduction in ferulic acid content in *sac1* is likely contributing to the increased saccharification of this mutant as these residues can increase cell wall recalcitrance through cross-linking of hemicellulose molecules and hemicellulose and lignin, forming an even stronger barrier between cellulose and the cellulase enzymes (158-160).

Chapter 6 - Final discussion

Considerable research is currently being carried out to improve lignocellulosic biomass as a feedstock for the production of cost-competitive sustainable bioethanol. The challenge will likely be met by innovations along the production and processing pipeline, and one of the key aspects of this will be to produce plants with cell walls that are more digestible by polysaccharide-degrading enzymes. The majority of studies looking into increasing lignocellulose digestibility have taken a reverse genetic approach, modifying the cell wall through alteration of specific cell wall biosynthesis genes. However, many of these studies have resulted in plants in which increased saccharification is accompanied by dwarfing, reduced stem strength, reduced fitness or an altered growth phenotype. Furthermore, many of the studies use Arabidopsis as a model plant. The cell walls of this annual dicot have many differences to those of the cereals and perennial grasses that form much of the proposed feedstock for bioethanol. The information gained for Arabidopsis may therefore not be translatable to these biofuel crops. The aim of the work carried out for this thesis was to use a forward genetic screen to investigate factors that impact on lignocellulose digestibility in a more empirical manner. We believe this is an important approach due to the complexity and large number of genes involved in cell wall synthesis, and the relatively few cell wall genes that have been functionally characterised (314). Furthermore, using such a screen has the potential to identify plants with increased cell wall digestibility but no notable yield penalty in terms of plant growth. This is extremely important in order to produce a biofuel crop that has the desired improvement in processability, combined with acceptable biomass and ease of cultivation in the field. The model grass *Brachypodium distachyon* was used for the screen, which has similar cell walls and significant co-linearity in gene organisation to the main candidates for bioethanol crops.

The forward genetic screen was carried out on a point mutation population of Brachypodium plants which were screened for saccharification using a high-throughput

175

assay. The screen revealed 12 mutant lines that showed heritable increases in saccharification, from 20 to 60% above the level of wild-type, which were named _saccharification1_ (_sac1_) to _sac12_. Interestingly, these 12 lines showed a range of alterations in cell wall composition although, noticeably, there appeared to be patterns in these alterations among the 12 lines. Lignin content, S:G ratio, % H monomers, ferulic acid content and crystalline cellulose content all only showed significant reductions among the _sac_ mutants. Hemicellulose content however was only significantly increased in the _sac_ mutants. This may be informative as to the effects of these components on cell wall digestibility, as discussed in section 3.9. For example, lignin, ferulic acid and crystalline cellulose have all been shown to negatively impact on lignocellulose saccharification in other plants.

Alternatively, these patterns in cell wall alterations may reveal information on how different ways of measuring saccharification/digestibility can affect the outcome. For example, our saccharification screen does not distinguish between breakdown of cellulose and hemicellulosic sugars when quantifying saccharification. It is therefore perhaps not surprising that none of the mutant lines showed reductions in hemicellulose content. Furthermore, our saccharification screen used an alkaline pretreatment, which is known to primarily solubilise lignin, particularly terminal units of the lignin polymer. The majority of terminal units in grass lignin are G and H units, while S units are predominantly internal. Perhaps, therefore, our screen selected for mutant plants with an increase in G and H units. The literature on saccharification and digestibility use a large range of different pretreatments, intensities and length of treatment, enzymes, and methods of quantification. It is therefore important to consider these differences, the affect they may have on the cell wall and how they may influence the outcome. Indeed, in this study, when four of the _sac_ mutants were tested with different pretreatments, no increase in saccharification was seen when an acid pretreatment was used. It would be interesting to see if the increased saccharification of these 12 mutant lines is retained

when conditions analogous to industrial pretreatments are used. Regarding this, collaboration has recently been proposed with the group of Claus Felby in Copenhagen, who have recently designed a robotic platform similar to the one used in this study but with harsher pretreatments (402). However, it is also important to consider that the plants screened in this study were a point mutation population. The mutations may therefore have caused only partial disruption of protein function, meaning that greater increases in saccharification may be possible with complete knockouts. The purpose of this study was to identify factors that impact on saccharification so that this information can be used for further research, not to produce plants that can be directly used as a feedstock for bioethanol production.

The screen conducted in this study did however have limitations. Firstly, the number of M_1 plants covered by the screen was relatively small. This means that it is unlikely that mutations at every locus within the genome have been covered by the screen and so genes that contribute to the recalcitrance of the cell wall may well have been missed. Furthermore, better statistical design of the original screen may have allowed for a less rigorous selection process for 'high saccharification' mutants, resulting in more mutant lines being followed for characterisation and mapping.

Importantly for use of the information gained in this study for altering biofuel crops, the majority of the *sac* mutants showed no negative impact on growth, development or fitness. Only one of the mutants showed a reduction in biomass, while two of the mutants took longer to germinate and flower. Interestingly, three of the mutants showed an increase in biomass, four mutants showed an increase in seed number, and none of the mutant lines showed a significant reduction in stem strength or stiffness, important for the prevention of lodging in the field. These results suggest that it is possible to achieve an improvement in processability without a concurrent negative impact on the potential of the plants to be cultivated in the field.

Once the 12 high saccharification mutant lines had been identified, it was then important to map the mutations responsible for the increase in saccharification. This would hopefully lead to identification of genes involved in cell wall digestibility which, subsequently, would enable the identification and disruption of homologues of these genes in cereal or biofuel crops. By whole genome sequencing the *sac4* mutant and searching for mutations in cell wall synthesis genes, a potential causal mutation was identified in a putative HCT enzyme, named BdHCT1 in this study. HCT is involved in lignin monomer biosynthesis and the cell wall alterations and increased saccharification of the *sac4* mutant closely matched that of HCT mutants reported in other plants. Furthermore, stem material from *sac4* plants showed reduced HCT activity, to about 70% of wild-type. Recombinant expression of mutant and wild-type versions of the HCT protein showed that the mutation appeared to almost entirely knock out the activity of the enzyme, leading to the hypothesis that there may be multiple copies of the HCT enzyme in Brachypodium, which act non-redundantly. Multiple HCT enzymes have been reported in other plants (375, 377, 378) and this hypothesis was therefore explored further. BLAST and phylogenetic analysis revealed two further putative HCT proteins in Brachypodium, named BdHCT2 and BdHCT3. RT-PCR showed that all three genes were expressed in the stem, although this was quite low for BdHCT3. Kinetic analysis revealed that BdHCT1 and BdHCT2 showed similar substrate preferences and activity to recombinant expressed HCT enzymes from other plants (377, 379, 380), while BdHCT3 did not. We hypothesise that BdHCT1 and BdHCT2 act in the monolignol synthesis pathway, possibly non-redundantly, whereas BdHCT3 most likely has a slightly different function. Phylogenetic analysis groups BdHCT3 with the rice HCT4 enzyme, which has been shown to have high activity with glycerol as a substrate. BdHCT3 does not show any activity with glycerol but it would be interesting to examine the activity of this enzyme with substrates utilised by other acyl transferases in order to assess the role of this enzyme.

Interestingly, the *sac4* plants did not show any alteration in their growth, development, fitness or stem strength, despite a reasonable increase in saccharification compared to wild-type. This is in contrast to HCT mutants reported in other plants which have shown severe dwarfing (218, 238, 239). This may well be due to the hypothesised redundancy within the HCT family in Brachypodium, while the plants in which HCT knockdown causes dwarfing contain just a single copy. BLAST analysis revealed multiple copies of putative HCT genes in other grasses and cereals that have been sequenced (Figure 33). This therefore demonstrates potential for increasing processability without affecting growth and development in biomass crops. However, further work is required in order to fully link the mutant genotype with the high saccharification and low HCT activity phenotypes of the *sac4* mutant. SNP genotyping of the *sac4* BCF$_2$ plants with homozygous mutant-, heterozygous- and homozygous wild-type-like saccharification phenotypes is currently being looked into.

The HCT enzyme has been hypothesised to also carry out the reverse reaction within the monolignol biosynthesis pathway; HCT converts *p*-coumaroyl CoA to *p*-coumaroyl shikimate which is converted to caffeoyl shikimate by C3H, and then HCT acts in reverse to convert caffeoyl shikimate to caffeoyl CoA (379). However, in Arabidopsis an additional enzyme, CSE, has been shown to replace the second HCT reaction by converting caffeoyl shikimate to caffeic acid (135). BLAST analysis of the Arabidopsis CSE protein against the Brachypodium proteome did not reveal a homologue of CSE in Brachypodium. It would be interesting to investigate the activity of the Brachypodium putative HCT enzymes identified in this study for the conversion of caffeoyl shikimate to caffeoyl CoA.

Mapping was carried out on the *sac1* and *sac2* mutant lines, which had the largest increases in saccharification. In both cases, the causal mutation was located to a narrow region of the genome. For *sac1*, a good candidate causal mutation for the increase in saccharification was identified (section 5.6.2). This mutation was in a GT61 gene which

grouped in to the same clade as a rice protein hypothesised to add xylosyl substitutions on to GAX (117). However, this function has not been fully validated. We hypothesise that the Brachypodium GT61 gene highlighted in this study may have a slightly different function, potentially in adding the ferulic acid substitutions on to GAX. It would be interesting to examine the activity of a recombinant expressed version of this protein. It would also be interesting to introduce this gene into a plant that does not contain these xylosyl substitutions and examine the effect. A similar approach has recently been taken to investigate the role of GT61 enzymes in the addition of arabinosyl substitutions on to the xylan backbone. The authors heterologously expressed members of the wheat GT61 family in Arabidopsis, which lacks arabinosyltransferase activity, and found that one member, named TaXAT1, introduced arabinose residues on to GAX (97). The GT61 gene identified in this study could be heterologously expressed simultaneously with *TaXAT1* in Arabidopsis to mimic the xylan with arabinose substitutions found in grasses. The structure and substitutions of the resulting xylan could then be examined.

The candidate region identified for the *sac2* mutant line did not contain any genes with immediately obvious roles related to the cell wall. However, a SNP within this region caused a change in amino acid sequence of an ATP-dependent RNA helicase, involved in RNA metabolism. Little change in cell wall composition was observed for the *sac2* mutant in this study. An RNA metabolism mutation could therefore fit, causing global alteration in cell wall synthesis/assembly rather than affecting a specific cell wall component. It would be interesting to assess the transcript profile of this gene in Brachypodium for compatibility with a role in cell wall synthesis. Another potential causal gene lay just outside the mapped region and has sequence similarity to a peroxidase involved in lignin synthesis in other plants. Mutant Arabidopsis plants of this gene showed an increase in % H monomers, as observed for the *sac2* mutant. It would be interesting to clone and sequence this gene in the wild-type and *sac2* plants to

see if any mutations are present. In this study, only SNPs that caused a change in amino acid were considered as causal mutations. This was a pragmatic approach taken in order to more easily rationalise and validate the effects that a causal mutation might have in the plant. However, it is possible that the causal mutations may be in non-coding regions and have an effect, for example, by altering expression or splicing of a gene.

Further work is required in order to fully validate the effect of the candidate mutations identified in this study on saccharification and the cell wall. Future work should include complementation of these mutants with the wild-type versions of the candidate genes to examine whether the saccharification and cell wall phenotypes are rescued. An alternative approach, as transformation of Brachypodium has proved difficult, could be silencing of the candidate genes in wild-type plants through virus-induced gene silencing (VIGS) and, again, examination of the saccharification and cell wall components. A further approach is to use Brachypodium T-DNA lines of the candidate genes. The T-DNA lines can be assessed for saccharification and cell wall phenotype, complemented with the wild-type gene and then reassessed. Unfortunately, the difficulties experienced with crossing Brachypodium meant that allelism tests between the *sac* mutants could not be carried out within the time frame of this PhD. This is something that should be tested in the future.

In addition to the *sac* mutants that have been mapped to candidate regions in this study, other *sac* mutants would be interesting to examine further. The *sac7* mutant in particular is interesting as it has no change in lignin content, but a large alteration in both cellulose and hemicellulose content. The *sac10* mutant is also interesting as it has no alteration in any of the cell wall components measured in this study. Microscopy of stem sections revealed much less uniformity of the metaxylem in this mutant compared to wild-type, suggesting a developmental phenotype. Interestingly, growth phenotyping and examination of stem strength of this mutant showed no difference compared to wild-type.

In conclusion, using a forward genetic approach, 12 mutant lines with heritable increases in saccharification, and mainly no negative impact on plant growth, fitness or stem strength, have been identified. The two mutants with highest saccharification have been mapped to narrow candidate regions. The potential causal mutation of the *sac1* mutant is located in a GT61 gene, which has been implicated in hemicellulose synthesis and saccharification. However, little research has been carried out on this gene and the exact function is yet to be determined. There are two potential causal mutations in the *sac2* mutant, neither of which has been examined in terms of the effect on saccharification in any plant species. This study has shown the potential of the high-throughput saccharification assay designed at the University of York to accurately and reliably measure saccharification and to be sufficiently robust to allow gene loci responsible for increased saccharification to be identified. Future work will concentrate on verifying the potential causal mutations identified in this study and on mapping the remaining *sac* mutants. The long term goal will be to identify homologues of the responsible genes in biofuel crops and investigate the effect of altering activity or expression on saccharification, cell wall composition, and growth and development of the plants.

Our work shows that forward genetic screening provides a powerful route to identify factors that impact on lignocellulose saccharification. This work gives a positive indication of what could be achieved to improve the saccharification in crop plants, in order to enhance the sustainability and efficiency of lignocellulosic bioethanol production. This is extremely important as world population continues to increase and industrialise, fossil fuel reserves are dwindling and the effects of global warming are becoming apparent. Lignocellulosic bioethanol has the potential to provide a renewable and sustainable energy source to replace liquid transportation fuels, due to the sources of lignocellulosic biomass being low input crops or surplus material from food crops. The replacement of fossil fuels with sustainable alternatives will likely need to be met

by a number of different solutions rather than one single thing. I believe that second generation bioethanol produced from lignocellulosic biomass has the potential to contribute to this, and the work carried out in this study, combined with reverse genetic approaches show the potential to increase the ease and extent of saccharification from this biomass. However, a number of challenges still need to be overcome. I think that the biggest of these is the translation of research that has been carried out in model plants into the relevant crop plants for bioethanol production and then investigating whether these changes are still observed in the field. A further challenge is to achieve an increase in saccharification without the concurrent effect on growth, development or pathogen resistance of the plant. This study gives a positive indication that this can be achieved.

List of abbreviations

4CL	4-coumarate: CoA ligase
AFEX	Ammonia fibre explosion
AGP	Arabinogalactan protein
AIR	Alcohol insoluble residue
Arabidopsis	*Arabidopsis thaliana*
AXX	Arabinoxylan-trisaccharide
AXY	ALTERED XYLAN
BAC	Bacterial artificial chromosome
BAHD	Superfamily named after the first four members of the family to be biochemically characterised (<u>B</u>EAT: benzylalcohol acetyltransferases, <u>A</u>HCT: anthocyanin hydroxycinnamoyl transferase, <u>H</u>CBT: anthranilate hydroxycinnamoyl/benzoyl transferase, <u>D</u>AT: deactylvindoline acetyltransferase)
bp	Base pairs
Brachypodium	*Brachypodium distachyon*
BSA	Bovine serum albumin
BSTFA	Bistrimethylsilyltrifluoroacetamide
C3H	*p*-coumarate 3-hydroxylase
C4H	Cinnamate-4-hydroxylase
CAD	Cinnamoyl alcohol dehydrogenase
CBM	Carbohydrate binding module
CBU	Cellobiase units
C-C	Carbon-carbon bond

CCoAOMT	Caffeoyl coenzyme A O-methyltransferase
CCR	Cinnamoyl-CoA reductase
cDNA	Complementary deoxyribonucleic acid
CDTA	1,2-Diaminocyclohexanetetraacetic acid
CESA	Cellulose synthase
cm	Centimetre
COMT	Caffeic acid O-methyltransferase
CSE	Caffeoyl shikimate esterase
CSL	Cellulose synthase-like
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EGU	Endoglucanase units
ER	Endoplasmic reticulum
ESK	ESKIMO
EST	Expressed sequence tag
F5H	Ferulate 5-hydroxylase
g	Gram
G (monolignol)	Guaiacyl
GAX	Glucuronoarabinoxylan
Gbp	Giga base pairs
GC-MS	Gas chromatography-mass spectrometry
GHG	Greenhouse gas

GlcA	Glucuronic acid
GPI	Glycosyl-phosphatidylinositol
GRP	Glycine-rich protein
GST	Glutathione-S-transferase
GT	Glycosyltransferase
GUX	GLUCURONIC ACID SUBSTITUTION OF XYLAN
GX	Glucuronoxylan
GXM	GLUCURONOXYLAN METHYLTRANSFERASE
h	Hour
H (monolignol)	*p*-hydroxyphenyl
H_2O	Water
H_2O_2	Hydrogen peroxide
H_2SO_4	Sodium hydroxide
HCl	Hydrochloric acid
HCT	Hydroxycinnamoyl-CoA shikimate/quinate transferase
HPAEC	High performance anion exchange chromatography
HPLC	High performance liquid chromatography
HRP	Horseradish peroxidise
IPTG	Isopropyl α-D-thiogalactoside
IRX	IRREGULAR XYLEM
kDa	Kilodalton
KOH	Potassium hydroxide
KOR	KORRIGAN

L	Litre
LB	Lysogeny broth
LC-MS	Liquid chromatography-mass spectrometry
LN	Liquid nitrogen
M	Molar
Mbp	Mega base pairs
MBTH	3-methyl-2-benzothiazolinone hydrazone
MeGlcA	Methyl glucuronic acid
mg	Milligram
min	Minute
ml	Millilitre
MLG	Mixed linkage glucan
mm	Millimetre
mM	Millimolar
MPa	Megapascal
n	Number of biological replicates (unless otherwise stated)
Na_2CO_3	Sodium carbonate
$NaBH_4$	Sodium borohydride
NaN_3	Sodium azide
NaOH	Sodium hydroxide
nm	Nanometre
nmol	Nanomole
OD	Optical density

p	Probability
PAL	Phenylalanine ammonia-lyase
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PRP	Proline-rich protein
PVPP	Polyvinylpolypyrrolidone
QTL	Quantitative trait locus
RG	Rhamnogalacturonan
RNA	Ribonucleic acid
rpm	Revolutions per minute
RT	Room temperature
RT-PCR	Real-time polymerase chain reaction
RWA	REDUCED WALL ACETYLATION
S (monolignol)	Sinapyl
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
sec	Seconds
SIFT	Sorts intolerant from tolerant
SNP	Single nucleotide polymorphism
SSF	Simultaneous saccharification and fermentation
TBL	Trichome birefringence-like
TBST	Tris-buffered saline-tween
T-DNA	Transferred deoxyribonucleic acid

TEMED	Tetramethylethylenediamine
TFA	Trifluoroacetic acid
TGAC	The Genome Analysis Centre
Tris	Tris(hydroxymethyl)aminomethane
UPLC	Ultra-performance liquid chromatography
UTR	Untranslated region
v/v	Volume to volume
w/v	Weight to volume
wild-type	Wild-type
XAT	XYLAN ARABINOSYLTRANSFERASE
XAX	XYLOSYL ARABINOSYL SUBSTITUTION OF XYLAN
X-Gal	5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside
β-O-4	β-aryl ether bond
μg	Microgram
μl	Microlitre
μm	Micron
μM	Micromolar
μmol	Micromole
[Me]GlcA	[methylated] glucuronic acid
°C	Degrees Celsius

Publications arising from this work

Marriott PE, Sibout R, Lapierre C, Fangel J, Willats WGT, Gomez, LD & McQueen-Mason SJ (2014) A range of cell wall alterations enhance saccharification in *Brachypodium distachyon* mutants. *Proceedings of the National Academy of Sciences of the United States of America* 11(40):14601-14606.

References

1.   International Energy Agency (2011) *World Energy Outlook* (OECD/IEA, Paris).

2.   UNPD (United Nations Population Division) (2013) *World Population Prospects: The 2012 Revision* (United Nations, New York).

3.   Quadrelli R & Peterson S (2007) The energy-climate challenge: Recent trends in CO_2 emissions from fuel combustion. *Energy Policy* 35(11):5938-5952.

4.   International Energy Agency (2008) *From 1st to 2nd Generation Biofuel Technologies* (OECD/IEA, Paris).

5.   Goldemberg J (2008) The Brazilian biofuels industry. *Biotechnology for Biofuels* 1:article number 6.

6.   Koh LP & Ghazoul J (2008) Biofuels, biodiversity, and people: Understanding the conflicts and finding opportunities. *Biological Conservation* 141(10):2450-2460.

7.   Poudel BN, Paudel KP, Timilsina G, & Zilberman D (2012) Providing Numbers for a Food versus Fuel Debate: An Analysis of a Future Biofuel Production Scenario. *Applied Economic Perspectives and Policy* 34(4):637-668.

8.   Hochman G, Rajagopal D, Timilsina GR, & Zilberman D (2011) *The Role of Inventory Adjustments in Quantifying Factors Causing Food Price Inflation. Policy Research Working Paper, No. 5744* (The World Bank, Washington, DC).

9.   Mueller SA, Anderson JE, & Wallington TJ (2011) Impact of biofuel production and other supply and demand factors on food price increases in 2008. *Biomass & Bioenergy* 35(5):1623-1632.

10.   Mitchell D (2008) *A Note on Rising Food Prices. Policy Research Working Paper 4682* (The World Bank Development Prospects Group, Washington, D.C.).

11.   Righelato R & Spracklen DV (2007) Environment - Carbon mitigation by biofuels or by saving and restoring forests? *Science* 317(5840):902-902.

12.   Havlik P, *et al.* (2011) Global land-use implications of first and second generation biofuel targets. *Energy Policy* 39(10):5690-5702.

13.   Karp A & Richter GM (2011) Meeting the challenge of food and energy security. *Journal of Experimental Botany* 62(10):3263-3271.

14.   de Vries SC, van de Ven GWJ, van Ittersum MK, & Giller KE (2010) Resource use efficiency and environmental performance of nine major biofuel crops, processed by first-generation conversion techniques. *Biomass & Bioenergy* 34(5):588-601.

15.   Hill J, Nelson E, Tilman D, Polasky S, & Tiffany D (2006) Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proceedings of the National Academy of Sciences of the United States of America* 103(30):11206-11210.

16.   Vogel J (2008) Unique aspects of the grass cell wall. *Current Opinion in Plant Biology* 11(3):301-307.

17. Copeland J & Turley D (2008) National and regional supply/demand balance for agricultural straw in Great Britain. (The National Non-Food Crops Centre, York, UK).

18. Greene N, *et al.* (2004) Growing Energy: how biofuels can help end America's oil dependence. (Natural Resources Defense Council, New York).

19. Dohleman FG, Heaton EA, Arundale RA, & Long SP (2012) Seasonal dynamics of above- and below-ground biomass and nitrogen partitioning in *Miscanthus* x *giganteus* and *Panicum virgatum* across three growing seasons. *Global Change Biology Bioenergy* 4(5):534-544.

20. Heaton EA, Dohleman FG, & Long SP (2008) Meeting US biofuel goals with less land: the potential of Miscanthus. *Global Change Biology* 14(9):2000-2014.

21. Monti A, Fazio S, & Venturi G (2009) Cradle-to-farm gate life cycle assessment in perennial energy crops. *European Journal of Agronomy* 31(2):77-84.

22. Fargione J, Hill J, Tilman D, Polasky S, & Hawthorne P (2008) Land clearing and the biofuel carbon debt. *Science* 319(5867):1235-1238.

23. McLaughlin SB & Walsh ME (1998) Evaluating environmental consequences of producing herbaceous crops for bioenergy. *Biomass & Bioenergy* 14(4):317-324.

24. Sims REH, Mabee W, Saddler JN, & Taylor M (2010) An overview of second generation biofuel technologies. *Bioresoures Technology* 101(6):1570-1580.

25. Harris D & DeBolt S (2010) Synthesis, regulation and utilization of lignocellulosic biomass. *Plant Biotechnology Journal* 8(3):244-262.

26. Mosier N, *et al.* (2005) Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresoures Technology* 96(6):673-686.

27. Albersheim P, Darvill A, Roberts K, Sederoff R, & Staehelin A (2011) *Plant Cell Walls* (Garland Science, Taylor & Francis Group, New York).

28. Somerville C (2006) Cellulose synthesis in higher plants. *Annual Review of Cell and Developmental Biology* 22:53-78.

29. Herth W (1983) Arrays of plasma-membrane rosettes involved in cellulose microfibril formation of *Spirogyra*. *Planta* 159(4):347-356.

30. Nishiyama Y, Langan P, & Chanzy H (2002) Crystal structure and hydrogen-bonding system in cellulose 1(beta) from synchrotron X-ray and neutron fiber diffraction. *Journal of the American Chemical Society* 124(31):9074-9082.

31. Nishiyama Y, Sugiyama J, Chanzy H, & Langan P (2003) Crystal structure and hydrogen bonding system in cellulose 1(alpha), from synchrotron X-ray and neutron fiber diffraction. *Journal of the American Chemical Society* 125(47):14300-14306.

32. Mueller SC & Brown RM (1980) Evidence for an intermembrane component asssociated with a cellulose microfibril-synthesising complex in higher plants. *The Journal of Cell Biology* 84(2):315-326.

33. Kimura S, *et al.* (1999) Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant *Vigna angularis*. *Plant Cell* 11(11):2075-2085.

34.   Arioli T*, et al.* (1998) Molecular analysis of cellulose biosynthesis in *Arabidopsis*. *Science* 279(5351):717-720.

35.   Taylor NG, Laurie S, & Turner SR (2000) Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in Arabidopsis. *Plant Cell* 12(12):2529-2539.

36.   Ding SY & Himmel ME (2006) The maize primary cell wall microfibril: A new model derived from direct visualization. *Journal of Agricultural Food Chemistry* 54(3):597-606.

37.   Delmer DP (1999) Cellulose biosynthesis: Exciting times for a difficult field of study. *Annual Review of Plant Physioly and Plant Molecular Biology* 50:245-276.

38.   Thomas LH*, et al.* (2013) Structure of Cellulose Microfibrils in Primary Cell Walls from Collenchyma. *Plant Physiology* 161(1):465-476.

39.   Fernandes AN*, et al.* (2011) Nanostructure of cellulose microfibrils in spruce wood. *Proceedings of the National Academy of Sciences of the United States of America* 108(47):E1195-E1203.

40.   Persson S*, et al.* (2007) Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* 104(39):15566-15571.

41.   Taylor NG, Howells RM, Huttly AK, Vickers K, & Turner SR (2003) Interactions among three distinct CesA proteins essential for cellulose synthesis. *Proceedings of the National Academy of Sciences of the United States of America* 100(3):1450-1455.

42.   Desprez T*, et al.* (2007) Organization of cellulose synthase complexes involved in primary cell wall synthesis in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* 104(39):15572-15577.

43.   Kurek I, Kawagoe Y, Jacob-Wilk D, Doblin M, & Delmer D (2002) Dimerization of cotton fiber cellulose synthase catalytic subunits occurs via oxidation of the zinc-binding domains. *Proceedings of the National Academy of Sciences of the United States of America* 99(17):11109-11114.

44.   Himmel ME*, et al.* (2007) Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science* 315(5813):804-807.

45.   Nicol F*, et al.* (1998) A plasma membrane-bound putative endo-1,4-beta-D-glucanase is required for normal wall assembly and cell elongation in *Arabidopsis*. *The EMBO Journal* 17(19):5563-5576.

46.   Peng LC, Kawagoe Y, Hogan P, & Delmer D (2002) Sitosterol-beta-glucoside as primer for cellulose synthesis in plants. *Science* 295(5552):147-150.

47.   Robert S, Mouille G, & Hofte H (2004) The mechanism and regulation of cellulose synthesis in primary walls: lessons from cellulose-deficient Arabidopsis mutants. *Cellulose* 11(3-4):351-364.

48.   Cosgrove DJ (2005) Growth of the plant cell wall. *Nature Reviews Molecular Cell Biology* 6(11):850-861.

49. Roudier F, *et al.* (2005) COBRA, an Arabidopsis extracellular glycosyl-phosphatidyl inositol-anchored protein, specifically controls highly anisotropic expansion through its involvement in cellulose microfibril orientation. *Plant Cell* 17(6):1749-1763.

50. Pagant S, *et al.* (2002) *KOBITO1* encodes a novel plasma membrane protein necessary for normal synthesis of cellulose during cell expansion in Arabidopsis. *Plant Cell* 14(9):2001-2013.

51. Valent BS & Albershe.P (1974) Structure of plant cell walls V: binding of xyloglucan to cellulose fibers. *Plant Physiology* 54(1):105-108.

52. McCann MC, Wells B, & Roberts K (1990) Direct visualisation of cross-links in the primary plant cell wall. *Journal of Cell Science* 96:323-334.

53. Hayashi T & Maclachlan G (1984) Pea xyloglucan and cellulose 1: macromolecular organization. *Plant Physiology* 75(3):596-604.

54. Scheller HV & Ulvskov P (2010) Hemicelluloses. *Annual Review of Plant Biology* 61:263-289.

55. Chanliaud E, De Silva J, Strongitharm B, Jeronimidis G, & Gidley MJ (2004) Mechanical effects of plant cell wall enzymes on cellulose/xyloglucan composites. *Plant Journal* 38(1):27-37.

56. Hayashi T, Marsden MPF, & Delmer DP (1987) Pea xyloglucan and cellulose V: xyloglucan-cellulose interaction *in vitro* and *in vivo*. *Plant Physiology* 83(2):384-389.

57. Hayashi T (1989) Xyloglucans in the primary cell wall. *Annual Review of Plant Physiology and Plant Molecular Biology* 40:139-168.

58. Kato Y & Nevins DJ (1986) Fine structure of (1- 3),(1- 4)-BETA-D-glucan from *Zea* shoot cell walls. *Carbohydrate Research* 147(1):69-85.

59. Luttenegger DG & Nevins DJ (1985) Transient nature of a (1- 3),(1- 4)-BETA-D-glucan in *Zea mays* coleoptile cell walls. *Plant Physiology* 77(1):175-178.

60. Gibeaut DM & Carpita NC (1991) Tracing cell wall biogensis in intact cells and plants: selective turnover and alteration of soluble and cell wall polysaccharides in grasses. *Plant Physiology* 97(2):551-561.

61. Buckeridge MS (2010) Seed Cell Wall Storage Polysaccharides: Models to Understand Cell Wall Biosynthesis and Degradation. *Plant Physiology* 154(3):1017-1023.

62. Northcote DH & Pickett-Heaps JD (1966) A function of glogi apparatus in polysaccharide synthesis and transport in root-cap cells of wheat. *Biochemical Journal* 98(1):159-&.

63. Ray PM, Eisinger WR, & Robinson DG (1976) Organelles involved in cell wall polysaccharide formation and transport in pea cells. *Berichte Der Deutschen Botanischen Gesellschaft* 89(1):121-146.

64. Wilson SM, *et al.* (2006) Temporal and spatial appearance of wall polysaccharides during cellularization of barley (*Hordeum vulgare*) endosperm. *Planta* 224(3):655-667.

65. Gordon R & Maclachlan G (1989) Incorporation of UDP-[C-14] glucose into xyloglucan by pea mebranes. *Plant Physiology* 91(1):373-378.

66. Camirand A & Maclachlan G (1986) Biosynthesis of the fucose-containing xyloglucan nonasaccharide by pea microsomal membranes. *Plant Physiology* 82(2):379-383.

67. Faik A, Chileshe C, Sterling J, & Maclachlan G (1997) Xyloglucan galactosyl- and fucosyltransferase activities from pea epicotyl microsomes. *Plant Physiology* 114(1):245-254.

68. Persson S, *et al.* (2007) The *Arabidopsis irregular xylem8* mutant is deficient in glucuronoxylan and homogalacturonan, which are essential for secondary cell wall integrity. *Plant Cell* 19(1):237-255.

69. Brown DM, *et al.* (2007) Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. *Plant Journal* 52(6):1154-1168.

70. 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 Journal* 57(4):732-746.

71. Lee C, O'Neill MA, Tsumuraya Y, Darvill AG, & Ye Z-H (2007) The *irregular xylem9* mutant is deficient in xylan xylosyltransferase activity. *Plant and Cell Physiology* 48(11):1624-1634.

72. Wu A-M, *et al.* (2009) The Arabidopsis IRX10 and IRX10-LIKE glycosyltransferases are critical for glucuronoxylan biosynthesis during secondary cell wall formation. *Plant Journal* 57(4):718-731.

73. Lee C, Zhong R, & Ye Z-H (2012) Arabidopsis Family GT43 Members are Xylan Xylosyltransferases Required for the Elongation of the Xylan Backbone. *Plant and Cell Physiology* 53(1):135-143.

74. 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 Physiology* 153(2):526-541.

75. Wu A-M, *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(2):542-554.

76. Lee C, *et al.* (2007) The PARVUS gene is expressed in cells undergoing secondary wall thickening and is essential for glucuronoxylan biosynthesis. *Plant and Cell Physiology* 48(12):1659-1672.

77. Gille S & Pauly M (2012) *O*-acetylation of plant cell wall polysaccharides. *Frontiers in Plant Science* 3:article number 12.

78. Bastawde KB (1992) Xylan structure, microbial xylanases, and their mode of action. *World Journal of Microbiology and Biotechnology* 8(4):353-368.

79. Manabe Y, *et al.* (2011) Loss-of-Function Mutation of *REDUCED WALL ACETYLATION2* in Arabidopsis Leads to Reduced Cell Wall Acetylation and Increased Resistance to *Botrytis cinerea*. *Plant Physiology* 155(3):1068-1078.

80. Lee C, Teng Q, Zhong R, & Ye Z-H (2011) The Four *Arabidopsis REDUCED WALL ACETYLATION* Genes are Expressed in Secondary Wall-Containing Cells and Required for the Acetylation of Xylan. *Plant and Cell Physiology* 52(8):1289-1301.

81. Manabe Y*, et al.* (2013) Reduced Wall Acetylation Proteins Play Vital and Distinct Roles in Cell Wall *O*-Acetylation in Arabidopsis. *Plant Physiology* 163(3):1107-1117.

82. Yuan Y, Teng Q, Zhong R, & Ye Z-H (2013) The Arabidopsis DUF231 Domain-Containing Protein ESK1 Mediates 2-*O*- and 3-*O*-Acetylation of Xylosyl Residues in Xylan. *Plant and Cell Physiology* 54(7):1186-1199.

83. Xiong G, Cheng K, & Pauly M (2013) Xylan *O*-Acetylation Impacts Xylem Development and Enzymatic Recalcitrance as Indicated by the *Arabidopsis* Mutant *tbl29*. *Molecular Plant* 6(4):1373-1375.

84. Gille S*, et al.* (2011) *O*-Acetylation of *Arabidopsis* Hemicellulose Xyloglucan Requires AXY4 or AXY4L, Proteins with a TBL and DUF231 Domain. *Plant Cell* 23(11):4041-4053.

85. Janbon G, Himmelreich U, Moyrand F, Improvisi L, & Dromer F (2001) Cas1p is a membrane protein necessary for the *O*-acetylation of the *Cryptococcus neoformans* capsular polysaccharide. *Molecular Microbiology* 42(2):453-467.

86. Arming S*, et al.* (2011) The human Cas1 protein: A sialic acid-specific *O*-acetyltransferase? *Glycobiology* 21(5):553-564.

87. Bernard E*, et al.* (2011) Characterization of *O*-Acetylation of *N*-Acetylglucosamine A novel structural variation of bacterial peptidoglycan. *The Journal of Biological Chemistry* 286(27):23950-23958.

88. Franklin MJ & Ohman DE (2002) Mutant analysis and cellular localization of the AlgI, AlgJ, and AlgF proteins required for *O*-acetylation of alginate in *Pseudomonas aeruginosa*. *Journal of Bacteriology* 184(11):3000-3007.

89. Moynihan PJ & Clarke AJ (2010) O-Acetylation of Peptidoglycan in Gram-negative Bacteria: identification and characterisation of peptidoglycan *O*-acetyltransferase in *Neisseria gonorrhoeae*. *The Journal of Biological Chemistry* 285(17):13264-13273.

90. Laaberki M-H, Pfeffer J, Clarke AJ, & Dworkin J (2011) *O*-Acetylation of Peptidoglycan Is Required for Proper Cell Separation and S-layer Anchoring in *Bacillus anthracis*. *The Journal of Biological Chemistry* 286(7):5278-5288.

91. Obel N*, et al.* (2009) Microanalysis of Plant Cell Wall Polysaccharides. *Molecular Plant* 2(5):922-932.

92. Pauly M & Scheller HV (2000) *O*-Acetylation of plant cell wall polysaccharides: identification and partial characterization of a rhamnogalacturonan *O*-acetyl-transferase from potato suspension-cultured cells. *Planta* 210(4):659-667.

93. Pauly M*, et al.* (2013) Hemicellulose biosynthesis. *Planta* 238(4):627-642.

94. Mortimer JC*, et al.* (2010) Absence of branches from xylan in Arabidopsis *gux* mutants reveals potential for simplification of lignocellulosic biomass.

Proceedings of the National Academy of Sciences of the United States of America 107(40):17409-17414.

95. Lee C, Teng Q, Zhong R, & Ye Z-H (2012) Arabidopsis GUX Proteins Are Glucuronyltransferases Responsible for the Addition of Glucuronic Acid Side Chains onto Xylan. *Plant and Cell Physiology* 53(7):1204-1216.

96. Lee C, *et al.* (2012) Three Arabidopsis DUF579 Domain-Containing GXM Proteins are Methyltransferases Catalyzing 4-*O*-Methylation of Glucuronic Acid on Xylan. *Plant and Cell Physiology* 53(11):1934-1949.

97. Anders N, *et al.* (2012) Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto xylan in grasses. *Proceedings of the National Academy of Sciences of the United States of America* 109(3):989-993.

98. Porchia AC, Sorensen SO, & Scheller HV (2002) Arabinoxylan biosynthesis in wheat. Characterization of arabinosyltransferase activity in Golgi membranes. *Plant Physiology* 130(1):432-441.

99. Schultink A, Cheng K, Park YB, Cosgrove DJ, & Pauly M (2013) The Identification of Two Arabinosyltransferases from Tomato Reveals Functional Equivalency of Xyloglucan Side Chain Substituents. *Plant Physiology* 163(1):86-94.

100. Harholt J, *et al.* (2006) ARABINAN DEFICIENT 1 is a putative arabinosyltransferase involved in biosynthesis of Pectic Arabinan in Arabidopsis. *Plant Physiology* 140(1):49-58.

101. Egelund J, *et al.* (2007) Molecular characterization of two *Arabidopsis thaliana* glycosyltransferase mutants, *rra1* and *rra2*, which have a reduced residual arabinose content in a polymer tightly associated with the cellulosic wall residue. *Plant Molecular Biology* 64(4):439-451.

102. Gille S, Haensel U, Ziemann M, & Pauly M (2009) Identification of plant cell wall mutants by means of a forward chemical genetic approach using hydrolases. *Proceedings of the National Academy of Sciences of the United States of America* 106(34):14699-14704.

103. Buanafina MMD (2009) Feruloylation in Grasses: Current and Future Perspectives. *Molecular Plant* 2(5):861-872.

104. Fry SC, Willis SC, & Paterson AEJ (2000) Intraprotoplasmic and wall-localised formation of arabinoxylan-bound diferulates and larger ferulate coupling-products in maize cell-suspension cultures. *Planta* 211(5):679-692.

105. Yoshida-Shimokawa T, Yoshida S, Kakegawa K, & Ishii T (2001) Enzymic feruloylation of arabinoxylan-trisaccharide by feruloyl-CoA : arabinoxylan-trisaccharide *O*-hydroxycinnamoyl transferase from *Oryza sativa*. *Planta* 212(3):470-474.

106. Mitchell RAC, Dupree P, & Shewry PR (2007) A novel bioinformatics approach identifies candidate genes for the synthesis and feruloylation of arabinoxylan. *Plant Physiology* 144(1):43-53.

107. D'Auria JC (2006) Acyltransferases in plants: a good time to be BAHD. *Current Opinion in Plant Biology* 9(3):331-340.

108. Piston F, *et al.* (2010) Down-regulation of four putative arabinoxylan feruloyl transferase genes from family PF02458 reduces ester-linked ferulate content in rice cell walls. *Planta* 231(3):677-691.

109. Bartley LE, *et al.* (2013) Overexpression of a BAHD acyltransferase, *OsAt10*, alters rice cell wall hydroxycinnamic acid content and saccharification. *Plant Physiology* 161(4):1615-1633.

110. Molinari HBC, Pellny TK, Freeman J, Shewry PR, & Mitchell RAC (2013) Grass cell wall feruloylation: istribution of bound ferulate and candidate gene expression in *Brachypodium distachyon*. *Frontiers in Plant Science* 4:article number 50.

111. Konishi T, *et al.* (2007) A plant mutase that interconverts UDP-arabinofuranose and UDP-arabinopyranose. *Glycobiology* 17(3):345-354.

112. Myton KE & Fry SC (1994) Intraprotoplasmic feruloylation of arabinoxylans in *Festuca arundinacea* cell cultures. *Planta* 193(3):326-330.

113. Obel N, Porchia AC, & Scheller HV (2003) Intracellular feruloylation of arabinoxylan in wheat: evidence for feruloyl-glucose as precursor. *Planta* 216(4):620-629.

114. Mastrangelo L, Lenucci M, Piro G, & Dalessandro G (2009) Evidence for intra- and extra-protoplasmic feruloylation and cross-linking in wheat seedling roots. *Planta* 229(2):343-355.

115. Wende G & Fry SC (1997) 2-O-beta-D-xylopyranosyl-(5-O-feruloyl)-L-arabinose, a widespread component of grass cell walls. *Phytochemistry* 44(6):1019-1030.

116. Hoije A, *et al.* (2006) Evidence of the presence of 2-O-beta-D-xylopyranosyl-alpha-L-arabinofuranose side chains in barley husk arabinoxylan. *Carbohydrate Research* 341(18):2959-2966.

117. Chiniquy D, *et al.* (2012) XAX1 from glycosyltransferase family 61 mediates xylosyltransfer to rice xylan. *Proceedings of the National Academy of Sciences of the United States of America* 109(42):17117-17122.

118. Mohnen D (2008) Pectin structure and biosynthesis. *Current Opinion in Plant Biology* 11(3):266-277.

119. Chanliaud E & Gidley MJ (1999) In vitro synthesis and properties of pectin/*Acetobacter xylinus* cellulose composites. *Plant Journal* 20(1):25-35.

120. Popper ZA & Fry SC (2008) Xyloglucan-pectin linkages are formed intra-protoplasmically, contribute to wall-assembly, and remain stable in the cell wall. *Planta* 227(4):781-794.

121. Nakamura A, Furuta H, Maeda H, Takao T, & Nagamatsu Y (2002) Analysis of the molecular construction of xylogalacturonan isolated from soluble soybean polysaccharides. *Bioscience Biotechnology and Biochemistry* 66(5):1155-1158.

122. Showalter AM (1993) Strcuture and function of plant cell wall proteins. *Plant Cell* 5(1):9-23.

123. Bradley DJ, Kjellbom P, & Lamb CJ (1992) Elicitor-induced and wound-induced oxidative cross-linking of a proline-rich plant cell wall protein: a novel, rapid defense response. *Cell* 70(1):21-30.

124. Mellon JE & Helgeson JP (1982) Interaction of a hydroxyproline-rich glycoprotein from tobacco callus with potential pathogens. *Plant Physiology* 70(2):401-405.

125. Vanholst GJ & Varner JE (1984) Reinforced polyproline II conformation in a hyroxyproline-rich cell wall glycoprotein from carrot root. *Plant Physiology* 74(2):247-251.

126. Epstein L & Lamport DTA (1984) An intramolecular linkage involving isodityrosine in extensin. *Phytochemistry* 23(6):1241-1246.

127. Everdeen DS*, et al.* (1988) Enzymic cross-linkage of monomeric extensin precursors *in vitro*. *Plant Physiology* 87(3):616-621.

128. Youl JJ, Bacic A, & Oxley D (1998) Arabinogalactan-proteins from *Nicotiana alata* and *Pyrus communis* contain glycosylphosphatidylinositol membrane anchors. *Proceedings of the National Academy of Sciences of the United States of America* 95(14):7921-7926.

129. Knox JP, Linstead PJ, Peart J, Cooper C, & Roberts K (1991) Developmentally regulated epitopes of cell-surface arabinogalactan proteins and their relation to root tissue pattern formation. *Plant Journal* 1(3):317-326.

130. Jones L, Ennos AR, & Turner SR (2001) Cloning and characterization of *irregular xylem4* (*irx4*): a severely lignin-deficient mutant of *Arabidopsis*. *Plant Journal* 26(2):205-216.

131. Sattler SE & Funnell-Harris DL (2013) Modifying lignin to improve bioenergy feedstocks: strengthening the barrier against pathogens? *Frontiers in Plant Science* 4:article number 70.

132. Davin LB & Lewis NG (2000) Dirigent proteins and dirigent sites explain the mystery of specificity of radical precursor coupling in lignan and lignin biosynthesis. *Plant Physiology* 123(2):453-461.

133. Boerjan W, Ralph J, & Baucher M (2003) Lignin biosynthesis. *Annual Review of Plant Biology* 54:519-546.

134. Ralph J, Peng JP, Lu FC, Hatfield RD, & Helm RF (1999) Are lignins optically active? *Journal of Agricultutral and Food Chemistry* 47(8):2991-2996.

135. Vanholme R*, et al.* (2013) Caffeoyl Shikimate Esterase (CSE) Is an Enzyme in the Lignin Biosynthetic Pathway in *Arabidopsis*. *Science* 341(6150):1103-1106.

136. Chen H-C*, et al.* (2011) Membrane protein complexes catalyze both 4-and 3-hydroxylation of cinnamic acid derivatives in monolignol biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* 108(52):21253-21258.

137. Li LG*, et al.* (2001) The last step of syringyl monolignol biosynthesis in angiosperms is regulated by a novel gene encoding sinapyl alcohol dehydrogenase. *Plant Cell* 13(7):1567-1585.

138. Barakate A*, et al.* (2011) Syringyl Lignin Is Unaltered by Severe Sinapyl Alcohol Dehydrogenase Suppression in Tobacco. *Plant Cell* 23(12):4492-4506.

139. Egertsdotter U*, et al.* (2004) Gene expression during formation of earlywood and latewood in loblolly pine: Expression profiles of 350 genes. *Plant Biology* 6(6):654-663.

140. Ehlting J, *et al.* (2005) Global transcript profiling of primary stems from *Arabidopsis thaliana* identifies candidate genes for missing links in lignin biosynthesis and transcriptional regulators of fiber differentiation. *Plant Journal* 42(5):618-640.

141. Hertzberg M, *et al.* (2001) A transcriptional roadmap to wood formation. *Proceedings of the National Academy of Sciences of the United States of America* 98(25):14732-14737.

142. Miao YC & Liu CJ (2010) ATP-binding cassette-like transporters are involved in the transport of lignin precursors across plasma and vacuolar membranes. *Proceedings of the National Academy of Sciences of the United States of America* 107(52):22728-22733.

143. Alejandro S, *et al.* (2012) AtABCG29 Is a Monolignol Transporter Involved in Lignin Biosynthesis. *Current Biology* 22(13):1207-1212.

144. Sasaki S, Baba Ki, Nishida T, Tsutsumi Y, & Kondo R (2006) The cationic cell-wall-peroxidase having oxidation ability for polymeric substrate participates in the late stage of lignification of *Populus alba* L. *Plant Molecular Biology* 62(6):797-807.

145. Takahama U (1995) Oxidation of hydroxycinnamic avid and hydroxycinnamyl alcohol derivatives by laccase and peroxidase: interactions among *p*-hydroxyphenyl, guaiacyl and syringyl groups during the oxidation reactions. *Physiologia Plantarum* 93(1):61-68.

146. Bao W, Omalley DM, Whetten R, & Sederoff RR (1993) A laccase associated with lignification in lololly pine xylem. *Science* 260(5108):672-674.

147. Driouich A, Laine AC, Vian B, & Faye L (1992) Characterisation and loacalization of laccase forms in stem and cell cultures of sycamore. *Plant Journal* 2(1):13-24.

148. Musel G, *et al.* (1997) Structure and distribution of lignin in primary and secondary cell walls of maize coleoptiles analyzed by chemical and immunological probes. *Planta* 201(2):146-159.

149. Otter T & Polle A (1997) Characterisation of acidic and basic apoplastic peroxidases from needles of Norway spruce (*Picea abies*, L, Karsten) with respect to lignifying substrates. *Plant and Cell Physiology* 38(5):595-602.

150. Quiroga M, *et al.* (2000) A tomato peroxidase involved in the synthesis of lignin and suberin. *Plant Physiology* 122(4):1119-1127.

151. Blee KA, *et al.* (2003) A lignin-specific peroxidase in tobacco whose antisense suppression leads to vascular tissue modification. *Phytochemistry* 64(1):163-176.

152. Berthet S, *et al.* (2011) Disruption of *LACCASE4* and *17* results in tissue-specific alterations to lignification of *Arabidopsis thaliana* stems. *Plant Cell* 23(3):1124-1137.

153. Li YH, Kajita S, Kawai S, Katayama Y, & Morohoshi N (2003) Down-regulation of an anionic peroxidase in transgenic aspen and its effect on lignin characteristics. *Journal of Plant Research* 116(3):175-182.

154. Hatfield R & Vermerris W (2001) Lignin formation in plants. The dilemma of linkage specificity. *Plant Physiology* 126(4):1351-1357.

155. Onnerud H, Zhang LM, Gellerstedt G, & Henriksson G (2002) Polymerization of monolignols by redox shuttle-mediated enzymatic oxidation: A new model in lignin biosynthesis I. *Plant Cell* 14(8):1953-1962.

156. Ralph J, *et al.* (2004) Lignins: Natural polymers from oxidative coupling of 4-hydroxyphenylpropanoids. *Phytochemistry Reviews* 3(1-2):29-60.

157. Campbell MM & Sederoff RR (1996) Variation in lignin content and composition - Mechanism of control and implications for the genetic improvement of plants. *Plant Physiology* 110(1):3-13.

158. Hatfield RD, Ralph J, & Grabber JH (1999) Cell wall cross-linking by ferulates and diferulates in grasses. *Journal of the Science of Food and Agriculture* 79(3):403-407.

159. Ralph J, Quideau S, Grabber JH, & Hatfield RD (1994) Identification and synthesis of new ferulic acid dehydrodimers present in grass cell walls. *Journal of the Chemical Society-Perkin Transactions 1* (23):3485-3498.

160. Scalbert A, Monties B, Lallemand JY, Guittet E, & Rolando C (1985) Ether linkage between phenolic-acids and lignin fractions from wheat straw. *Phytochemistry* 24(6):1359-1362.

161. Ralph J, Grabber JH, & Hatfield RD (1995) Lignin-feruate crosslinks in grasses. Active incorporation of ferulate polysaccharide esters into ryegrass lignins. *Carbohydrate Research* 275(1):167-178.

162. Jung HJG (2003) Maize stem tissues: ferulate deposition in developing internode cell walls. *Phytochemistry* 63(5):543-549.

163. MacAdam JW & Grabber JH (2002) Relationship of growth cessation with the formation of diferulate cross-links and p-coumaroylated lignins in tall fescue leaf blades. *Planta* 215(5):785-793.

164. Kamisaka S, Takeda S, Takahashi K, & Shibata K (1990) Diferulic and ferulic acid in the cell wall of *Avena* coleoptiles - their relationships to mechanical properties of the cell wall. *Physiologia Plantarum* 78(1):1-7.

165. Tan KS, Hoson T, Masuda Y, & Kamisaka S (1992) Involvement of cell wall-bound diferulic acid in light-induced decrease in growth-rate and cell wall extensibility of *Oryza* coleoptiles. *Plant and Cell Physiology* 33(2):103-108.

166. Ikegawa T, Mayama S, Nakayashiki H, & Kato H (1996) Accumulation of diferulic acid during the hypersensitive response of oat leaves to *Puccinia coronata* f.sp *avenae* and its role in the resistance of oat tissues to cell wall degrading enzymes. *Physiological and Molecular Plant Pathology* 48(4):245-255.

167. Bily AC, *et al.* (2003) Dehydrodimers of ferulic acid in maize grain pericarp and aleurone: Resistance factors to *Fusarium graminearum*. *Phytopathology* 93(6):712-719.

168. Lyons PC, Hipskind J, Vincent JR, & Nicholson RL (1993) Phenylpropanoid dissemination in maize resistnat or susceptible to *Helminthosporium maydis*. *Maydica* 38(3):175-181.

169. Santiago R, *et al.* (2006) Diferulate content of maize sheaths is associated with resistance to the Mediterranean corn borer *Sesamia nonagrioides* (Lepidoptera : Noctuidae). *Journal of Agricultural and Food Chemistry* 54(24):9140-9144.

170. Fry SC (1982) Phenolic components of the primary cell wall - feruloylated disaccharides of D-galactose and L-arabinose from spinach polysaccharide. *Biochemical Journal* 203(2):493-504.

171. Colquhoun IJ, Ralet MC, Thibault JF, Faulds CB, & Williamson G (1994) Structure identification of feruloylated oligosaccharides from sugar-beet pulp by NMR-spectroscopy. *Carbohydrate Research* 263(2):243-256.

172. Harada H & Cote WA (1985) Structure of wood. *Biosynthesis and biodegradation of wood components*, ed Higuchi T (Academic Press, Orlando).

173. Lord EM & Mollet JC (2002) Plant cell adhesion: A bioassay facilitates discovery of the first pectin biosynthetic gene. *Proceedings of the National Academy of Sciences of the United States of America* 99(25):15843-15845.

174. Marry M, *et al.* (2006) Cell-cell adhesion in fresh sugar-beet root parenchyma requires both pectin esters and calcium cross-links. *Physiologia Plantarum* 126(2):243-256.

175. Abe H, Funada R, Imaizumi H, Ohtani J, & Fukazawa K (1995) Dynamic changes in the arrangement of cortical microtubules in conifer tracheids during differentiation. *Planta* 197(2):418-421.

176. Prodhan A, Funada R, Ohtani J, Abe H, & Fukazawa K (1995) Orientation of microtubules in developing tension-wood fibers of japanese ash (*Fraxinus mandshurica* var. *japonica*). *Planta* 196(3):577-585.

177. Suslov D & Verbelen JP (2006) Cellulose orientation determines mechanical anisotropy in onion epidermis cell walls. *Journal of Experimental Botany* 57(10):2183-2192.

178. Plomion C, Leprovost G, & Stokes A (2001) Wood formation in trees. *Plant Physiology* 127(4):1513-1523.

179. Donaldson LA (2001) Lignification and lignin topochemistry - an ultrastructural view. *Phytochemistry* 57(6):859-873.

180. Terashima N, Fukushima K, & Tsuchiya S (1986) Heterogeneity in formation of lignin VII: an autoradiographic study on the formation of guaiacyl and syringyl lignin in poplar. *Journal of Wood Chemistry and Technology* 6(4):495-504.

181. Whitney SEC, Gothard MGE, Mitchell JT, & Gidley MJ (1999) Roles of cellulose and xyloglucan in determining the mechanical properties of primary plant cell walls. *Plant Physiology* 121(2):657-663.

182. Carpita NC & Gibeaut DM (1993) Structural models of primary cell walls in flowering plants - consistency of molecular structure with the physical properties of the walls during growth. *Plant Journal* 3(1):1-30.

183. Giddings TH, Brower DL, & Staehelin LA (1980) Visualisation of particle complexes in the plasma membrane of *Micrasterias denticulata* associated with the formation of cellulose fibrils in primary and secondary cell walls. *The Journal of Cell Biology* 84(2):327-339.

184. Levy S, York WS, Stuikeprill R, Meyer B, & Staehelin LA (1991) Simulations of the static and dynamic molecular conformations of xyloglucan - the role of the fucosylated side-chain in surface-specific side-chain folding. *Plant Journal* 1(2):195-215.

185. Bromley JR*, et al.* (2013) GUX1 and GUX2 glucuronyltransferases decorate distinct domains of glucuronoxylan with different substitution patterns. *Plant Journal* 74(3):423-434.

186. Nakamura A, Furuta H, Maeda H, Takao T, & Nagamatsu Y (2002) Structural studies by stepwise enzymatic degradation of the main backbone of soybean soluble polysaccharides consisting of galacturonan and rhamnogalacturonan. *Bioscience Biotechnology and Biochemistry* 66(6):1301-1313.

187. Ishii T & Matsunaga T (2001) Pectic polysaccharide rhamnogalacturonan II is covalently linked to homogalacturonan. *Phytochemistry* 57(6):969-974.

188. McCann MC & Roberts K (1991) Architecture of the Primary Cell Wall. *The Cytoskeletal Basis of Plant Growth and Form*, ed Lloyd CW (London Academic Press).

189. Carpita NC (1983) Hemicellulosic polymers of cell walls of *Zea* coleoptiles. *Plant Physiology* 72(2):515-521.

190. Knox JP, Linstead PJ, King J, Cooper C, & Roberts K (1990) Pectin esterification is spatially regulated both within cell walls and between developing tissues of root apices. *Planta* 181(4):512-521.

191. Carpita NC (1989) Pectic polysaccharides of maize coleoptiles and proso millet cells in liquid culture. *Phytochemistry* 28(1):121-125.

192. Tassinari T, Macy C, Spano L, & Ryu DDY (1980) Energy-requirements and process design considerations in a compresion-milling pretreatment of cellulosic wastes for enzymatic hydrolysis. *Biotechnology amd Bioengineering* 22(8):1689-1705.

193. Kumar R & Wyman CE (2009) Does change in accessibility with conversion depend on both the substrate and pretreatment technology? *Bioresource Technology* 100(18):4193-4202.

194. Galbe M & Zacchi G (2007) Pretreatment of lignocellulosic materials for efficient bioethanol production. *Biofuels*, ed Olsson L (Springer-Verlag Berlin, Berlin), Vol 108, pp 41-65.

195. Cara C, Ruiz E, Ballesteros I, Negro MJ, & Castro E (2006) Enhanced enzymatic hydrolysis of olive tree wood by steam explosion and alkaline peroxide delignification. *Process Biochemistry* 41(2):423-429.

196. Saha BC, Iten LB, Cotta MA, & Wu YV (2005) Dilute acid pretreatment, enzymatic saccharification and fermentation of wheat straw to ethanol. *Process Biochemistry* 40(12):3693-3700.

197. Alvira P, Tomas-Pejo E, Ballesteros M, & Negro MJ (2010) Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review. *Bioresource Technology* 101(13):4851-4861.

198. Wyman CE, *et al.* (2005) Comparative sugar recovery data from laboratory scale application of leading pretreatment technologies to corn stover. *Bioresource Technology* 96(18):2026-2032.

199. Zhao XB, Cheng KK, & Liu DH (2009) Organosolv pretreatment of lignocellulosic biomass for enzymatic hydrolysis. *Applied Microbiology and Biotechnology* 82(5):815-827.

200. Tanahashi M (1990) Characterisation and degradation mechanisms of wood components by steam explosion and utilization of exploded wood. *Wood Research* (77):49-117.

201. Glasser WG & Wright RS (1998) Steam-assisted biomass fractionation. II. Fractionation behavior of various biomass resources. *Biomass & Bioenergy* 14(3):219-235.

202. Laureano-Perez L, Teymouri F, Alizadeh H, & Dale BE (2005) Understanding factors that limit enzymatic hydrolysis of biomass. *Applied Biochemistry and Biotechnology* 121:1081-1099.

203. Jorgensen H, Kristensen JB, & Felby C (2007) Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. *Biofuels Bioproducts & Biorefining-Biofpr* 1(2):119-134.

204. Alfani F, Gallifuoco A, Saporosi A, Spera A, & Cantarella M (2000) Comparison of SHF and SSF processes for the bioconversion of steam-exploded wheat straw. *Journal of Industrial Microbiology and Biotechnology* 25(4):184-192.

205. Wingren A, Galbe M, & Zacchi G (2003) Techno-economic evaluation of producing ethanol from softwood: Comparison of SSF and SHF and identification of bottlenecks. *Biotechnology Progress* 19(4):1109-1117.

206. Zhu JY & Pan XJ (2010) Woody biomass pretreatment for cellulosic ethanol production: Technology and energy consumption evaluation. *Bioresource Technology* 101(13):4992-5002.

207. Zhang YHP, Himmel ME, & Mielenz JR (2006) Outlook for cellulase improvement: Screening and selection strategies. *Biotechnology Advances* 24(5):452-481.

208. Holtzapple M, Cognata M, Shu Y, & Hendrickson C (1990) Inhibition of *Trichoderma reesei* cellulase by sugars and solvents. *Biotechnology and Bioengineering* 36(3):275-287.

209. Xiao ZZ, Zhang X, Gregg DJ, & Saddler JN (2004) Effects of sugar inhibition on cellulases and beta-glucosidase during enzymatic hydrolysis of softwood substrates. *Applied Biochemistry and Biotechnology* 113:1115-1126.

210. Panagiotou G & Olsson L (2007) Effect of compounds released during pretreatment of wheat straw on microbial growth and enzymatic hydrolysis rates. *Biotechnology and Bioengineering* 96(2):250-258.

211. Cantarella M, Cantarella L, Gallifuoco A, Spera A, & Alfani F (2004) Effect of inhibitors released during steam-explosion treatment of poplar wood on subsequent enzymatic hydrolysis and SSF. *Biotechnology Progress* 20(1):200-206.

212. Mansfield SD, Mooney C, & Saddler JN (1999) Substrate and enzyme characteristics that limit cellulose hydrolysis. *Biotechnology Progress* 15(5):804-816.

213. Ragauskas AJ*, et al.* (2006) The path forward for biofuels and biomaterials. *Science* 311(5760):484-489.

214. Yang B & Wyman CE (2008) Pretreatment: the key to unlocking low-cost cellulosic ethanol. *Biofuels Bioproducts & Biorefining-Biofpr* 2(1):26-40.

215. Gu TY, Held MA, & Faik A (2013) Supercritical CO_2 and ionic liquids for the pretreatment of lignocellulosic biomass in bioethanol production. *Environmental Technology* 34(13-14):1735-1749.

216. Lee SH, Doherty TV, Linhardt RJ, & Dordick JS (2009) Ionic Liquid-Mediated Selective Extraction of Lignin From Wood Leading to Enhanced Enzymatic Cellulose Hydrolysis. *Biotechnology and Bioengineering* 102(5):1368-1376.

217. Vogel KP, Pedersen JF, Masterson SD, & Toy JJ (1999) Evaluation of a filter bag system for NDF, ADF, and IVDMD forage analysis. *Crop Science* 39(1):276-279.

218. Chen F & Dixon RA (2007) Lignin modification improves fermentable sugar yields for biofuel production. *Nature Biotechnology* 25(7):759-761.

219. Van Acker R*, et al.* (2013) Lignin biosynthesis perturbations affect secondary cell wall composition and saccharification yield in *Arabidopsis thaliana*. *Biotechnology for Biofuels* 6:article number 46.

220. Lacombe E*, et al.* (1997) Cinnamoyl-CoA reductase, the first committed enzyme of the lignin branch biosynthetic pathway: Cloning, expression and phylogenetic relationships. *Plant Journal* 11(3):429-441.

221. Boudet AM, Kajita S, Grima-Pettenati J, & Goffner D (2003) Lignins and lignocellulosics: a better control of synthesis for new and improved uses. *Trends in Plant Science* 8(12):576-581.

222. Van Acker R*, et al.* (2014) Improved saccharification and ethanol yield from field-grown transgenic poplar deficient in cinnamoyl-CoA reductase. *Proceedings of the National Academy of Sciences of the United States of America* 111(2):845-850.

223. Fu C*, et al.* (2011) Downregulation of Cinnamyl Alcohol Dehydrogenase (CAD) Leads to Improved Saccharification Efficiency in Switchgrass. *Bioenergy Research* 4(3):153-164.

224. Vailhe MAB*, et al.* (1998) Effect of down-regulation of cinnamyl alcohol dehydrogenase on cell wall composition and on degradability of tobacco stems. *Journal of the Science of Food and Agriculture* 76(4):505-514.

225. Min D, Li Q, Jameel H, Chiang V, & Chang H-M (2012) The Cellulase-Mediated Saccharification on Wood Derived from Transgenic Low-Lignin Lines of Black Cottonwood (*Populus trichocarpa*). *Applied Biochemistry and Biotechnology* 168(4):947-955.

226. Mansfield SD, Kang K-Y, & Chapple C (2012) Designed for deconstruction - poplar trees altered in cell wall lignification improve the efficacy of bioethanol production. *New Phytologist* 194(1):91-101.

227. Zhang K, *et al.* (2012) An Engineered Monolignol 4-O-Methyltransferase Depresses Lignin Biosynthesis and Confers Novel Metabolic Capability in *Arabidopsis*. *Plant Cell* 24(7):3135-3152.

228. Duceppe M-O, *et al.* (2012) Assessment of Genetic Variability of Cell Wall Degradability for the Selection of Alfalfa with Improved Saccharification Efficiency. *Bioenergy Research* 5(4):904-914.

229. Stabile SS, *et al.* (2012) Expression of genes from the lignin synthesis pathway in guineagrass genotypes differing in cell-wall digestibility. *Grass Forage Science* 67(1):43-54.

230. Goff BM, Murphy PT, & Moore KJ (2012) Comparison of common lignin methods and modifications on forage and lignocellulosic biomass materials. *Journal of the Science of Food and Agriculture* 92(4):751-758.

231. Liu J, *et al.* (2013) Fungal pretreatment of switchgrass for improved saccharification and simultaneous enzyme production. *Bioresource Technology* 135:39-45.

232. Xu B, *et al.* (2011) Silencing of 4-coumarate:coenzyme A ligase in switchgrass leads to reduced lignin content and improved fermentable sugar yields for biofuel production. *New Phytologist* 192(3):611-625.

233. Guo DG, *et al.* (2001) Improvement of in-rumen digestibility of alfalfa forage by genetic manipulation of lignin *O*-methyltransferases. *Transgenic Research* 10(5):457-464.

234. Studer MH, *et al.* (2011) Lignin content in natural *Populus* variants affects sugar release. *Proceedings of the National Academy of Sciences of the United States of America* 108(15):6300-6305.

235. Li X, *et al.* (2010) Lignin monomer composition affects *Arabidopsis* cell-wall degradability after liquid hot water pretreatment. *Biotechnology for Biofuels* 3:article number 27.

236. Ziebell A, *et al.* (2010) Increase in 4-Coumaryl Alcohol Units during Lignification in Alfalfa (*Medicago sativa*) Alters the Extractability and Molecular Weight of Lignin. *Journal of Biolical Chemistry* 285(50):38961-38968.

237. Russell WR, Forrester AR, Chesson A, & Burkitt MJ (1996) Oxidative coupling during lignin polymerization is determined by unpaired electron delocalization within parent phenylpropanoid radicals. *Archives of Biochemistry and Biophysics* 332(2):357-366.

238. Hoffmann L, *et al.* (2004) Silencing of hydroxycinnamoy-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. *Plant Cell* 16(6):1446-1465.

239. Shadle G, *et al.* (2007) Down-regulation of hydroxycinnamoyl-CoA: Shikimate hydroxycinnamoyl transferase in transgenic alfalfa affects lignification, development and forage quality. *Phytochemistry* 68(11):1521-1529.

240. Bonawitz ND, *et al.* (2014) Disruption of Mediator rescues the stunted growth of a lignin-deficient *Arabidopsis* mutant. *Nature* 509(7500):376-380.

241. Russell WR, Burkitt MJ, Scobbie L, & Chesson A (2006) EPR investigation into the effects of substrate structure on peroxidase-catalyzed phenylpropanold oxidation. *Biomacromolecules* 7(1):268-273.

242. Riboulet C, Lefevre B, Denoue D, & Barriere Y (2008) Genetic variation in maize cell wall for lignin content, lignin structure, *p*-hydroxycinnamic acid content, and digestibility in set of 19 lines at silage harvest maturity. *Maydica* 53(1):11-19.

243. Vanholme R, *et al.* (2012) A Systems Biology View of Responses to Lignin Biosynthesis Perturbations in *Arabidopsis*. *Plant Cell* 24(9):3506-3529.

244. Abdulrazzak N, *et al.* (2006) A *coumaroyl-ester-3-hydroxylase* insertion mutant reveals the existence of nonredundant *meta*-hydroxylation pathways and essential roles for phenolic precursors in cell expansion and plant growth. *Plant Physiology* 140(1):30-48.

245. Coleman HD, Samuels AL, Guy RD, & Mansfield SD (2008) Perturbed Lignification Impacts Tree Growth in Hybrid Poplar-A Function of Sink Strength, Vascular Integrity, and Photosynthetic Assimilation. *Plant Physiology* 148(3):1229-1237.

246. Schilmiller AL, *et al.* (2009) Mutations in the *cinnamate 4-hydroxylase* gene impact metabolism, growth and development in Arabidopsis. *Plant Journal* 60(5):771-782.

247. Reddy MSS, *et al.* (2005) Targeted down-regulation of cytochrome P450 enzymes for forage quality improvement in alfalfa (*Medicago sativa* L.). *Proceedings of the National Academy of Sciences of the United States of America* 102(46):16573-16578.

248. Prashant S, *et al.* (2011) Down-regulation of *Leucaena leucocephala* cinnamoyl-CoA reductase (LlCCR) gene induces significant changes in phenotype, soluble phenolic pools and lignin in transgenic tobacco. *Plant Cell Reports* 30(12):2215-2231.

249. Leple JC, *et al.* (2007) Downregulation of cinnamoyl-coenzyme a reductase in poplar: Multiple-level phenotyping reveals effects on cell wall polymer metabolism and structure. *Plant Cell* 19(11):3669-3691.

250. O'Connell A, *et al.* (2002) Improved paper pulp from plants with suppressed cinnamoyl-CoA reductase or cinnamyl alcohol dehydrogenase. *Transgenic Research* 11(5):495-503.

251. Huang J, *et al.* (2010) Functional Analysis of the Arabidopsis *PAL* Gene Family in Plant Growth, Development, and Response to Environmental Stress. *Plant Physiology* 153(4):1526-1538.

252. Kajita S, Hishiyama S, Tomimura Y, Katayama Y, & Omori S (1997) Structural characterization of modified lignin in transgenic tobacco plants in which the activity of 4-coumarate:coenzyme A ligase is depressed. *Plant Physiology* 114(3):871-879.

253. Stout A, *et al.* (2014) Growth under field conditions affects lignin content and productivity in transgenic *Populus trichocarpa* with altered lignin biosynthesis. *Biomass Bioenergy* 68:228-239.

254. Penning BW*, et al.* (2014) Genetic determinants for enzymatic digestion of lignocellulosic biomass are independent of those for lignin abundance in a maize recombinant inbred population. *Plant Physiology* epub ahead of print: pp.114.242446.

255. Jeoh T*, et al.* (2007) Cellulase digestibility of pretreated biomass is limited by cellulose accessibility. *Biotechnol. Bioeng.* 98(1):112-122.

256. Hall M, Bansal P, Lee JH, Realff MJ, & Bommarius AS (2010) Cellulose crystallinity - a key predictor of the enzymatic hydrolysis rate. *FEBS Journal* 277(6):1571-1582.

257. Harris DM*, et al.* (2012) Cellulose microfibril crystallinity is reduced by mutating C-terminal transmembrane region residues CESA1^{A903V} and CESA3^{T942I} of cellulose synthase. *Proceedings of the National Academy of Sciences of the United States of America* 109(11):4098-4103.

258. Su Y, Zhao G, Wei Z, Yan C, & Liu S (2012) Mutation of Cellulose Synthase Gene Improves the Nutritive Value of Rice Straw. *Asian-Australasian Journal of Animal Sciences* 25(6):800-805.

259. Harris D, Stork J, & Debolt S (2009) Genetic modification in cellulose-synthase reduces crystallinity and improves biochemical conversion to fermentable sugar. *Global Change Biology Bioenergy* 1(1):51-61.

260. Din N*, et al.* (1991) Non-hydrolytic disruption of cellulose fibers by the binding domain of a bacterial cellulase. *Bio-Technology* 9(11):1096-1099.

261. Shpigel E, Roiz L, Goren R, & Shoseyov O (1998) Bacterial cellulose-binding domain modulates in vitro elongation of different plant cells. *Plant Physiology* 117(4):1185-1194.

262. Abramson M, Shoseyov O, & Shani Z (2010) Plant cell wall reconstruction toward improved lignocellulosic production and processability. *Plant Science* 178(2):61-72.

263. Baker JO*, et al.* (2000) Investigation of the cell-wall loosening protein expansin as a possible additive in the enzymatic saccharification of lignocellulosic biomass. *Applied Biochemistry and Biotechnology* 84-6:217-223.

264. Sahoo DK, Stork J, DeBolt S, & Maiti IB (2013) Manipulating cellulose biosynthesis by expression of mutant Arabidopsis *proM24::CESA3^{ixr1-2}* gene in transgenic tobacco. *Plant Biotechnology Journal* 11(3):362-372.

265. Gillmor CS, Poindexter P, Lorieau J, Palcic MM, & Somerville C (2002) Alpha-glucosidase I is required for cellulose biosynthesis and morphogenesis in *Arabidopsis*. *The Journal of Cell Biology* 156(6):1003-1013.

266. Fujita M*, et al.* (2013) The *anisotropy1* D604N Mutation in the Arabidopsis Cellulose Synthase1 Catalytic Domain Reduces Cell Wall Crystallinity and the Velocity of Cellulose Synthase Complexes. *Plant Physiology* 162(1):74-85.

267. Williamson RE*, et al.* (2001) Morphology of *rsw1*, a cellulose-deficient mutant of *Arabidopsis thaliana*. *Protoplasma* 215(1-4):116-127.

268. Beeckman T*, et al.* (2002) Genetic complexity of cellulose synthase A gene function in Arabidopsis embryogenesis. *Plant Physiology* 130(4):1883-1893.

269. Kokubo A, Sakurai N, Kuraishi S, & Takeda K (1991) Culm brittleness of Barley (*Hordeum vulgare* L.) mutants is caused by smaller number of cellulose molecules in cell wall. *Plant Physiology* 97(2):509-514.

270. Joshi CP, *et al.* (2011) Perturbation of Wood Cellulose Synthesis Causes Pleiotropic Effects in Transgenic Aspen. *Molecular Plant* 4(2):331-345.

271. Burton RA, *et al.* (2010) A Customized Gene Expression Microarray Reveals That the Brittle Stem Phenotype *fs2* of Barley Is Attributable to a Retroelement in the *HvCesA4* Cellulose Synthase Gene. *Plant Physiology* 153(4):1716-1728.

272. Tanaka K, *et al.* (2003) Three distinct rice cellulose synthase catalytic subunit genes required for cellulose synthesis in the secondary wall. *Plant Physiology* 133(1):73-83.

273. Petersen PD, *et al.* (2012) Engineering of plants with improved properties as biofuels feedstocks by vessel-specific complementation of xylan biosynthesis mutants. *Biotechnology for Biofuels* 5:article number 84.

274. Brown D, *et al.* (2011) Arabidopsis genes *IRREGULAR XYLEM* (*IRX15*) and *IRX15L* encode DUF579-containing proteins that are essential for normal xylan deposition in the secondary cell wall. *Plant Journal* 66(3):401-413.

275. Lee CH, Teng Q, Huang WL, Zhong RQ, & Ye ZH (2009) Down-Regulation of PoGT47C Expression in Poplar Results in a Reduced Glucuronoxylan Content and an Increased Wood Digestibility by Cellulase. *Plant and Cell Physiology* 50(6):1075-1089.

276. Kaida R, *et al.* (2009) Loosening Xyloglucan Accelerates the Enzymatic Degradation of Cellulose in Wood. *Molecular Plant* 2(5):904-909.

277. Moxley G, Gaspar AR, Higgins D, & Xu H (2012) Structural changes of corn stover lignin during acid pretreatment. *Journal of Industrial Microbiology and Biotechnology* 39(9):1289-1299.

278. Yang B & Wyman CE (2004) Effect of xylan and lignin removal by batch and flowthrough pretreatment on the enzymatic digestibility of corn stover cellulose. *Biotechnology and Bioengineering* 86(1):88-95.

279. Wang ZJ, Zhu JY, Zalesny RS, Jr., & Chen KF (2012) Ethanol production from poplar wood through enzymatic saccharification and fermentation by dilute acid and SPORL pretreatments. *Fuel* 95(1):606-614.

280. Zhu W, Houtman CJ, Zhu JY, Gleisner R, & Chen KF (2012) Quantitative predictions of bioconversion of aspen by dilute acid and SPORL pretreatments using a unified combined hydrolysis factor (CHF). *Process Biochemistry* 47(5):785-791.

281. Zhu JY, *et al.* (2010) Ethanol production from SPORL-pretreated lodgepole pine: preliminary evaluation of mass balance and process energy efficiency. *Applied Microbiology and Biotechnology* 86(5):1355-1365.

282. Pena MJ, *et al.* (2007) *Arabidopsis irregular xylem8* and *irregular xylem9*: Implications for the complexity of glucuronoxylan biosynthesis. *Plant Cell* 19(2):549-563.

283. Xu N, *et al.* (2012) Hemicelluloses negatively affect lignocellulose crystallinity for high biomass digestibility under NaOH and H2SO4 pretreatments in *Miscanthus. Biotechnology for Biofuels* 5:artcle number 58.

284. Appeldoorn MM, Kabel MA, Van Eylen D, Gruppen H, & Schols HA (2010) Characterization of oligomeric xylan structures from corn fiber resistant to pretreatment and simultaneous saccharification and fermentation. *Journal of Agricultural and Food Chemistry* 58(21):11294-11301.

285. Grohmann K, Mitchell DJ, Himmel ME, Dale BE, & Schroeder HA (1989) The role of ester groups in resistance of plant cell polysaccharides to enzymatic hydrolysis. *Applied Biochemistry and Biotechnology* 20-1:45-61.

286. Chang VS & Holtzapple MT (2000) Fundamental factors affecting biomass enzymatic reactivity. *Applied Biochemistry and Biotechnology* 84-6:5-37.

287. Kong FR, Engler CR, & Soltes EJ (1992) Effects of cell wall acetate, xylan backbone, and lignin on enzymatic hydrolysis of aspen wood. *Applied Biochemistry and Biotechnology* 34-5:23-35.

288. Biely P, Mackenzie CR, Puls J, & Schneider H (1986) Cooperativity of esterases and xylanases in the enzymatic degradation of acetyl xylan. *Bio-Technology* 4(8):731-733.

289. Chen X, *et al.* (2012) The impacts of deacetylation prior to dilute acid pretreatment on the bioethanol process. *Biotechnology for Biofuels* 5:article number 8.

290. Chen X, Shekiro J, Elander R, & Tucker M (2012) Improved Xylan Hydrolysis of Corn Stover by Deacetylation with High Solids Dilute Acid Pretreatment. *Industrial & Engineering Chemistry Research* 51(1):70-76.

291. Pogorelko G, *et al.* (2011) Post-synthetic modification of plant cell walls by expression of microbial hydrolases in the apoplast. *Plant Molecular Biology* 77(4-5):433-445.

292. Olsson L & HahnHagerdal B (1996) Fermentation of lignocellulosic hydrolysates for ethanol production. *Enzyme and Microbial Technology* 18(5):312-331.

293. Franden MA, Pilath HM, Mohagheghi A, Pienkos PT, & Zhang M (2013) Inhibition of growth of *Zymomonas mobilis* by model compounds found in lignocellulosic hydrolysates. *Biotechnology for Biofuels* 6:article number 99.

294. Helle S, Cameron D, Lam J, White B, & Duff S (2003) Effect of inhibitory compounds found in biomass hydrolysates on growth and xylose fermentation by a genetically engineered strain of *S. cerevisiae. Enzyme and Microbial Technology* 33(6):786-792.

295. Watanabe T & Koshijima T (1988) Evidence for an ester linkage between lignin and glucuronnic acid in lignin carbohydrate complexes by DDQ-oxidation. *Agricultural and Biological Chemistry* 52(11):2953-2955.

296. Bi C, Rice JD, & Preston JF (2009) Complete Fermentation of Xylose and Methylglucuronoxylose Derived from Methylglucuronoxylan by *Enterobacter asburiae* Strain JDR-1. *Applied Environmental Microbiology* 75(2):395-404.

297. Lam TBT, Iiyama K, & Stone BA (2003) Hot alkali-labile linkages in the walls of the forage grass *Phalaris aquatica* and *Lolium perenne* and their relation to in vitro wall digestibility. *Phytochemistry* 64(2):603-607.

298. Hartley RD (1972) *p*-coumaric acid and ferulic acid components of cell walls of ryegrass and their relationships with lignin and digestibility. *Journal of the Science of Food and Agriculture* 23(11):1347-&.

299. Barros-Rios J, Malvar RA, Jung H-JG, Bunzel M, & Santiago R (2012) Divergent selection for ester-linked diferulates in maize pith stalk tissues. Effects on cell wall composition and degradability. *Phytochemistry* 83:43-50.

300. Grabber JH, Ralph J, & Hatfield RD (1998) Ferulate cross-links limit the enzymatic degradation of synthetically lignified primary walls of maize. *Journal of the Science of Food and Agriculture* 46(7):2609-2614.

301. Buanafina M, Langdon T, Hauck B, Dalton S, & Morris P (2008) Expression of a fungal ferulic acid esterase increases cell wall digestibility of tall fescue (*Festuca arundinacea*). *Plant Biotechnology Journal* 6(3):264-280.

302. Lionetti V, *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(2):616-621.

303. Francocci F, *et al.* (2013) Analysis of pectin mutants and natural accessions of *Arabidopsis* highlights the impact of de-methyl-esterified homogalacturonan on tissue saccharification. *Biotechnology for Biofuels* 6:article number 163.

304. Coenen GJ, Bakx EJ, Verhoef RP, Schols HA, & Voragen AGJ (2007) Identification of the connecting linkage between homo- or xylogalacturonan and rhamnogalacturonan type I. *Carbohydrate Polymers* 70(2):224-235.

305. Powell DA, Morris ER, Gidley MJ, & Rees DA (1982) Conformations and interactions of pectins: II. Influence of residue sequence on chian association in calcium pectate gels. *Journal of Molecular Biology* 155(4):517-531.

306. Jacquet G, Pollet B, & Lapierre C (1995) New ether-linked ferulic acid-coniferyl alcohol dimers identified in grass straws. *Journal of Agricultural and Food Chemistry* 43(10):2746-2751.

307. Vogel JP, *et al.* (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463(7282):763-768.

308. Bevan M, Garvin D, & Vogel J (2010) *Brachypodium distachyon* genomics for sustainable food and fuel production. *Current Opinion in Biotechnology* 21(2):211-217.

309. Huo NX, *et al.* (2006) Construction and characterization of two BAC libraries from *Brachypodium distachyon*, a new model for grass genomics. *Genome* 49(9):1099-1108.

310. Gu YQ, *et al.* (2009) A BAC-based physical map of *Brachypodium distachyon* and its comparative analysis with rice and wheat. *BMC Genomics* 10:article number 496.

311. Filiz E, *et al.* (2009) Molecular, morphological, and cytological analysis of diverse *Brachypodium distachyon* inbred lines. *Genome* 52(10):876-890.

312. Garvin DF, *et al.* (2010) An SSR-based genetic linkage map of the model grass *Brachypodium distachyon*. *Genome* 53(1):1-13.

313. Vain P, *et al.* (2008) *Agrobacterium*-mediated transformation of the temperate grass *Brachypodium distachyon* (genotype Bd21) for T-DNA insertional mutagenesis. *Plant Biotechnology Journal* 6(3):236-245.

314. Mewalal R, Mizrachi E, Mansfield SD, & Myburg AA (2014) Cell wall-related proteins of unknown function: missing links in plant cell wall development. *Plant Cell Physiology* 55(6):1031-1043.

315. Yang X, Ye C-Y, Bisaria A, Tuskan GA, & Kalluri UC (2011) Identification of candidate genes in *Arabidopsis* and *Populus* cell wall biosynthesis using text-mining, co-expression network analysis and comparative genomics. *Plant Science* 181(6):675-687.

316. Truntzler M, *et al.* (2010) Meta-analysis of QTL involved in silage quality of maize and comparison with the position of candidate genes. *Theoretical and Applied Genetics* 121(8):1465-1482.

317. Dalmais M, *et al.* (2013) A TILLING Platform for Functional Genomics in *Brachypodium distachyon*. *Plos One* 8(6):article number e65503.

318. Gomez LD, Whitehead C, Barakate A, Halpin C, & McQueen-Mason SJ (2010) Automated saccharification assay for determination of digestibility in plant materials. *Biotechnology for Biofuels* 3:article number 23.

319. Foster CE, Martin TM, & Pauly M (2010) Comprehensive compositional analysis of plant cell walls (Lignocellulosic biomass) part I: lignin. *Journal of visualized experiments* (37).

320. Fukushima RS & Hatfield RD (2004) Comparison of the acetyl bromide spectrophotometric method with other analytical lignin methods for determining lignin concentration in forage samples. *Journal of Agricultural and Food Chemistry* 52(12):3713-3720.

321. Foster CE, Martin TM, & Pauly M (2010) Comprehensive compositional analysis of plant cell walls (lignocellulosic biomass) part II: carbohydrates. *Journal of visualized experiments* (37).

322. Updegraf DM (1969) Semimicro Determination of Cellulose in Biological Materials. *Analytical Biochemistry* 32(3):420-424.

323. Fry S (1988) *The Growing Plant Cell Wall: Chemical and Metabolic Analysis* (The Blackburn Press, New Jersey).

324. Lapierre C, Pollet B, & Rolando C (1995) New Insights into the Molecular Architecture of Hardwood Lignins by Chemical Degradative Methods. *Research on Chemical Intermediates* 21(3-5):397-412.

325. Larkin MA, *et al.* (2007) Clustal W and clustal X version 2.0. *Bioinformatics* 23(21):2947-2948.

326. Page RDM (1996) TreeView: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* 12(4):357-358.

327. Bradford MM (1976) Rapid and senstive method for quantification of microgram quantities of protein utilizing principle of protein-dye binding. *Analytical Biochemistry* 72(1-2):248-254.

328. Hong S-Y, Seo PJ, Yang M-S, Xiang F, & Park C-M (2008) Exploring valid reference genes for gene expression studies in *Brachypodium distachyon* by real-time PCR. *BMC Plant Biology* 8.

329. Chambers JP, Behpouri A, Bird A, & Ng CKY (2012) Evaluation of the Use of the Polyubiquitin Genes, *Ubi4* and *Ubi10* as Reference Genes for Expression Studies in *Brachypodium distachyon*. *Plos One* 7(11).

330. Gomez LD, Steele-King CG, & McQueen-Mason SJ (2008) Sustainable liquid biofuels from biomass: the writing's on the walls. *New Phytologist* 178(3):473-485.

331. Leu S-Y & Zhu JY (2013) Substrate-Related Factors Affecting Enzymatic Saccharification of Lignocelluloses: Our Recent Understanding. *Bioenergy Research* 6(2):405-415.

332. Grabber JH, Ralph J, Lapierre C, & Barriere Y (2004) Genetic and molecular basis of grass cell-wall degradability. I. Lignin-cell wall matrix interactions. *Comptes Rendus Biologies* 327(5):455-465.

333. Simmons BA, Logue D, & Ralph J (2010) Advances in modifying lignin for enhanced biofuel production. *Current Opinion in Plant Biology* 13(3):313-320.

334. Li X, Weng J-K, & Chapple C (2008) Improvement of biomass through lignin modification. *Plant Journal* 54(4):569-581.

335. Taylor NG (2008) Cellulose biosynthesis and deposition in higher plants. *New Phytologist* 178(2):239-252.

336. Gomez LD, Bristow JK, Statham ER, & McQueen-Mason SJ (2008) Analysis of saccharification in *Brachypodium distachyon* stems under mild conditions of hydrolysis. *Biotechnology for Biofuels* 1(15):article number 15.

337. Kleinhofs A, Owais WM, & Nilan RA (1978) Azide. *Mutation Research* 55(3-4):165-195.

338. Olsen O, Wang XZ, & Von Wettstein D (1993) Sodium-azide mutagenesis - preferential generation of A.T -> G.C transitions in the barley *Ant18* gene. *Proceedings of the National Academy of Sciences of the United States of America* 90(17):8043-8047.

339. Yan L*, et al.* (2013) The heterologous expression in *Arabidopsis thaliana* of sorghum transcription factor *SbbHLH1* downregulates lignin synthesis. *Journal of Experimental Botany* 64(10):3021-3032.

340. Allison GG, Morris C, Clifton-Brown J, Lister SJ, & Donnison IS (2011) Genotypic variation in cell wall composition in a diverse set of 244 accessions of *Miscanthus*. *Biomass & Bioenergy* 35(11):4740-4747.

341. Hu WJ*, et al.* (1999) Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. *Nature Biotechnology* 17(8):808-812.

342. Jouanin L*, et al.* (2000) Lignification in transgenic poplars with extremely reduced caffeic acid O-methyltransferase activity. *Plant Physiology* 123(4):1363-1373.

343. Yi H & Puri VM (2014) Contributions of the mechanical properties of major structural polysaccharides to the stiffness of a cell wall network model. *American Journal of Botany* 101(2):244-254.

344. Matsushika A, Inoue H, Kodaki T, & Sawayama S (2009) Ethanol production from xylose in engineered *Saccharomyces cerevisiae* strains: current state and perspectives. *Applied Microbiology and Biotechnology* 84(1):37-53.

345. Protti-Alvarez F (2014) DSM succeeds in getting improved yields in cellulosic bioethanol fermentation. *Focus on catalysts* (3).

346. Olofsson K, Rudolf A, & Liden G (2008) Designing simultaneous saccharification and fermentation for improved xylose conversion by a recombinant strain of *Saccharomyces cerevisiae*. *Journal of Biotechnology* 134(1-2):112-120.

347. Christensen U, Alonso-Simon A, Scheller HV, Willats WGT, & Harholt J (2010) Characterization of the primary cell walls of seedlings of *Brachypodium distachyon* - A potential model plant for temperate grasses. *Phytochemistry* 71(1):62-69.

348. Zhang B, *et al.* (2009) A missense mutation in the transmembrane domain of CESA4 affects protein abundance in the plasma membrane and results in abnormal cell wall biosynthesis in rice. *Plant Molecular Biology* 71(4-5):509-524.

349. Turner SR & Somerville CR (1997) Collapsed xylem phenotype of Arabidopsis identifies mutants deficient in cellulose deposition in the secondary cell wall. *Plant Cell* 9(5):689-701.

350. Handakumbura PP, *et al.* (2013) Perturbation of *Brachypodium distachyon CELLULOSE SYNTHASE A4* or *7* results in abnormal cell walls. *BMC Plant Biology* 13:article number 131.

351. Huntley SK, Ellis D, Gilbert M, Chapple C, & Mansfield SD (2003) Significant increases in pulping efficiency in C4H-F5H-transformed poplars: Improved chemical savings and reduced environmental toxins. *Journal of Agricultural and Food Chemistry* 51(21):6178-6183.

352. Ciesielski PN, *et al.* (2014) Engineering plant cell walls: tuning lignin monomer composition for deconstructable biofuel feedstocks or resilient biomaterials. *Green Chemistry* 16:2627-2635.

353. Fu C, *et al.* (2011) Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. *Proceedings of the National Academy of Sciences of the United States of America* 108(9):3803-3808.

354. Skyba O, Douglas CJ, & Mansfield SD (2013) Syringyl-Rich Lignin Renders Poplars More Resistant to Degradation by Wood Decay Fungi. *Appl. Environmental Microbiology* 79(8):2560-2571.

355. Tu Y, *et al.* (2010) Functional Analyses of *Caffeic Acid O-Methyltransferase* and *Cinnamoyl-CoA-Reductase* Genes from Perennial Ryegrass (*Lolium perenne*). *Plant Cell* 22(10):3357-3373.

356. Piquemal J, *et al.* (2002) Down-regulation of caffeic acid *O*-methyltransferase in maize revisited using a transgenic approach. *Plant Physiology* 130(4):1675-1685.

357. Pichon M, *et al.* (2006) Variation in lignin and cell wall digestibility in caffeic acid *O*-methyltransferase down-regulated maize half-sib progenies in field experiments. *Molecular Breeding* 18(3):253-261.

358. Chen L, *et al.* (2004) Transgenic down-regulation of caffeic acid *O*-methyltransferase (COMT) led to improved digestibility in tall fescue (*Festuca arundinacea*). *Functional Plant Biology* 31(3):235-245.

359. Stewart JJ, Akiyama T, Chapple C, Ralph J, & Mansfield SD (2009) The Effects on Lignin Structure of Overexpression of Ferulate 5-Hydroxylase in Hybrid Poplar. *Plant Physiology* 150(2):621-635.

360. Lapierre C (2010) Determining lignin structure by chemical degradations. *Lignin and lignans - Advances in Chemistry*, eds Heitner C, Dimmel D, & Schmidt JA (CRC Press, Taylor & Francis Group, Boca Raton, USA), pp 11-48.

361. Lapierre C, Monties B, & Rolando C (1988) Thioacidolyses of diazomethane-methylated pine compression wood and wheat straw *in situ* lignins. *Holzforschung* 42(6):409-411.

362. Sarath G, *et al.* (2011) Ethanol yields and cell wall properties in divergently bred switchgrass genotypes. *Bioresource Technology* 102(20):9579-9585.

363. DeMartini JD, *et al.* (2013) Investigating plant cell wall components that affect biomass recalcitrance in poplar and switchgrass. *Energy & Environmental Science* 6(3):898-909.

364. Silveira RL, Stoyanov SR, Gusarov S, Skaf MS, & Kovalenko A (2013) Plant Biomass Recalcitrance: Effect of Hemicellulose Composition on Nanoscale Forces that Control Cell Wall Strength. *Journal of the American Chemical Society* 135(51):19048-19051.

365. Gómez LD, *et al.* (2014) Side by Side Comparison of Chemical Compounds Generated by Aqueous Pretreatments of Maize Stover, Miscanthus and Sugarcane Bagasse. *BioEnergy Research* epub ahead of print: DOI 10.1007/s12155-014-9480-2.

366. Kumar R, Mago G, Balan V, & Wyman CE (2009) Physical and chemical characterizations of corn stover and poplar solids resulting from leading pretreatment technologies. *Bioresource Technology* 100(17):3948-3962.

367. Silverstein RA, Chen Y, Sharma-Shivappa RR, Boyette MD, & Osborne J (2007) A comparison of chemical pretreatment methods for improving saccharification of cotton stalks. *Bioresource Technology* 98(16):3000-3011.

368. Ponte Rocha MV, Soares Rodrigues TH, de Macedo GR, & Goncalves LRB (2009) Enzymatic Hydrolysis and Fermentation of Pretreated Cashew Apple Bagasse with Alkali and Diluted Sulfuric Acid for Bioethanol Production. *Applied Biochemistry and Biotechnology* 155(1-3):407-417.

369. Lau MW, Gunawan C, & Dale BE (2009) The impacts of pretreatment on the fermentability of pretreated lignocellulosic biomass: a comparative evaluation

between ammonia fiber expansion and dilute acid pretreatment. *Biotechnology for Biofuels* 2:article number 30.

370. d'Yvoire MB, *et al.* (2013) Disrupting the *cinnamyl alcohol dehydrogenase 1* gene (*BdCAD1*) leads to altered lignification and improved saccharification in *Brachypodium distachyon*. *Plant Journal* 73(3):496-508.

371. Timpano H, *et al.* (2014) *Brachypodium* cell wall mutant with enhanced saccharification potential despite increased lignin content. *Bioenergy Research* Epub ahead of print, DOI 10.1007/s12155-014-9501-1.

372. Ng PC & Henikoff S (2001) Predicting deleterious amino acid substitutions. *Genome Research* 11(5):863-874.

373. Saxena IM, Brown RM, & Dandekar T (2001) Structure-function characterization of cellulose synthase: relationship to other glycosyltransferases. *Phytochemistry* 57(7):1135-1148.

374. Zhong RQ, Morrison WH, Freshour GD, Hahn MG, & Ye ZH (2003) Expression of a mutant form of cellulose synthase AtCesA7 causes dominant negative effect on cellulose biosynthesis. *Plant Physiology* 132(2):786-795.

375. Vanholme B, *et al.* (2013) Breeding with rare defective alleles (BRDA): a natural *Populus nigra* HCT mutant with modified lignin as a case study. *New Phytologist* 198(3):765-776.

376. Besseau S, *et al.* (2007) Flavonoid accumulation in *Arabidopsis* repressed in lignin synthesis affects auxin transport and plant growth. *Plant Cell* 19(1):148-162.

377. Kim IA, Kim B-G, Kim M, & Ahn J-H (2012) Characterization of hydroxycinnamoyltransferase from rice and its application for biological synthesis of hydroxycinnamoyl glycerols. *Phytochemistry* 76:25-31.

378. Shi R, *et al.* (2010) Towards a Systems Approach for Lignin Biosynthesis in *Populus trichocarpa*: Transcript Abundance and Specificity of the Monolignol Biosynthetic Genes. *Plant and Cell Physiology* 51(1):144-163.

379. Hoffmann L, Maury S, Martz F, Geoffroy P, & Legrand M (2003) Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. *Journal of Biological Chemistry* 278(1):95-103.

380. Kim B-G, Kim I-A, & Ahn J-H (2011) Characterization of Hydroxycinnamoyl-coenzyme A Shikimate Hydroxycinnamoyltransferase from *Populus euramericana*. *Journal of the Korean Society for Applied Biological Chemistry* 54(5):817-821.

381. Guillaumie S, *et al.* (2007) MAIZEWALL. Database and developmental gene expression profiling of cell wall biosynthesis and assembly in maize. *Plant Physiology* 143(1):339-363.

382. Schoch G, *et al.* (2001) CYP98A3 from *Arabidopsis thaliana* is a 3'-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. *Journal of Biolical Chemistry* 276(39):36566-36574.

383. Franke R, *et al.* (2002) Changes in secondary metabolism and deposition of an unusual lignin in the *ref8* mutant of Arabidopsis. *Plant Journal* 30(1):47-59.

384. Chen F, *et al.* (2006) Multi-site genetic modulation of monolignol biosynthesis suggests new routes for formation of syringyl lignin and wall-bound ferulic acid in alfalfa (*Medicago sativa* L.). *Plant Journal* 48(1):113-124.

385. Nair RB, Bastress KL, Ruegger MO, Denault JW, & Chapple C (2004) The *Arabidopsis thaliana REDUCED EPIDERMAL FLUORESCENCE1* gene encodes an aldehyde dehydrogenase involved in ferulic acid and sinapic acid biosynthesis. *Plant Cell* 16(2):544-554.

386. Derikvand MM, *et al.* (2008) Redirection of the phenylpropanoid pathway to feruloyl malate in *Arabidopsis* mutants deficient for cinnamoyl-CoA reductase 1. *Planta* 227(5):943-956.

387. Dauwe R, *et al.* (2007) Molecular phenotyping of lignin-modified tobacco reveals associated changes in cell-wall metabolism, primary metabolism, stress metabolism and photorespiration. *Plant Journal* 52(2):263-285.

388. Ehlting J, *et al.* (1999) Three 4-coumarate:coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *Plant Journal* 19(1):9-20.

389. Michelmore RW, Paran I, & Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis - a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences of the United States of America* 88(21):9828-9832.

390. Schneeberger K, *et al.* (2009) SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nature Methods* 6(8):550-551.

391. Zuryn S, Le Gras S, Jamet K, & Jarriault S (2010) A Strategy for Direct Mapping and Identification of Mutations by Whole-Genome Sequencing. *Genetics* 186(1):427-430.

392. Abe A, *et al.* (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. *Nature Biotechnology* 30(2):174-178.

393. Allen RS, Nakasugi K, Doran RL, Millar AA, & Waterhouse PM (2013) Facile mutant identification via a single parental backcross method and application of whole genome sequencing based mapping pipelines. *Frontiers in Plant Science* 4:article number 362.

394. James GV, *et al.* (2013) User guide for mapping-by-sequencing in *Arabidopsis*. *Genome Biology* 14(6):article number R61.

395. Jander G, *et al.* (2002) Arabidopsis map-based cloning in the post-genome era. *Plant Physiology* 129(2):440-450.

396. Aubourg S, Kreis M, & Lecharny A (1999) The DEAD box RNA helicase family in *Arabidopsis thaliana*. *Nucleic Acids Research* 27(2):628-636.

397. Stevenson RJ, Hamilton SJ, MacCallum DE, Hall PA, & Fuller-Pace FV (1998) Expression of the 'DEAD box' RNA helicase p68 is developmentally and growth regulated and correlates with organ differentiation/maturation in the fetus. *The Journal of Pathology* 184(4):351-359.

398. Schmid SR & Linder P (1992) D-E-A-D protein family of putative RNA helicases. *Molecular Microbiology* 6(3):283-292.

399. Herrero J, *et al.* (2013) Bioinformatic and functional characterization of the basic peroxidase 72 from *Arabidopsis thaliana* involved in lignin biosynthesis. *Planta* 237(6):1599-1612.

400. Gabaldon C, Lopez-Serrano M, Pedreno MA, & Barcelo AR (2005) Cloning and molecular characterization of the basic peroxidase isoenzyme from *Zinnia elegans*, an enzyme involved in lignin biosynthesis. *Plant Physiology* 139(3):1138-1154.

401. Gabaldon C, *et al.* (2006) Characterization of the last step of lignin biosynthesis in *Zinnia elegans* suspension cell cultures. *FEBS Letters* 580(18):4311-4316.

402. Lindedam J, *et al.* (2014) Evaluation of high throughput screening methods in picking up differences between cultivars of lignocellulosic biomass for ethanol production. *Biomass & Bioenergy* 66:261-267.