Using QTL analysis of *Brachypodium distachyon* to understand the genetic basis of grass cell wall saccharification

Caragh Bryony Whitehead

PhD

University of York

Biology

2016

Abstract

Second generation biofuels are seen as a sustainable solution to the problem of dwindling fossil fuels stocks. However, the process of converting lignocellulosic biomass to sugars for fermentation is expensive due the recalcitrance of these materials to enzymatic digestion. The identification of quantitative trait loci (QTL) in the model grass species *Brachypodium distachyon* was undertaken in order to improve our understanding of genes that affect straw digestibility. Initially, the study focused on the analysis of natural accessions to determine if there was variation, in terms of digestibility and cell wall composition, within the species and to identify lines suitable for producing recombinant inbred lines (RILs). This information was successfully used to initiate the production of a RIL population that can be used in future research. I made use of a pre-existing RIL population produced previously from a bi-parental cross between Bd21 and Bd3.1 to study pathogen resistance. This RIL population was screened for straw digestibility using a semi-automated robotic platform. This data together with the genotype data was used to identify QTL linked to digestibility. A single QTL was detected on chromosome 5 together with a further QTL on chromosome 3 that acted in epistasis. A candidate gene for each of the QTLs was identified by reviewing those located within the QTL regions. The chromosome 5 candidate gene encodes a glycosyl hydrolase family 43 family protein likely involved in xylan biosynthesis and the chromosome 3 candidate gene is a cellulose synthase-like subfamily A protein that has a possible glucomannan 4-beta-mannosyltransferase function. Functionality was analysed by studying the cell wall composition of selected RILs and corresponding *Arabidopsis thaliana* T-DNA lines to determine any differences in the secondary cell wall structure. The results indicated that the differences in digestibility are associated with subtle differences in cell wall composition.

Table of Contents

List of Tables

List of Figures

Acknowledgements

Firstly I would like to thank my supervisor, Simon McQueen-Mason for all his support, advice and guidance. I would also like to thank Richard Waites and Frans Maathuis from my TAP committee members for their helpful suggestions throughout the project. Thank you to the Burgess family who supplied the funding for my PhD and the University of York with providing all the facilities.

I would also like to thank David Garvin, University of Minnesota for providing the *Brachypodium distachyon* RIL populatation Bd21 x Bd3-1, without it none of this research could have been conducted.

I'm grateful for the help I received from various people at INRA, France. Firstly, Richard Sibout and Sylvain Legay for suppling plant material when I need it. Secondly, Matthieu Reymond for providing QTL analysis training and Paul Robert for supplying guidance when analysing ATR-FTIR data.

There are a number of people I wish to thank from the University of York. Firstly, Leonardo Gomez and Rachael Hallam for all their guidance in the laboratory related to the various cell wall composition techniques. Secondly, Yi Li and Zhesi He with regards to all the support they provided in terms of bioinformatics. Tengyo Jiang for running my XRD samples at Green Chemistry. Finally, thanks to all the support staff in the horticulture team that kept my plants growing as well as the laboratory support on M2 who made sure all the consumables were available when I needed them.

Many thanks to all the members of the McQueen-Mason team who have made every day on M2 a wonderful place to work. You have all supported and motivated me to keep on going.

Last but not least I thank my husband, Vincent for providing the support I needed in so many different ways. Without you I would not have been able to do this. Thank you for always believing in me. Finally, I would like to dedicate this to my son, Erik, a little bonus that came into my life during the PhD.

Author's Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Chapter 1: Introduction

1.1 Biofuels

1.1.1 Fossil fuels

Fossil fuels are the result of compression and heating of organic matter, such as plant material and animal remains, over a large geological time scale. This organic matter is reduced to a range of hydrocarbons found as gases, liquids and solids. There are a number of types of fossil fuels including; tar sand, oil shale, petroleum, coal, bitumen and natural gas. Petroleum, also known as crude oil, is a thick greenish or dark brown liquid composed of alkane and aromatic compounds. In 2005 it was reported that petroleum was the largest energy source consumed (Demirbas 2009).

Oil refineries are used to clean and separate petroleum into various products using a heating process known as distillation. These products include transportation fuels such as gasoline, diesel and jet fuel as well as raw products that are used for the production of petrochemicals, such as olefins and aromatics. The price of the fuel produced is dependent of the grade of the fuel, the location of the refinery, the demand as well as the current supply available and the perceived supply of fuel in the future (Demirbas 2009, Naik *et al* 2010).

At present, petroleum is used to produce 98% of all transportation fuels. However, the unsustainable nature of fossil fuels requires the need to look for other possible alternatives, one of which is biofuels (Gomez *et al* 2008, Koh *et al* 2008, Naik *et al* 2010). This unsustainability is the result of fossil fuels being non-renewable as well as an ever increasing demand for petroleum, which is being driven by developing countries like China and India (Figure 1.1). This demand cannot be met indefinitely as there are limited reserves available. Some of the small oil reserves are nearly depleted and it has been forecast that the larger reserves could be depleted within the next 50 years (Demirbas 2009, US Energy Information Administration).
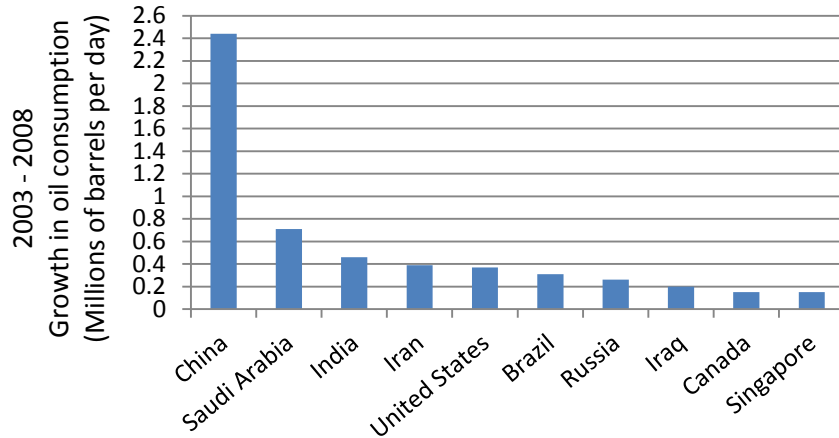
Figure 1.1 Growth in oil consumption, in terms of millions of barrels used per day, from 2003 to 2008 (data from the US Energy Information Administration).

Energy security is also a large driving force to switching to other forms of transportation fuels. The oil reserves are not evenly distributed around the world and a large percentage are found in politically unstable regions. It has been reported that 64% of the reserves are located in the Middle East and a number of these are found in countries that are ranked in the top 10 countries containing the largest oil reserves (Figure 1.2). This top 10 list includes countries, such as Iraq and Kuwait that have been involved in a recent conflict, the first Gulf War, which led to a global increase in oil prices (Koe *et al* 2008, Hoekman 2009, US Energy Information Administration).
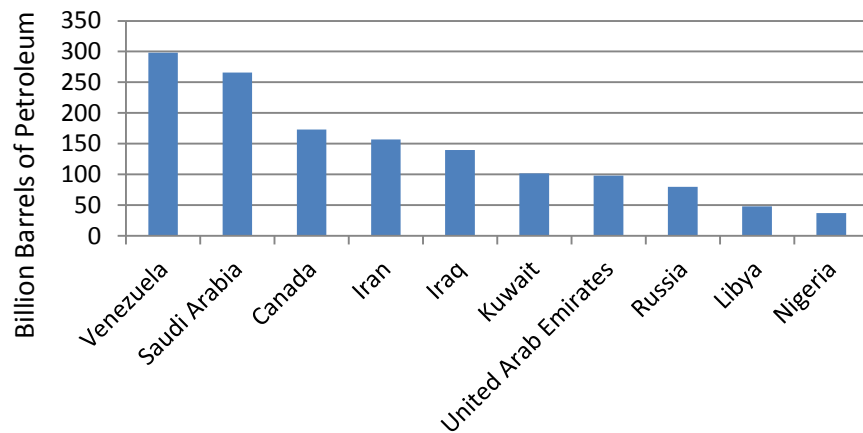
Figure 1.2 Top 10 holders of petroleum reserves, January 2014 (data from the US Energy Information Administration).

Another important issue related to the use of fossil fuels is that they have a negative impact on the environment because they release carbon dioxide (CO_2) into the atmosphere when they are burned. This then adds to the atmospheric load of CO_2 and causes an increase in global warming, which exacerbates climate change. Approximately, 30% of all CO_2 emissions from industrialised countries are due to the combustion of liquid transportation fuels (Gomez *et al* 2008, Carroll *et al* 2009). Worldwide there was an overall increase in CO_2 emissions of 0.5% between 2013 and 2014 (Figure 1.3). This increase was driven by developing countries like India who had an increase in CO_2 emissions of 8.1%, whereas the European Union had an overall decrease of 3.8% (BP Statistical Review of World Energy 2015).
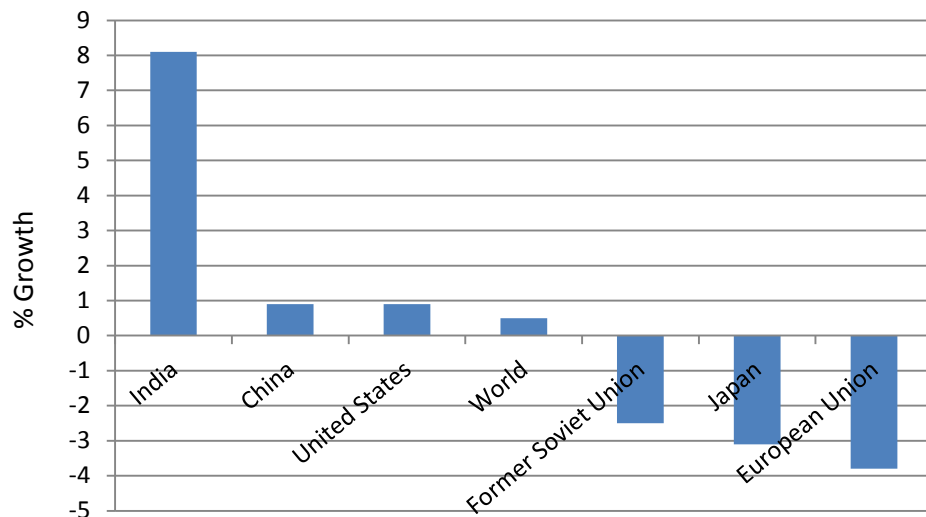
Figure 1.3 Percentage growth or decline in CO_2 emissions from energy use, 2013 to 2014 (data from BP Statistical Review of World Energy 2015).

1.1.2 First generation biofuels

Currently, first generation biofuels are produced from starch, sucrose and plant oil crops by fermentation or transesterification (Carroll *et al* 2009, Naik *et al* 2010). This has been successfully achieved in Brazil with sugarcane, in the USA with corn and soybean and in Europe with wheat, sugar beet and oil seed rape. Bioethanol is the most widely used liquid biofuel with 60% being produced from sugarcane and 40% form other crops. It is predominantly used in light-duty vehicles as a blend of

10% together with gasoline, otherwise known as E10. Some countries such as Sweden and India incorporate it as a 5% blend (E5) whereas Brazil has flexi-fuel vehicles that can run on 85% bioethanol blends (Koh *et al* 2008, Demirbas 2009).

The main drawback of first generation biofuels is that they are not sustainable as they compete for feedstocks that also serve as food, either directly or indirectly, for human consumption. This leads to increased pressure on global food security at a time of population expansion and increased consumption rates (Koh *et al* 2008, Carroll *et a*l 2009, Vega-Sanchez *et al* 2010). It is estimated that by 2050 there will be 9.5 billion people in the world with the least developed countries doubling their population to 1.8 billion by 2050. Not only will this increase in population size put pressure on food security but also the fact that people's average kilo calorie intake per day is also expected to continue to increase (Figure 1.4). To keep up with this increase in demand it has been calculated that agricultural production will need to increase by 60% by 2050. This means that the production of cereals would have to increase by 940 million tonnes. Using current increases in crop yields it can be determined that current growth rate of maize, rice, wheat and soybean yields would fall short of this 60% (CCAFS-CGIAR).
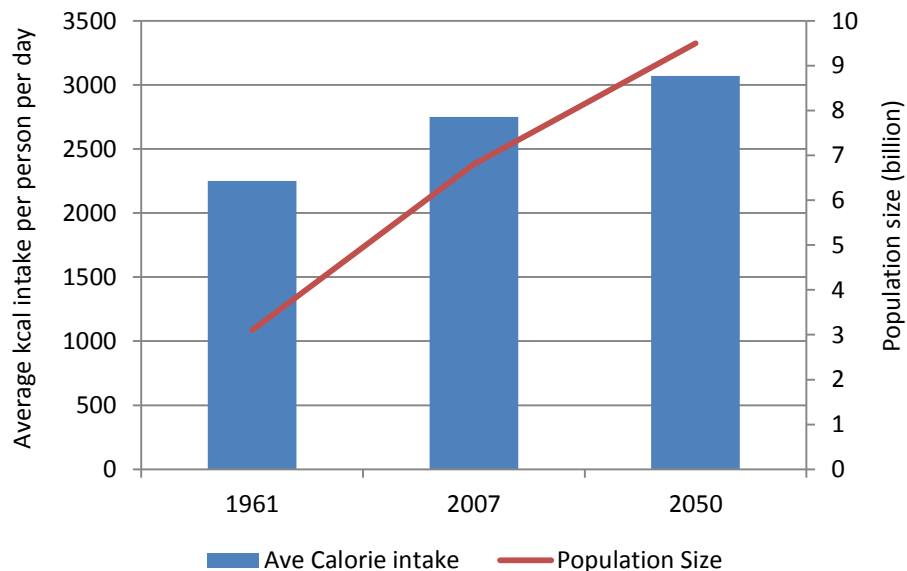
Figure 1.4 Estimated population growth and increase in the consumption of kilo calories per person per day (data supplied by CCAFS-CGIAR).

Another factor that contributes to the unsuitability of first generation biofuels as a fossil fuel replacement is that these crops require large inputs of fertilisers and pesticides for growth. These inputs themselves are known to have a considerable carbon footprint which then weakens the carbon mitigation effectiveness of such fuels (Gomez *et al* 2008).

1.1.3 Second generation biofuels

Focus has now shifted to second generation biofuels as a source of energy for transportation due to the concerns raised related to first generation biofuels. Second generation biofuel is produced from woody, non-food lignocellulosic materials that can be found as agricultural bi-products or can be from dedicated biomass crops grown on unused or marginal land (Hoekman 2009, Naik *et al* 2010). The main advantage of this form of biofuel is that the crops require lower inputs for growth than food crops thereby making them more sustainable. It is also possible that they could provide an additional income stream to farmers by providing new revenue (Gomez *et al* 2008, Carroll *et al* 2009). The disadvantage of lignocellulosic biofuels is that there are a number of technical hurdles that need to be overcome before the conversion technologies can produce a product that is sufficiently cost effective to compete directly with first generation biofuels and fossil fuels (Gomez *et al* 2008, Hoekman 2009, Naik *et al* 2010). The price of second generation biofuels depends on operating costs resulting from the pretreatment and enzymes used during production, distribution costs of delivering the feedstock to the biorefinery and distributing the bioethanol as well as the cost of feedstocks (Demirbas 2009).

There are two main methods for the production of biofuels from lignocellulosic material, namely thermochemical and biochemical conversion (Figure 1.5). The thermochemical method involves heating the biomass at high temperatures in the absence or partial absence of oxygen to generate pyrolysis gases, liquids and solids (Koh *et al* 2008, Carroll *et al* 2009, Hoekman 2009, Naik *et al* 2010). This method has the advantage that it is less dependent on the composition of the biomass. However, it has very high capital costs associated, and also requires extremely large volumes of biomass to operate effectively (Gomez *et al* 2008).
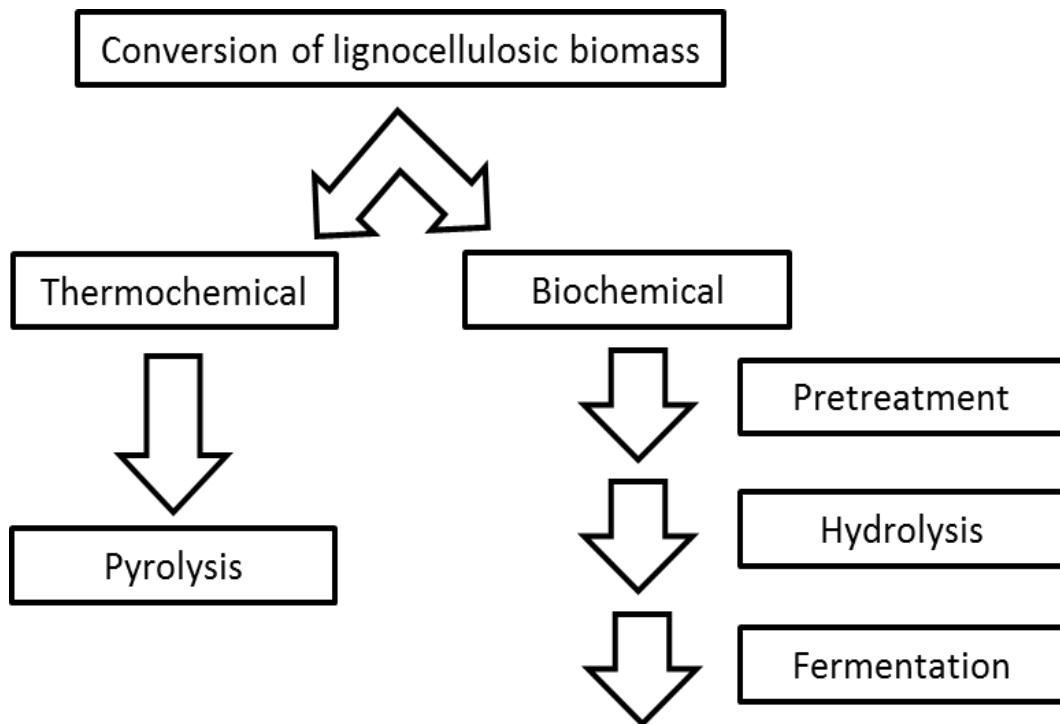
Figure 1.5 Methods of converting lignocellulosic material into biofuels.

The biochemical method involves the conversion of lignocellulosic material to fermentable sugars (Foust *et al* 2008, Koh *et al* 2008). This method involves the following steps; pretreatment of the biomass, enzyme hydrolysis to release fermentable sugars and fermentation to produce alcohols, such as ethanol (Ding *et al* 2008, Carroll *et al* 2009).

The advantage of using biochemical conversion methods is that it can be set up on a smaller scale than thermochemical processing (Gomez *et al* 2008, Carroll *et al* 2009). However, it is currently not cost effective because of the large amount of energy, chemicals and expensive enzymes that are needed to release the fermentable sugars from the lignocellulosic biomass (Carroll *et al* 2009). Currently, a large percentage of the cost of production is spent on firstly pretreating the biomass with stringent thermal and/or chemical methods and secondly the enzymatic hydrolysis which involves a relatively large amount of expensive commercial enzymes (Gomez *et al* 2008, Shishir *et al* 2008, Carroll *et al* 2009).

1.1.4 Lignocellulosic feedstocks

Lignocellulosic biomass includes all biodegradable organic plant material, which is able to store sunlight as energy, making it the most abundant energy resource available (Demibras 2009, Naik *et al* 2010). There are a number of different categories that lignocellulosic feedstocks can be grouped into, namely:

- Forestry products: logging residues, wood chips, sawdust.
- Agricultural waste: crop residues, bagasse.
- Energy crops: grasses, short rotation and herbaceous woody crops.
- Aquatic plants: reeds and rushes.
- Municipal waste: landfill, urban wood and organic waste.

Currently, these feedstocks have limited applications or are considered waste products. Some applications include producing compost from municipal green waste, or wood pellets and briquettes from forestry residues or animal feed and bedding from agricultural straw residues (Di Maio *et al* 2014).

In the UK there are approximately 16 million tons of possible feedstock available in terms of forestry products, agricultural residues, energy crops as well as green and paper industry waste. The main component of this feedstock is green waste. Another major contributor is the 25% of crop residues that are currently ploughed back into the fields. Availability of the various feedstocks differs by region, however it has been determined that the Southern and Eastern areas of England have the most biomass available (Di Maio *et al* 2014).

It is thought that perennial crops such as short rotation coppices and grasses provide the best source of lignocellulosic biomass as they produce high yields of material at low cost and have a low impact on the environment (Demibras 2009). Short rotation coppices are energy crops that have been planted especially to produce biomass to be used for the generation of heat and power or potentially for the production of lignocellulosic biofuel. In the UK the main energy crops are Miscanthus, willow and poplar. These plants are currently only being grown on a small area of land with

Miscanthus being harvested yearly whereas poplar and willow are only harvested every three years. The majority of the agricultural waste residue in the UK is from wheat, barley, oilseed rape and a small amount of oats. Currently barley and oat straw is used in animal feed to add roughage whereas wheat straw is used as animal bedding. A large percentage of wheat straw is ploughed back into the fields because of the cost of haulage. Therefore, currently oilseed rape and wheat straw are an underutilised resource (Di Maio *et al* 2014).

A challenge of using lignocellulosic biomass to produce biofuel is that it is a heterogeneous substrate whose cell wall concentration as well as composition affects digestibility. This heterogeneity is dependent on a number of factors including plant species and genotype, tissue and cell type as well as the developmental stage of the plant, cell or tissue (Slavov *et al* 2013, Loqué *et al* 2015). This feature of the material needs to be accounted for during processing, which means a better understanding of the makeup of the material is needed. This includes particle size, ash and moisture content, structural components in terms of cellulose, hemicellulose and lignin as well as its elemental constituents such as carbon, hydrogen, oxygen and nitrogen (Demibras 2009). There are a number of other factors influencing the use of energy crops in the production of lignocellulosic bioethanol in the UK namely; convincing farmers to grow the crops, the capacity for planting and harvesting as well as the compatibility with current technology. In the case of agricultural waste straw the industry will have to compete with the current practice of ploughing the straw back into the soil as a method of improving its organic matter content as well as overcome the time limit for collection of straw from the field   (Di Maio *et al* 2014).

1.2 Secondary plant cell walls

The resistance of lignocellulosic material to digestion is the main technical hurdle to overcome to make second generation biofuels a cost-competitive replacement for fossil fuels. This is predominantly due to the structure of the material itself, which has evolved to be strong and flexible as well as to protect itself against infection from pathogens (Gomez *et al* 2008, Himmel *et al* 2008). Lignocellulosic plant

biomass mostly comprises of secondary plant cell walls (Figure 1.6). The cell wall is a complex structure made predominantly of cellulose, hemicellulose and lignin. The majority of the cell wall consists of cellulose that is found as microfibrils. These cellulose microfibrils are covered by hemicelluloses, which together with the cellulose forms a matrix of polysaccharides. Finally, this matrix is covered and impregnated with lignin, which is a complex polyphenolic compound (Gomez *et al* 2008).
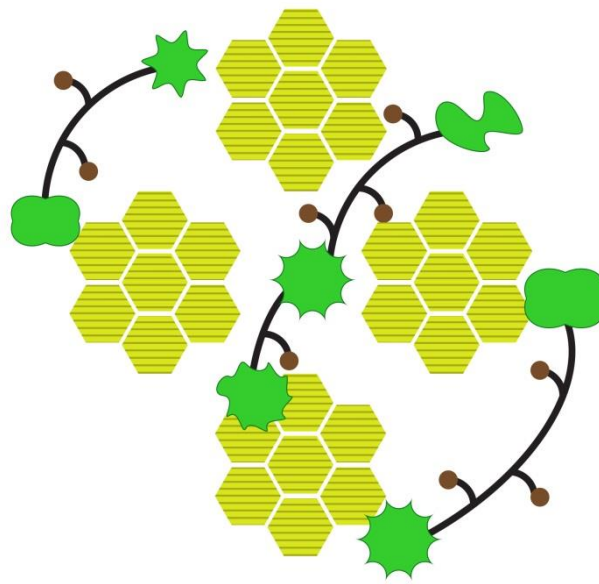
Figure 1.6 Structure of the secondary plant cell wall showing the cellulose microfibrils (yellow), hemicellulose (black) with sidechains (brown), and lignin (green).

The components of the cell wall can be found in differing quantities depending on the plant source (Table 1.1). In dicots the main fraction is made up of cellulose, followed by hemicellulose and finally lignin. It appears that grasses on average have a higher percentage of hemicellulose as well as more lignin, which means there is less cellulose than that found in dicots (Marriott *et al* 2015).

Table 1.1 The typical percentage of cell wall components cellulose, hemicellulose and lignin found within different plant sources (Marriott *et al* 2015).

	Cellulose (%)	Hemicellulose (%)	Lignin (%)
Dicots	45 - 50	20 - 30	7 - 10
Grasses	35 - 45	40 - 50	20
Softwood	25 - 50	20 - 30	25 - 35
Hardwood	40 - 55	20 - 35	18 - 25

1.2.1 Cellulose

Cellulose microfibrils are long crystalline assemblies of β-1,4 glucans composed of repeating cellobiose units, which give the cell wall its inherent strength (Figure 1.7). The microfibrils of primary cell walls have been shown to be approximately 3 nm in diameter and consist of around 36 β-1,4 glucan chains that lie in parallel to each other (Lerouxel *et al* 2006). Cellulose is thus a potentially rich source of glucose, the preferred sugar for fermentation (Gomez *et al* 2008). However, the cellulose microfibrils are insoluble in most solvents and enzymes struggle to access the β-1,4 glycosidic linkages to break it down to single units of glucose (Demibras 2009).
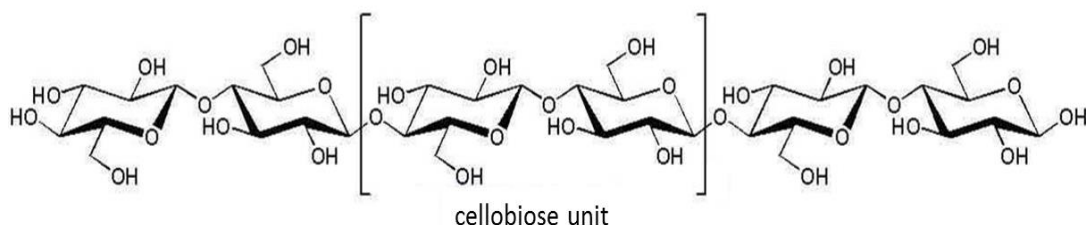
cellobiose unit

Figure 1.7 Structure of cellulose found within the plant cell wall indicating the repeating cellobiose units.

Previous studies in *Arabidopsis thaliana* (Arabidopsis) have shown that cellulose microfibrils are synthesised by cellulose synthase complexes within the plasma membrane (Leroxel *et al* 2006). The complex contains multiple CesA protein subunits encoded by 10 *CesA* genes (Persson *et al* 2005, Leroxel *et al* 2006). These assembled subunits are known as rosettes and are made up of six hexagonal units, each probably containing six catalytic subunits from three different isoforms.

Different CesA proteins are active in the synthesis of either the primary or secondary cell wall. The rosettes are thought to be involved in linking glucose monomers by β-1,4 glycosidic bonds to form β-1,4-glucan chains, which join together into linear chains outside of the plasma membrane (Leroxel *et al* 2006, Ding *et al* 2008). However, regions of reduced crystallinity have been identified and are assumed to play a role in providing a structure for hemicellulose to bind too (Gomez *et al* 2008).

1.2.2 Hemicellulose

Hemicellulose consists of glucose or glucose-like sugars that are β-1,4 linked. However, it differs from cellulose in that it is shorter than cellulose and it also contains side-branches that prevent it from forming a crystalline structure (Gomez *et al* 2008). It is believed that the production of non-cellulosic polysaccharides involves the endoplasmic reticulum and Golgi. The hemicellulose is then secreted within vesicles and travels to the plasma membrane where it is polymerised and integrated with the cellulose microfibrils (Carpita 2012). The hemicellulose fraction of the cell wall represents a relatively large source of sugars for fermentation. Alkaline pretreatments are able to partially solubilise hemicellulose thereby making it accessible to hydrolytic enzymes which release hexoses such as glucose and mannose as well as pentoses such as xylose and arabinose. However, current commercial yeast strains are not able to ferment all the released sugars (Demibras 2009).

Figure 1.8 shows the different types of hemicellulose that can be classed according to their structure into the following groups: xyloglucans, xylans, mannans or glucomannans, and β-(1-3,1-4)-glucans (Scheller *et al* 2010). The composition of hemicellulose differs depending whether it is found in the primary or secondary cell wall. In the case of Arabidopsis primary cell walls it contains predominantly xyloglucans, whereas secondary cell walls are made up of glucuronoxylan (Goubet *et al* 2009). The composition also differs depending on the plant species and environmental conditions (Scheller *et al* 2010, Sills *et al* 2012). For example, angiosperms contain mainly xylans and arabinoxylans within their secondary cell walls. These polymers have a backbone of β-1,4 xylan and may be decorated with

glucoronic acid side chains. On the other hand, grass arabinoxylans often have ferulic acid residues as side chains. The various side chains are thought to form crosslinks with lignin (Gomez *et al* 2008).

Xyloglucans are found in all land plant species. It is the most abundant primary cell wall hemicellulose in most plants except in grasses. The xyloglucan structure can vary but all have a β-1,4 glucan backbone decorated with xylosyl side chains (Figure 1.8A). The side chains can differ in the patterns of substitutions as well as in further decorations. A number of different glycosyltransferases (GT) have been identified as playing a role in the biosynthesis of xyloglucans. It is thought that hydrolases may also be involved but to date none have been identified (Scheller *et al* 2010). Previous work in Arabidopsis has determined that cellulose synthase like (*Csl*) genes, which form a superfamily of genes, play a role in the synthesis of different hemicellulose backbones. It appears that each gene in this family plays a specific role. In the case of xyloglucan backbone synthesis it is thought that *CslC4* is responsible (Yin *et al* 2009, Scheller *et al* 2010). It has been hypothesised that the CslC4 protein together with other proteins forms a complex to achieve this (Oikawa *et al* 2013).
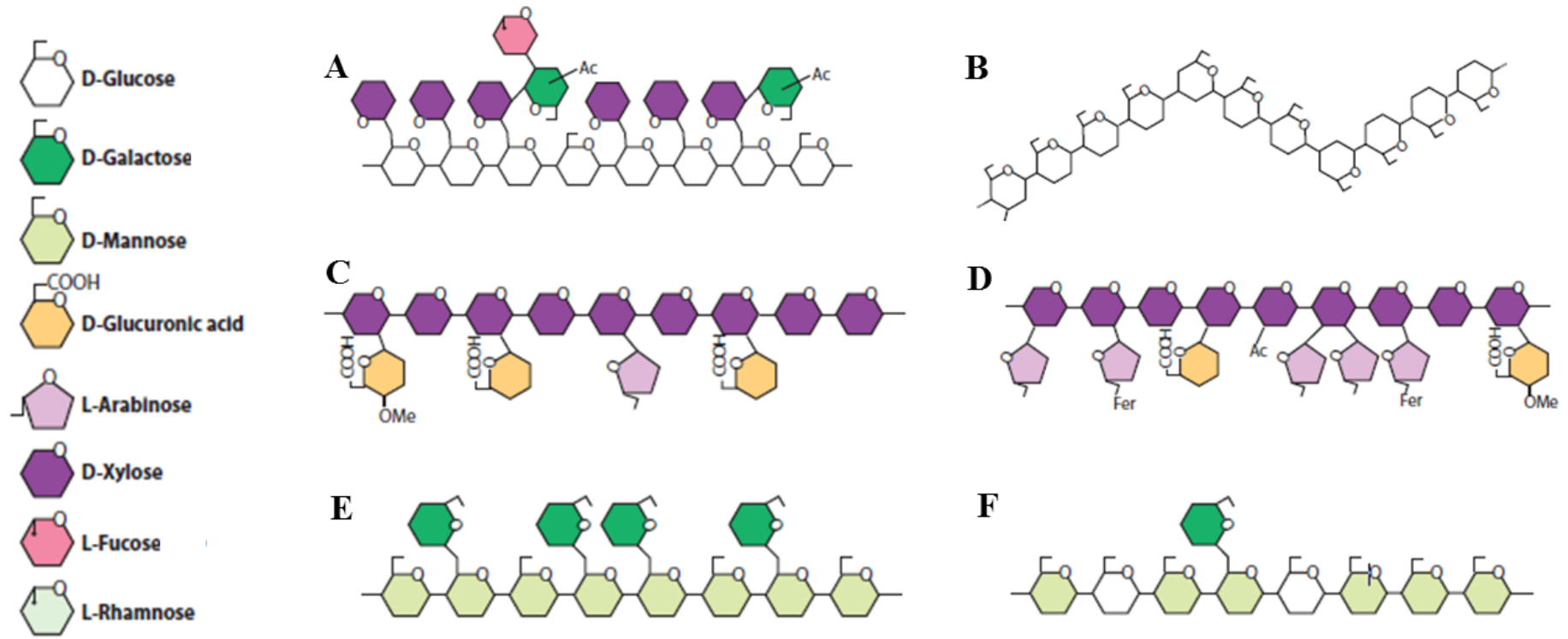
Figure 1.8 The structure of various types of hemicellulose including (A) xyloglucan, (B) β-(1-3,1-4)glucan, (C) glucuronoxylan, (D) glucuronoarabinoxylan, (E) galactomannan and (F) galactoglucomannan (adapted from Scheller *et al* 2010).

Mannans and galactomannans have β-(1,4)-linked backbones consisting of only mannose, whereas glucomannans and galactoglucomannans have a backbone of mannose and glucose units in a non-repeating pattern (Figure 1.8E and F). It is thought that mannans have a diverse functionality because various types of mannans have been found in the majority of Arabidopsis cell types. It has been shown that glucomannans in the cell wall play a structural role whereas glucomannans found in the embryo are involved in embryo development. In some cases of palms and legumes mannans have been shown to function as a carbohydrate storage reserve (Goubet *et al* 2009). Previous studies have indicated that various *CslA* genes are involved in the biosynthesis of mannan and glucomannan (Scheller *et al* 2010, Carpita 2012, Dhugga 2012). In Arabidopsis primary and secondary cell walls of inflorescent stems it has been reported that the CslA9 enzyme is the main producer whereas CslA2 and 3 only produce a small amount of glucomannan. However, all three enzymes are needed for successful synthesis (Goubert *et al* 2009). It has also been shown that CslD has a glucomannan synthase activity with CslD2 and CslD3 functioning together in a complex (Oikawa *et al* 2013).

The β-(1-3,1-4)-linked glucans, also known as mixed linkage glucans, are often found in grasses but have not been observed in dicots. Their backbone is predominantly made up of cellotriosyl and cellotetrasyl units (Figure 1.8B). The biosynthesis of mixed linkage glucans has been reported to involve the *Csl* gene subfamilies *CslF* and *CslH* (Carpita 2012, Dhugga 2012). Both of these genes have not been shown to be present in Arabidopsis, therefore supporting the theory that mixed linkage glucans are absent in dicots. It also doesn't appear that the CslF and CslH proteins work together as a complex as they don't both need to be present for biosynthesis to take place (Scheller *et al* 2010).

Xylans are the major hemicellulosic component found in many plant secondary cell walls and are characterised by a backbone made up of β-(1,4)-linked xylosyl residues (Figure 1.8C and D). A modification, known as glucuronoxylans, is often seen in dicot secondary cell walls involving the substitution of α-(1,2)-linked glucuronosyl and 4-*O*-methyl glucuronosyl residues (Chiniquy *et al* 2013). Xylan is the key component of both primary and secondary cell walls of grasses but it normally

contains arabinosyl sidechains and therefore is known as arabinoxylan or glucuronoarabinoxylan. An important characteristic of grass xylans is the presence of ferulic acid esters attached to some of the arabinofuranosyl residues. These ferulic acid esters are able to dimerise forming covalent attachments between neighbouring xylan chains and also forming a link between hemicellulose and lignin, thereby making the cell wall more indigestible (Scheller *et al* 2010). It was originally thought that the biosynthesis of the xylan backbone would involve *Csl* genes due to the structure of the backbone being similar to other β-1,4 glucans. However, this is not the case as a number of GTs (GT8, 43 and 47) has been shown to carry out this function (Carpita 2012).

1.2.3 Lignin

Lignin has been studied in various plants such as Arabidopsis, maize and alfalfa (Chen *et al* 2007, Li *et al* 2010, Brenner *et al* 2012). It has been shown to have a complex polyphenolic structure consisting of three monomers namely, *p*-hydroxyphenyl, syringyl and guaiacyl (H, S and G) that can be bound together in a number of possible patterns one of which is presented in Figure 1.9 (Gomez *et al* 2008, Li *et al* 2010).

The monomers are formed via the phenylpropanoid pathway from *p*-coumaryl, sinapyl and coniferyl alcohols respectively and are linked by β-O-4 ether bonds (Figure 1.10). The phenylpropanoid pathway begins with the amino acid phenylalanine, which is transformed to *p*-coumaric acid by the removal of an ammonia molecule by phenylalanine ammonia lyase (PAL). A number of steps involving various enzymes then takes place until cinnamyl alcohol dehydrogenase (CAD) converts several substrates into the specific monolignols (Brenner *et al* 2010). Some of these enzymes such as C4H, C3H and F5H, are membrane bound in the endoplasmic reticulum and are thought to be active on the cytosolic side. It is unclear how the monolignols are transported to the secondary cell wall were they are polymerized by oxidative radicalization of phenols (Vanholme *et al* 2010).

27
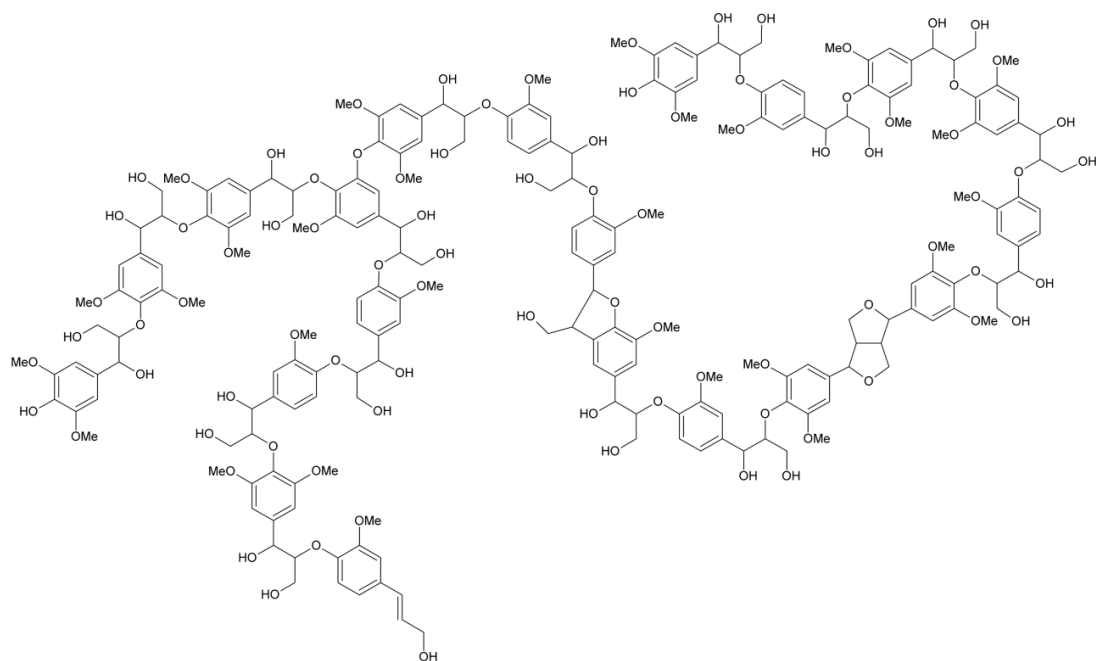
Figure 1.9 A hypothetical structure of lignin found within plant cell walls (Albersheim *et al* 2011).

In previous research it has been shown that the quantity of the various monomers differs between plant species. It appears that maize containing mainly G and S units whereas dicots tend to have five times more H units than G and S (Barrière *et al* 2008, Brenner *et al* 2012). Research has also indicated that not only the amount of lignin present in the cell wall can affect the rate of digestibility but also the ratio of the monomers present in the polymer (Chen *et al* 2007, Brenner *et al* 2010, Li *et al* 2010). However, studies have also indicated that mutants of the genes involved in the phenylpropanoid pathway can lead to undesirable traits together with increased digestibility. This has been seen with the maize brown-midrib (*bm*) mutants, which show an improvement in digestibility but the yield of both grain and stover is decreased and there is also an increase of stalk breakage (Brenner *et al* 2010).
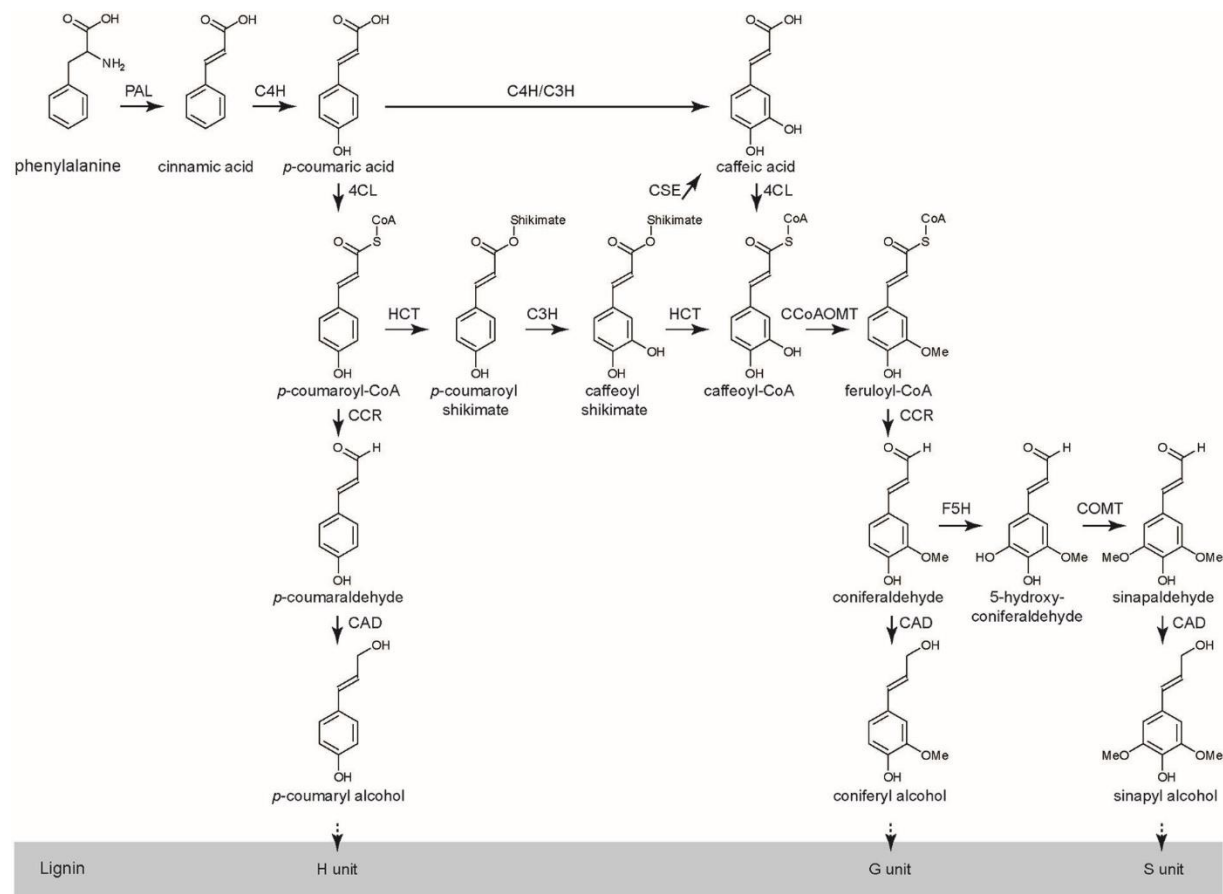
Figure 1.10 The phenylpropanoid pathway [phenylalanine-ammonia-lyase (PAL), cinnamate-4hydroxylase (C4H), 4-courmarate-3-hydroxylase (C3H), *O*-methyltranserase (OMT), ferulate-5-hydroxylase (F5H), hydroxycinnamate-CoA ligase (4CL), , 4-hydroxycinnamoyl-CoA (CCoA-3H), 5-hydroxyferuloyl-CoA-*O*-methltransferase (CCoA-OMT), cinnamoyl-CoA-reductase (CCR), cinnamyl-alcohol dehydrogenase (CAD)] (Miedes *et al* 2014).

1.2.4 Digestion of plant biomass

Although lignocellulosic biomass is rich in polysaccharides, its recalcitrant nature requires the use of high energy (and cost) pretreatments to open up the structure of the cell walls to allow hydrolytic enzymes to access their substrates. There are various methods that can be broadly classified as physical, chemical or biochemical pretreatments and the choice of which is dependent on the final objective, economics and the environmental impact of the process. This study focused on thermochemical pretreatments, predominantly those that have been shown to be the more cost effective or are used predominantly within industry, including methods involving water, acids or alkalis (Harmsen *et al* 2010).

Thermochemical pretreatments involve using water at high temperatures and pressure with or without added acid or alkali chemicals. Examples of this method are liquid hot water (LWA) pretreatment and steam explosion. In the case of steam explosion, the steam is injected into the reactor in the form of high-pressure saturated steam. The temperature of the reactor is increased to 160 – 260 ℃ and then the pressure is suddenly released. This change in pressure causes physical disruption of the material, while the high temperatures lead to the degradation of hemicellulose and disruption of lignin. Addition of mild acid solutions to the hot water treatment can result in hemicellulose hydrolysis, allowing these sugars to be removed prior to enzyme hydrolysis if required, but this also can lead to the formation of fermentation inhibitors by sugar dehydration (Harmsen *et al* 2010). Concentrated acids can also be used for total hydrolysis of the biomass polysaccharides without the need of enzymes, but this is costly in terms of chemicals and the equipment needed to deal with corrosive chemicals. (Kumar *et al* 2009, Harmsen *et al* 2010, Li *et al* 2010, Chaturvedi *et al* 2013).

The most common alkali pretreatment used involves sodium or calcium hydroxide under mild conditions. This method results in the removal of lignin and the hydrolysis of acetyl and uronic substitutions of hemicellulose as well as the breakage of ester linkages found in xylan. The advantage of this method is that the mild conditions prevent the lignin from condensing as well as avoiding the formation of

fermentation inhibitors. However, the reaction times can be relatively long and the salts also need to be removed before enzyme hydrolysis can take place, which results in a large amount of water being used (Harmsen *et al* 2010, Chaturvedi *et al* 2013). In this study, the pretreatment protocol used was based on a mild alkali pretreatment. Ammonia fibre explosion (AFEX) is another alkali pretreatment that is being trialled. This involves treating the biomass with aqueous ammonia at 60 - 100 ºC for approximately five minutes at a high pressure. This method leads to the lignin remaining intact, however some of the hemicellulose is hydrolysed and the cellulose is decrystallised. To keep costs down the ammonia has to be recovered and reused. This pretreatment has been used successfully on various types of lignocellulosic material from agricultural to municipal waste, however, it not suitable to biomass with a high lignin content (Chaturvedi *et al* 2013). Scaling AFEX to commercial applications is challenging from the costs of equipment and handling of hazardous chemicals.

Enzyme hydrolysis is used after pre-treating the lignocellulosic material in the biorefinery process to further degrade the biomass. A cocktail of enzymes are used that contain cellulases, hemicellulases and some accessory enzymes. These enzymes break down both the hemicellulose and cellulose found within the cell wall to monosaccharides ready for fermentation. The cellulolytic enzymes consist of endogluconases, exogluconases and β-glucosidases. The endo- and exocellulases work together to produce oligosaccharides and cellobiose, which the β-glucosidases converts to glucose. The oligosaccharides and cellobiose are produced by the endocellulases breaking the internal glycosidic linkages of cellulose whereas the exocellulases remove the cellobiose from the reducing and non-reducing ends of the polysaccharide (Mohanram *et al* 2013).

There are a number of different types of hemicellulases that can be classed as glycoside hydrolases, polysaccharide lyases, endo-hemicellulases and carbohydrate esterases. They include enzymes such as β-xylosidase, β-mannanase, endo-1.5-β-xylanse and acetylxylan esterase. They all have specific roles in hydrolysing hemicellulose by either breaking ester or glycosidic bonds or detaching side chains and chain substituents (Mohanram *et al* 2013).

The C_5 and C_6 sugars that are released from cellulose and hemicelluloses during pretreatment and enzyme hydrolysis are then fermented to produce the bioethanol used in second generation biofuels. The fermentation is usually conducted using an anaerobic method involving specially engineered yeast (Demibras 2009). These microorganisms have to be resistant to various compounds that are produced during the pretreatment and enzymatic hydrolysis steps, such as furfural and methyl-furfural which can be toxic to the yeast (Carroll *et al* 2009).

1.3   Targets for reducing lignocellulose recalcitrance

A possible route to improving the cost effectiveness of the production of biofuels is to generate lignocellulosic biomass with improved digestibility characteristics. This could be achieved by altering the plant cell wall properties through plant breeding (Foust *et al* 2008), transgenic manipulation (Vanholme *et al* 2008) and mutation identification (Halpin *et al* 1998). However, before this can be achieved a better understanding of the composition and structure of the cell wall is needed. This knowledge would help to predict how any changes in cell wall will impact the plant's physiology, development and resistance to diseases (Slavov *et al* 2013).

The most studied target is lignin in terms of quantity and composition. Alterations in this area can lead to an increase in accessibility to cellulose and hemicelluloses by the enzymes involved in cell wall degradation. Various lignin mutants have been identified in a number of important crop species. These mutants cover a range of enzymatic steps in the polyphenolic pathway and include the *bm* mutants in maize (Halpin *et al* 1998, Loqué *et al* 2015). However, these mutations tend to lead to undesirable plant growth effects such as low yield (Chen *et al* 2007, Vega-Sanchez *et al* 2010). In terms of lignin composition, it has been reported that grasses that contain a decrease in their S/G ratios have a better digestibility potential (Torres *et al* 2015). Another method under investigation is to incorporate unconventional monolignoids into the lignin matrix. This way a customised lignin polymer can be created which has the desired properties such as an increase in digestibility as well as possibly an increase in chemical value (Torres *et al* 2015).

Cellulose is a recalcitrant material due to it having a crystalline structure with a high degree of polymerisation. An understanding of the complex cellulose pathway is needed to be able to manipulate this trait. It has been hypothesised that a number of plasma membrane-associated proteins interact with the cellulose synthase complex and is responsible for the assembly, crystallisation and orientation of the cellulose microfibril (Torres *et al* 2015). Genetic manipulation in Arabidopsis has shown that it is possible to decrease cellulose crystallinity (Harris *et al* 2009, Vega-Sanchez *et al* 2010). However, this can lead to undesirable traits if the change is too severe. A possible solution to this is to identify milder allelic variations that produce the same desired effect (Torres *et al* 2015).

Previously, hemicellulose as a target for improved digestibility received less attention. It is now thought that the acetyl esters, methyl ester and ethers found within the polysaccharide contribute to the recalcitrant nature of lignocellulosic material by inhibiting the access of enzymes. When these esters are released they can inhibit fermentation (Loqué *et al* 2015, Torres *et al* 2015). It is known that hemicellulose also contains a significant amount of pentoses such as xylose that aren't efficiently fermented by current yeast strains. A possible target for improvement is to increase the hexose/pentose ratio found within the biomass (Loqué *et al* 2015). What still isn't clear is which enzyme complexes are involved in the production and mobilisation of the hemicellulose from the ER and Golgi. However, this is slowly beginning to improve with the knowledge gained from studying the various roles that the numerous *Csl* genes play (Torres *et al* 2015).

Another possible target that could be considered together with hemicellulose are the ferulate cross links that are found as either xylan-xylan ferulate bridges or ferulate-lignin crosslinks. These bonds have been shown to have a negative influence on saccharification potential. To date no genes have been identified that play a role in the esterification of the ferulate (Torres *et al* 2015).

Transcription factors have recently become a target for improving digestibility of lignocellulosic materials. It has been reported that a number of Arabidopsis transcription factors belonging to the NAC protein family play a regulatory role in controlling secondary cell wall formation. It is not surprising that transcription

factors are a possible target as it is known that they play a role in regulating complex metabolic pathways as well as defining tissue and organ differentiation (Slavov *et al* 2013, Torres *et al* 2015).

Most of the research to date has relied on knowledge we already have about plant cell wall biosynthesis. However there is still a lot of information about this topic that is unknown (Fagard *et al* 2003, Cardinal *et al* 2003, Persson *et al* 2005).

1.3.1 Brachypodium: a model plant for grass research

Past research has focused predominantly on Arabidopsis for studying plant cell walls. However this work is difficult to transfer to the majority of agriculturally important crop plants as the cell wall composition of grasses is significantly different from dicots. Grasses have Type II cell walls whereas dicots have Type I. These cell wall types differ in a number ways and the information is summarised in Table 1.2 (Vogel 2008, Bevan *et al* 2010). A model grass was needed for the study of plant cell walls due to these significant differences in cell wall composition as a large proportion of important agricultural crops are grasses.

Table 1.2 The difference between Type I and Type II plant cell walls.

	Type I Cell Walls	Type II Cell Walls
Type of plants	Dicots, Noncommelinoid Monocots, Gymnosperms	Commelinoid Monocots (grasses)
Cellulose	Coated by xyloglucan, pectin and structural proteins.	Coated in glucuronoarabinoxylans, hydroxycinnamates and mixed linkage glucans
Hemicellulose	Key component is xyloglucan No mixed linkage glucans	Key component is glucuronoarabinoxylan Mixed linkage glucans
Lignin	Composed predominantly of G and S units	Composed predominantly of G and S units but also has a significant amount of H units
Structural proteins and pectins	Both present in significant amounts	Ferulic and coumaric acid act like structural proteins in that is provides links between lignin and hemicellulose

Brachypodium distachyon (Brachypodium), whose common name is purple false brome, was identified as a possible model (Christensen *et al* 2010). It is a wild annual grass that is prevalent around the Mediterranean Sea and into India. It belongs to the Pooidaea subfamily of the Poaceae grass family (Figure 1.11). The Pooidaea subfamily is the largest and it consists of the majority of the cool season cereal, forage and turf grasses. Brachypodium has been shown to be closely related to other agriculturally important grasses such as maize and rice. It also shows co-linearity of gene organisation (Garvin *et al* 2008, The International Brachypodium Initiative 2010)
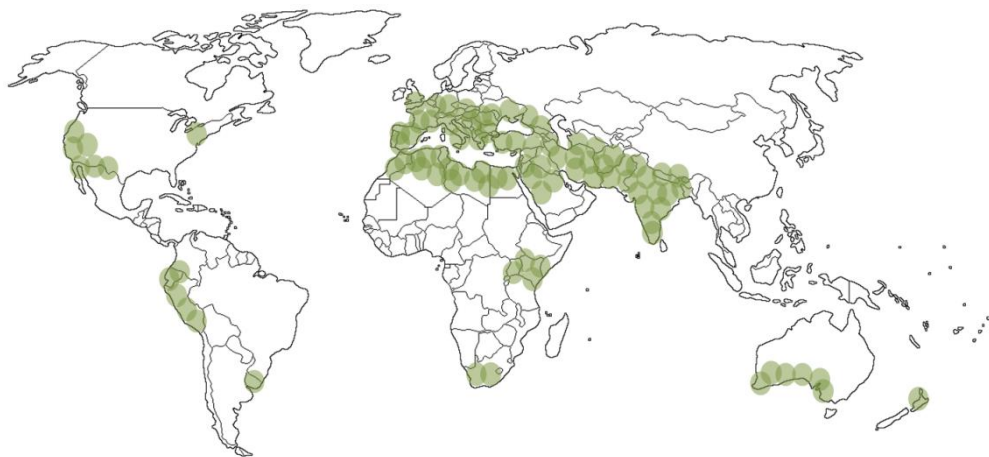
Figure 1.11 The world wide distribution of Brachypodium shown in green (adapted from Garvin *et al* 2008).

Many agricultural crops have big complex genomes and are large plants with relatively long generation times which make them difficult to use in genetic studies. However, the characteristics of Brachypodium makes it an ideal model in that it is easy to grow in a dense fashion, it has a relatively short life cycle, it is self-pollinating, it has a small diploid genome (275 Mbp) that has been fully sequenced (The International Brachypodium Initiative 2010), there are also a number of varieties available and various genomic tools have been developed (Bevan *et al* 2010, Christensen *et al* 2010, Vain 2011).

1.3.2 Research method

The approach followed in this project is to attempt to identify factors that can be manipulated to produce grasses with more digestible lignocellulose. The approach uses association genetics to identify regions within the genome of Brachypodium that have an impact on lignocellulose digestibility; so called quantitative trait loci (QTL). A lot of previous research has taken a reverse genetics approach, which generates hypotheses based on prior knowledge of the cellulose, hemicellulose and lignin biosynthesis pathways and uses targeted gene disruption or overexpression to test these hypotheses (Fagard *et al* 2000, Cardinal *et al* 2003). Association genetic studies in contrast take an empirical, hypothesis free, approach to examine the genetic elements that underpin natural variation in a trait of interest.

The first step of QTL analysis is the development of a mapping population. In this study we decided it would be best to obtain a recombinant inbred line (RIL) population that had already been developed for Brachypodium. This decision was taken because it is time-consuming to develop a population. It takes 6 to 8 generations of self-pollinating F2 lines to produce a RIL population (Figure 1.12). A population had already been developed by David Garvin that was used for the identification of a Barley Stripe Mosaic Virus resistance gene (*bsr1*) and contained 165 lines, which had been genotyped with 768 single nucleotide polymorphic (SNP) markers (Cui *et al* 2012). A RIL population was chosen as it is a homozygous true-breeding population that allows the use of the same population over time without losing information. It also has the advantage that it can use dominant and co-dominant markers systems (Fazio *et al* 2003).
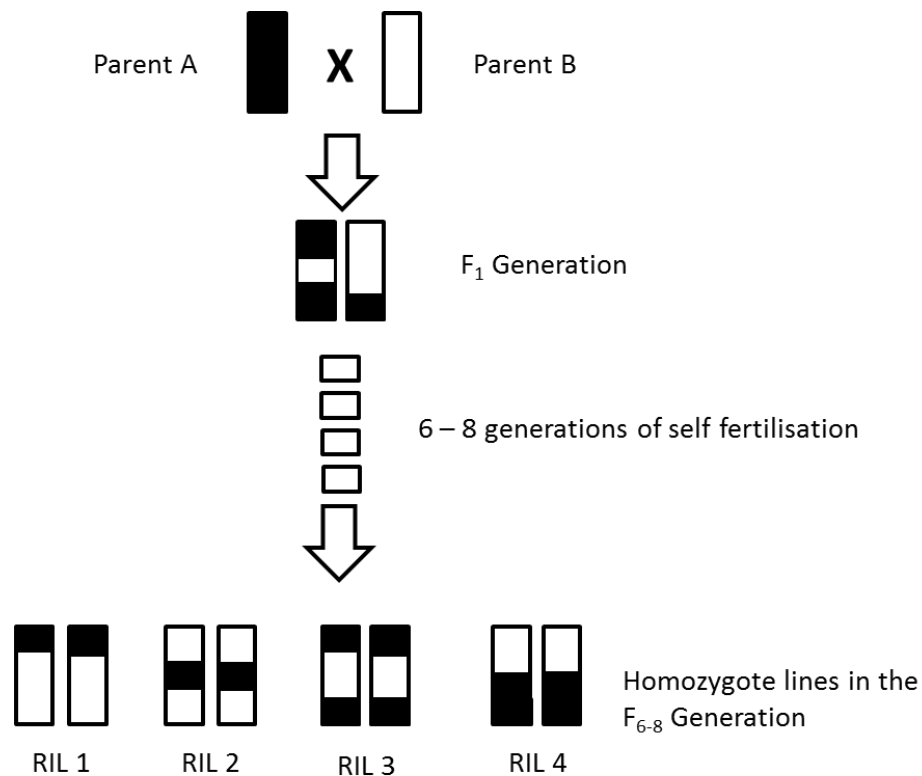
Figure 1.12 The formation of a recombinant inbred population.

There are a number of factors that can influence the success of a QTL experiment. The first one being the genetic properties of the population studied. It has been shown that only QTLs with a large phenotypic effect as well as those that are closely linked to a marker will be detected using this method. Secondly, there are a number of different sources of variation within an experiment that can have a negative effect on identifying a QTL. Finally, the population size can also have an effect on the success of QTL detection. It has been reported that the larger the population the more likely the detection of QTLs with smaller effects (Collard *et al* 2005).

The final step of QTL analysis is to identify candidate genes within the regions of interest that have been highlighted by the analysis. The advantage of using Brachypodium as the model instead of an agriculturally important plant, such as maize, is that it has a fully sequenced genome. The sequence annotation of the identified regions can help to pinpoint potential candidate genes. The target gene sequences of the two parental lines can be analysed to determine if it contains any allelic variations that could be used to explain the difference in function of the gene

(Borevitz *et al* 2004). The target genes can then be further validated functionally by producing plants in which the gene has been modified such as creating T-DNA insertion lines. This thereby allows for the impact on plant phenotype and chemotype to be assessed.

1.4   Aims

The need to decrease our dependency on fossil fuels and specifically for transportation has been shown to be an important target for a number of countries especially those within the European Union (EU). The European Parliament and Council have set out a number of targets within the Renewable Energy Directive (Directive 2009/28/EC) including that 20% of all energy needs must be met by renewables by 2020 and more specifically that 10% of all transport fuels must be renewable. This has led to a demand in biofuels but as previously discussed the production of second generation biofuels is not cost competitive in comparison to first generation biofuels or fossil fuels. To improve this situation more knowledge is needed concerning all areas of cell wall biosynthesis so that appropriate feedstocks can be generated in the future.

The aim of this study was therefore to identify genes involved in cell wall biosynthesis that have an impact on the recalcitrance of lignocellulosic biomass to digestion. To undertake this aim it was decided to use Brachypodium as a model plant as it is similar in terms of cell wall composition to a variety of important agronomical crops. Initially a population of Brachypodium accessions collected from various sites around the Mediterranean were analysed to determine if there was sufficient natural variation in terms of cell wall composition and digestibility to make it worth producing a RIL population from a bi-parental cross. This was experimentally determined by analysing lignin content, saccharification potential, sugars released during enzyme hydrolysis as well as silica content. In future, this information can be used to select lines for crossing to produce RIL populations for cell wall studies.

A forward genetic approach involving association genetics was also undertaken to identify QTLs linked to digestibility. In this case the aim I used already stabilised RIL population (generated initially to study disease resistance) to identify possible QTL for saccharification. These QTLs were confirmed and candidate genes identified as well as further experiments performed to validate their causative association with the observed QTL.

Chapter 2: Cell wall characterisation of *Brachypodium distachyon* accessions for the creation of a RIL population specifically for saccharification analysis.

2.1 Introduction

Plant cell wall biosynthesis and composition is a growing area of interest due to the increasing need to replace traditional transportation fossil fuels with more sustainable biofuels. However, the current methods for producing lignocellulosic bioethanol are expensive. This is due to the high input costs arising from the need for large amounts of energy and expensive enzymes that are necessary within the process. A possible route to decreasing these costs is to develop a better understanding to what factors contribute to the recalcitrant nature of the biomass material and identify genes and markers for its improvement (Gomez *et al* 2008, Carroll *et al* 2009).

In the past, the majority of plant research was conducted using Arabidopsis as a model species. However, as discussed in chapter one, it has been acknowledged that Arabidopsis may not be the best plant model for studying grasses due to the significant differences found within the structure of the cell wall as well as genetic differences (Feuillet *et al* 2002). Therefore, Brachypodium has been selected as a model species for important agricultural crops such as maize, rice, wheat, barley and oats as well as the biofuel crop, switchgrass (Bevan *et al* 2010).

Brachypodium was first recommended as a grass crop model plant by Draper *et al* (2001) as it has a number of properties that make it suitable. The International Brachypodium Initiative (IBI) whose aim is to develop and distribute genetic and genomic resources was formed in 2005. The formation of this group has led to the sequencing of the genome of the diploid inbred line Bd21, as well as the creation of T-DNA mutant lines, reference genotypes, BAC libraries, genetic markers, mapping populations and a transformation system (Garvin *et al* 2008, The International Brachypodium Initiative 2010, Vain 2011).

Research has predominantly focused on Bd21 and a few selected Brachypodium accessions. This has come about by the way the inbred lines were constructed from the original germplasm collections. Initially there were two main germplasm collections. One contained 30 accessions and was labelled with the prefix "PI". It was held by the United States Department of Agriculture (USDA). The second collection contained 38 accessions and was kept in Aberystwyth, Wales. It was identified with the prefix "ABR". There is some overlap between both collections. Twenty-eight accessions were created by single-seed descent from the "PI collection" and were given the label "Bd". This group of plants contained both diploid and polyploid individuals and was later used to create segregating populations such as the recombinant inbred lines from Bd3.1 x Bd21, which were used in creating a molecular map. It was at this time that Bd21 became the recognized standard genotype, or wild type, as it was used as the source for various large scale projects such as whole genome sequencing (Garvin *et al* 2008).

Early on it was noted that there was phenotypic variation between the "Bd" lines in terms of vernalisation requirements, flowering date, plant height, shattering and seed size (Garvin 2007, Schwartz *et al* 2010, Pacheco-Villalobos *et al* 2012). Molecular variation was also noted at a later date with the creation of Bd21-3 (Garvin *et al* 2008). Brachypodium was identified as a possible resource for understanding the interaction between plants and pathogens in a wild plant population because various accessions reacted differently to a range of fungal pathogens (Opanowicz *et al* 2008). Recently, research has been conducted using natural genetic variation to understand drought tolerance in Brachypodium (Lou *et al* 2011). However, there is not much known about whether there are significant differences between natural varieties in terms of secondary cell wall composition and if these differences can be exploited in understanding cell wall digestibility. If there is natural genetic variation present for digestibility it could be exploited to identify genes underlying this complex trait. This could be achieved by using QTL analysis, which has been shown to be successful in the past in Arabidopsis when looking at the function of CBF transcription factors (Gery *et al* 2011). It has also been used to identify large effect loci related to root architecture in Arabidopsis and Brachypodium (Pacheco-Villalobos *et al* 2012).

Saccharification is a method of measuring the digestibility of plant material. It recreates the process of producing sugars for fermentation followed within a lignocellulosic biorefinery but at a laboratory scale. The steps involved in saccharification begin with grinding biomass to reduce the particle size of the material. This is followed with a mild hot water, acid or alkaline pretreatment to open up the structure of the biomass. The next step is an enzymatic hydrolysis step, which releases the sugars from the biomass (Decker *et al* 2009, Studer *et al* 2010). In a biorefinery these sugars would then be used in a fermentation process to produce ethanol however in saccharification analysis the released sugars are simply measured using colourmetric, enzymatic or chromatographic methods (Anthon *et al* 2002) to determine the digestibility potential of the plant material. The University of York has developed a high throughput saccharification system carried out in 96-well plates using a semi-automatic robotic platform. This system is flexible which allows for a water or alkaline pretreatment to be used (Gomez *et al* 2010, Gomez *et al* 2011).

There are a number of factors that influence the digestibility potential of lignocellulosic biomass. These factors are linked with the structure of the material namely the composition and quantity of lignin; the composition of hemicellulose, as well as the crystallinity of the cellulose (Vega-Sanchez *et al* 2010). Traditionally, it was believed that lignin had the most influence on recalcitrance and therefore it has been the main area of cell wall composition research (Timpano *et al* 2014). Lignin has a complex polyphenolic structure composed of monolignols that are produced via the phenylpropanoid pathway that results in the formation of the three monomers. In past Arabidopsis research it was noted that the different ratio of these monomers as well as the total amount of lignin present within the cell wall had an effect on recalcitrance (Li *et al* 2010). However, in a number of cases when these factors were adjusted it led to undesirable plant traits such as low yield and weak stems (Vanholme *et al* 2010). It has recently been shown that lignin is not necessarily the main influencing factor as Wu *et al* (2013) discovered that for rice and wheat the main factors influencing digestibility were cellulose crystallinity and the xylose-arabinose bond in hemicellulose. Rollin *et al* (2011) were able to show that even when a relatively large amount of lignin was still present in switchgrass biomass after pretreatment, the accessibility to the cellulose was the most important factor in improving enzyme hydrolysis.

Another factor that affects recalcitrance and has been studied in terms of feed digestibility in ruminants is the presence of silica (Si). It is believed that Si forms either a physical barrier or produces inhibitors to the stomach enzymes responsible for digestion (Agbagla-Dohnani *et al* 2003). More recently there has been a growing need for plants with lower Si concentrations due to the increased demand for replacement energy sources. This is because when plants with high Si concentrations are burnt they release particles that are dangerous to human health as well as fouling the burners, which reduces productivity (Reidinger *et al* 2012). The role of Si in plants is to protect them against abiotic and biotic stressors such as providing drought tolerance, defence against herbivores and protection against fungal diseases. It is thought that Si is able to do this by providing physical and biochemical responses (Currie *et al* 2007, Reidinger *et al* 2012).

Si is found naturally in the soil as monosilic acid (H_4SiO_4) at a concentration of 0.1 – 0.6mM. This makes it the second most abundant element found in soil and as common as other major organic nutrients such as K^+, Ca^{2+} and SO_4^- as well as exceeding the concentration of phosphates (Epstein 1994). H_4SiO_4 can be transported into the plant either actively or passively and it follows the transpiration stream via the xylem to the end of the transpiring organs such as leaves, stems and inflorescence bracts found in cereals or into hairs and trichomes. The Si is deposited as an amorphous hydrated opal (SiO_2), which is polymerized as either a thick layer found externally of the epidermis or intracellularly where it can be associated with the components found within the cell wall (Epstein 1994, Currie *et al* 2007, Agbagla-Dohnani *et al* 2003).

Si can be found at concentrations of 0.1 – 10% of plant dry weight, which at its lowest makes it as abundant as some other plant macronutrients such as P, S, Ca and Mg. At its highest concentration of 10% it is more prevalent than the nutrients K and N. Even though it is found at such concentrations within the plant it is not considered an essential element, except in the case of Equisitaceae (Epstein 1994, Currie *et al* 2008). This large variation in Si accumulation led Jones and Handreck (1967) to initially classify crop plants as either dicotyledons that had low Si concentrations of 0.1%, dryland grasses with 1% Si or wetland grasses that have 5% or higher Si. In

1977 Takahashi and Miyake broadened the classification to Si accumulators (>1% Si concentration) and non-accumulators (<1% Si concentration). It has been reported that mainly monocotyledons are Si accumulators, especially Poaceae, Equisetaceae and Cyperaceae (Currie *et al* 2007). Rice is known to be an efficient Si accumulator with a very high Si concentration of up to 10% plant dry weight (Agbagla-Dohnani *et al* 2003).

The aim of this study is to determine if there is natural variation present within a selection of Brachypodium accessions in terms of cell wall digestibility and composition. This knowledge will be used to select suitable accessions for the development of a recombinant inbred line that can be used for further research into cell wall biosynthesis and indicate some of the major factors influencing digestibility in this species.

2.2   Material and Methods

2.2.1 Preparation of plant material

A selection of 22 natural accessions of Brachypodium was grown to maturity and allowed to desiccate at INRA-Versailles, France. Dried stems were harvested, stripped of leaves and heads and then ground into a powder to be used for further analysis. These samples were sent to the University of York together with information concerning their chromosomal ploidy and the Klason lignin content, which is the measure of the acid-insoluble fraction of lignin present within the cell wall. A summary of this information can be found in Table 2.1 and Figure 2.2.

A population of 93 F_2 plants was developed by INRA-Versailles, France. This population was developed by crossing the chosen natural accessions, Bd21 and BdTR1.1F. Once the F_2 plants reached maturity, the dried material was harvested and sent to The University of York for saccharification analysis. The plant biomass from the Bd21 x BdTr1.1f population was prepared for analysis by cutting the internodes into small fragments and placing them into 2ml tubes together with three ball bearings for grinding in the grinding and weighing robot (Gomez *et al* 2010, Gomez

et al 2011). The nodes of the stems were excluded together with the lowest and highest internodes.

Table 2.1: The Brachypodium natural accessions that were selected for further analysis together with information about their origin and chromosomal ploidy (data supplied by INRA-Versailles, France).

Accession	Origin	Ploidy
Abrc1d	Turkey	2n
Abrc2b	France	2n
Abrc5a	Spain	2n
Abrc7d	Spain	2n
Abrc8d	Italy	2n
Bd18-1	Turkey	2n
Bd21	Iraq	2n
Bd21-3	Iraq	2n
Bd3-1	Iraq	2n
BdR18	Turkey	4n
BdTr1.1f	Turkey	2n
BdTr11.1c	Turkey	2n
BdTr13.1g	Turkey	2n
BdTr3.1a	Turkey	2n
BdTr4.1d	Turkey	4n
BdTr6.1b	Turkey	4n
BdTr7.1a	Turkey	2n
BdTr9.1a	Turkey	2n
Cre0	Crete	4n
Esp0	Spain	2n
Esp1	Spain	4n
Est0	France	4n

2.2.2 Saccharification analysis

The samples for each population were randomised before formatting took place. For each sample, four replicates of 4 mg were dispensed into the 96-well plates. The samples were screened using the liquid handling robot, which utilised a programme involving a pretreatment of 0.5 N sodium hydroxide (NaOH) or water at 90 ºC for 30 minutes and an enzymatic incubation time of 8 hours at 50 ºC. The enzyme used is commercially available and consists of Celluclast and Novozyme 188 (Novozymes A/S, Bagsvaerd, Denmark) at a ratio of 4:1. An enzyme loading of 6.3 FPU/g material was used. The reducing sugars released from the material were detected using an adapted 3-methyl-2benzothiazolinonehydrazone (MBTH) method (Gomez *et al* 2010, Gomez *et al* 2011). The population of 22 natural accessions was screened twice using either a NaOH or water pretreatment, whereas the Bd21 x BdTr1.1f population was screened twice using only the NaOH pretreatment.

2.2.3 Lignin analysis

The Klason lignin determination method was used to quantify the amount of acid-insoluble lignin present in the natural accessions (Dence 1992). This method was conducted at INRA-Versailles, France on dried stem material that had been ground.

2.2.4 ATR-FTIR analysis

Powder from the three biological replicates of each natural accession was analysed three times using attenuated total reflectance Fourier transform infrared (ATR-FTIR) spectroscopy (PerkinElmer, UK). The method involved placing the sample directly onto the diamond and applying the pressure arm until it reaches a value of 100 to insure contact between the sample and the diamond. The spectrum from 1850 – 850 cm^{-1} was collected at a resolution of 4 cm^{-1}. A total of 256 scans were taken during each run to reduce any background noise. The principle component analysis (PCA) of the results was determined using Unscrambler X software (CAMO Software, Norway), which is a multi-variant analysis program. Within this program it is also possible to normalise the data before analysis. In this case the data was pre-treated

using peak normalisation. This normalised data was then used for the PCA analysis to determine any groups of data.

2.2.5 Monosaccharide analysis

The sugars present in the hydrolysate after saccharification were analysed by high-performance anion exchange chromatography (HPAEC) on a Dionex ICS-3000, with a Carbopac P20 column (Dionex, UK) (Jones *et al* 2003). The dried down saccharification hydrolysate from the three biological replicates of each accession, together with monosaccharide standards, was treated with 2 M trifluoroacetic acid. The chemical hydrolysate was dehydrated and washed twice with isopropyl alcohol, which was evaporated. The resulting monosaccharides were finally suspended in water and then filtered using a 4 mm syringe driven filter (MerckMillipore, Germany) before loading onto the Dionex.

2.2.6 Silica analysis

Powder from three biological replicates of each of the Brachypodium natural accessions was pelleted in a 10 ton manual hydraulic press (Specac, UK) for two seconds. The pellets were analysed using a portable x-ray fluorescence spectrometer (Niton XL3t900 GOLDD Analyzer; Thermo Scientific, UK) to determine the amount of silicon present (Reidinger *et al* 2012). A standard curve was produced using the following certified reference materials (CRM); tea, spinach, bush and phalaris. The sample pellets were each read once, whereas the CRM were scanned four times.

2.3   Results and discussion

2.3.1 Saccharification analysis of natural accessions

The 22 natural accessions were screened twice using two different pretreatments. In the first instance a mild alkaline pretreatment of NaOH was used. This produced an average of 57.02 nmol reducing groups released/mg material.hr^{-1}(Figure 2.1A). Plant accession Abrc8d released the most sugar whereas accession Bd18-1 released the lowest amount of sugar.

The second pretreatment tested was hot water (Figure 2.1B). In this case an average of 18.56 nmol reducing groups released/mg material.hr^{-1}, which is approximately three times lower than the amount of sugar released when the sodium hydroxide pretreatment was used. Plant accession Abrc8d and Bd18-1 remained the most and least digestible respectively even though the overall amount of released sugar was lower.

These results therefore indicate that the sodium hydroxide pretreatment is more effective in disrupting the lignocellulosic structure and making it more digestible than the hot water pretreatment as more sugar was released. For some plant lines, like AbrC8d and Bd18-1, the type of pretreatment used did not affect digestion potential. However other lines, such as Est0 and Esp1, were more digestible with hot water than with mild alkaline pretreatment.
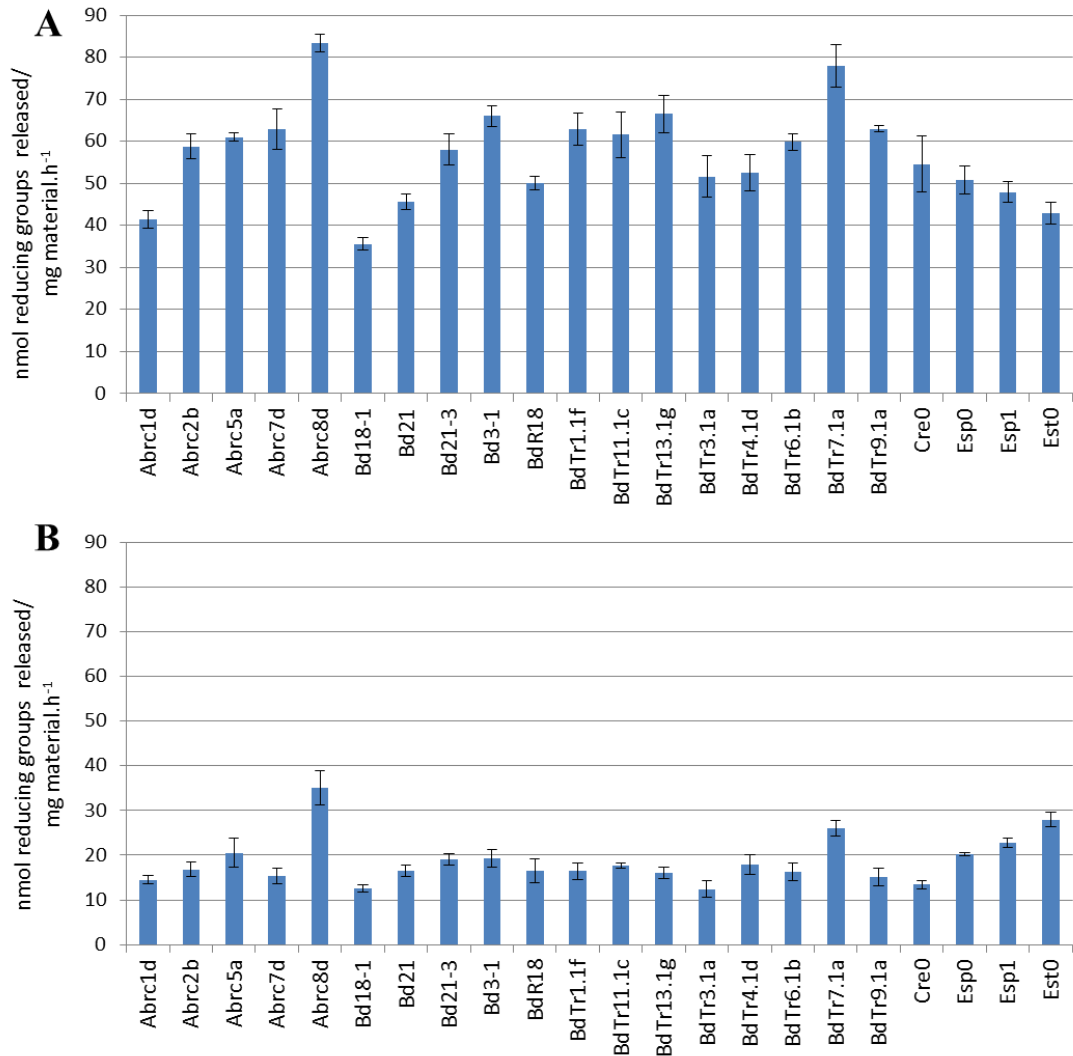
Figure 2.1: Saccharification analysis of the ground stem internodes of the Brachypodium accessions using (A) a 0.5 N NaOH pretreatment or (B) a water pretreatment at 90 °C for 30 minutes followed by digestion with a commercial cellulase coctail for eight hours at 50 °C. Results are the means and the standard deviations from two replicates.

The amount of acid-insoluble lignin detected in the natural accessions varied between 18 and 23% with an average of 21% (Figure 2.2). It has previously been reported that Bd21-3 contained between 13.5% and 19.6% when using the Klason lignin method (Timpano *et al* 2014, Ho-Yu-Kuan *et al* 2015).
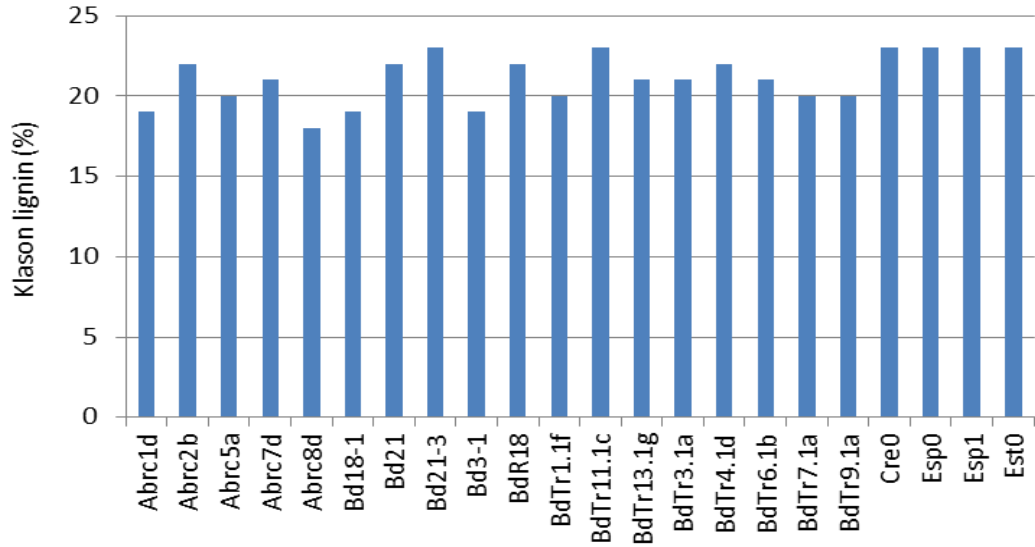
Figure 2.2: Klason lignin content of the Brachypodium natural accessions. Data supplied by INRA-Versailles, France. Briefly describe how data were obtained andif it was replicated- why are there no error bars? See 2.1 comments

The results that were obtained from the saccharification analysis were compared with the Klason lignin content to determine if there was a correlation between the two traits. When the results were compared for either pretreatment no correlation was found between the digestibility of the accession and the amount of lignin present in the sample (Figure 2.3). It can be concluded from the results that the lignin content of these Brachypodium accessions do not have an influence on digestibility but suggest other factors may rather play a key role.
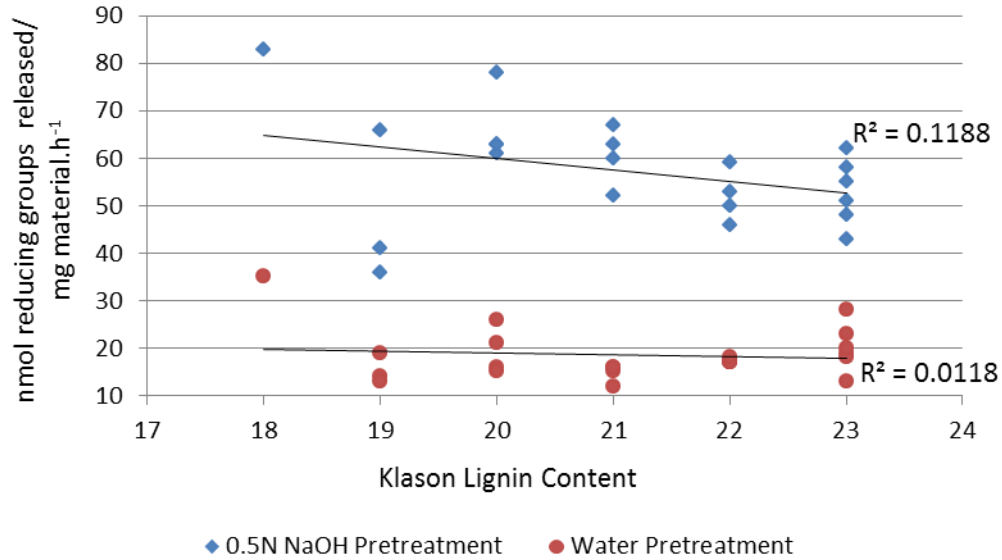
Figure 2.3: Correlation between the amount of reducing groups released during saccharification with either the 0.5N NaOH or water pretreatment and the Klason lignin content of the Brachypodium accessions.

2.3.2 ATR-FTIR analysis of natural accessions

Powder from the 22 Brachypodium natural accessions was screened using ATR-FTIR to determine if there was a difference in the composition of the secondary plant cell wall. It was hoped that this method of analysis would highlight any plant accessions that were significantly different from the rest or indicate what fraction of the cell wall further analysis should focus on.

Initially, the spectra data from all the accessions were analysed together using PCA to determine if any of the accessions formed clusters together thereby indicating that they had a similar composition (Figure 2.4). From the results it was determined that there was one main cluster and three smaller clusters. The main regions of difference observed in the spectra appear to be those that have been previously linked to changes in lignin ($1700 - 1500$ cm^{-1}), hemicellulose ($1200 - 800$ cm^{-1}) and sugar ($950 - 750$ cm^{-1}) composition (Kačuráková *et al* 2000, Kačuráková *et al* 2001).
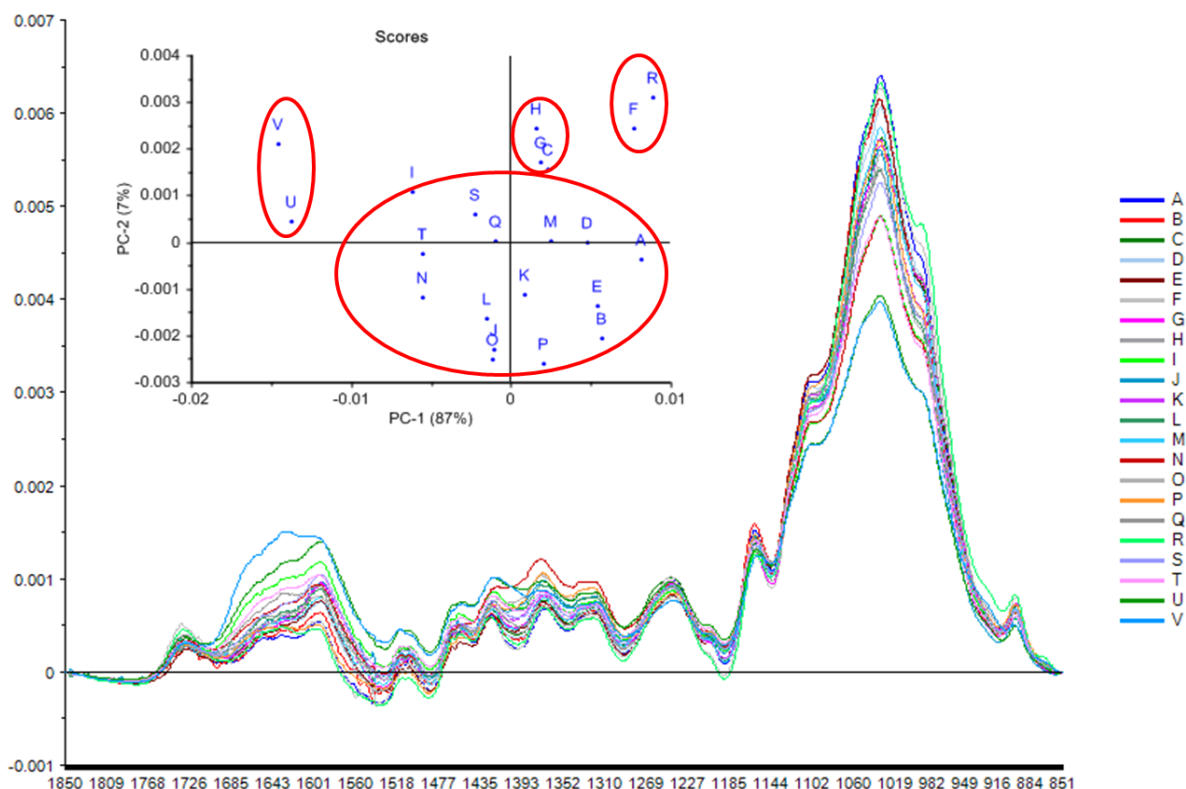
Figure 2.4: ATR-FTIR analysis of the spectra from all the Brachypodium accessions using PCA. The spectrum is an average of three replicates per sample analysed over the region 1850 – 850 cm[-1] at a resolution of 4 cm[-1] for 256 scans. All data was peak normalised before analysis using UnscramblerX. The red circles indicate the groupings of samples identified during PCA analysis. Sample identification: (A) Est0, (B) Esp1, (C) Esp0, (D) Cre0, (E) BdTr9.1a, (F) BdTr7.1a, (G) BdTr6.1b, (H) BdTr4.1d, (I) BdTr3.1a, (J) BdTr13.1g, (K) BdTr11.1c, (L) BdTr1.1f, (M) BdR18, (N) Bd3-1, (O) Bd21-3, (P) Bd21, (Q) Bd18-1, (R) Abrc8d, (S) Abrc7d, (T) Abrc5a, (U) Abrc2b, (V) Abrc1d.

The outliers were then further analysed to get a better understanding as to how these lines differ from each other (Figure 2.5). The accessions selected were Abrc1d (V), Abrc2b (U), BdTr7.1a (F) and Abrc8d (R). From the spectra is can be determined that accessions V and U have large differences in lignin (1700 – 1500 cm[-1]) and sugar (950 – 750 cm[-1]) composition (Kačuráková *et al* 2001). There also appears to be some differences in the regions related to hemicellulose.
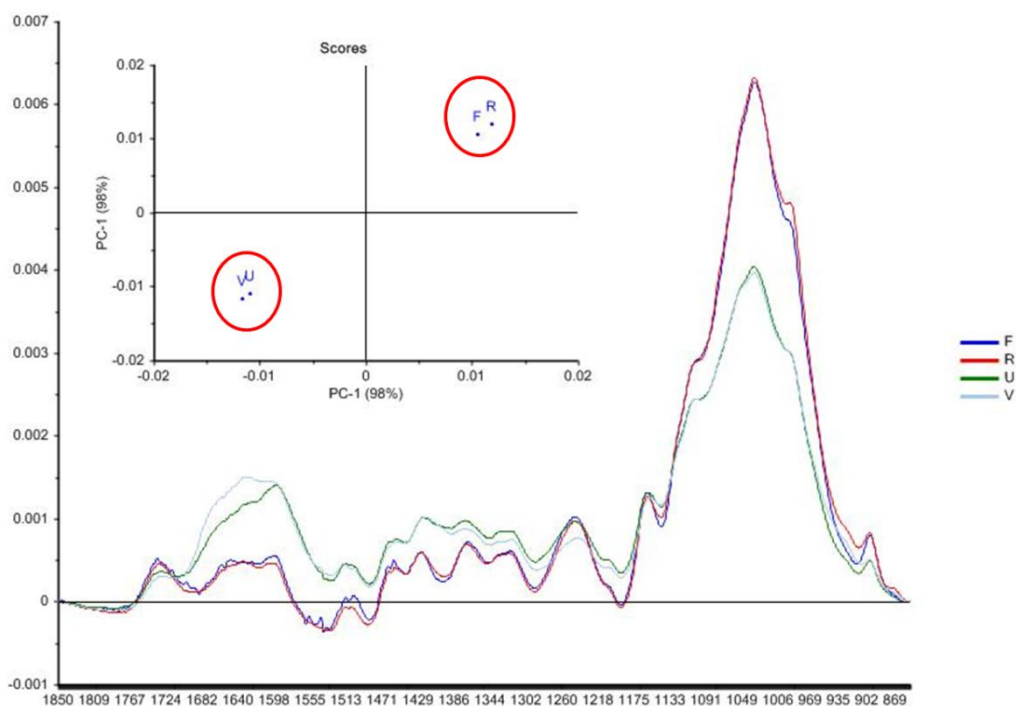
Figure 2.5: ATR-FTIR analysis of the spectra from the outlieing Brachypodium accessions using PCA. The spectrum is an average of three replicates per sample analysed over the region 1850 – 850 cm^{-1} at a resolution of 4 cm^{-1} for 256 scans. All data was peak normalised before analysis using UnscramblerX. The red circles indicate the groupings of samples identified during PCA analysis. Sample identification: (F) BdTr7.1a, (R) Abrc8d, (U) Abrc2b, (V) Abrc1d.

Due to the lignin region being highlighted as containing differences in composition as well as having laboratory data about lignin content it was decided to compare the spectra of those accessions that had either a high [Cre0 (D), Esp0 (C), Esp1 (B), Est0 (A), Bd21-3 (O) and BdTr11.1c (K)] or low [Abrc8d (R), Abrc1d (V), Bd18-1 (Q) and Bd3-1 (N)] lignin content using PCA analysis (Figure 2.6). The analysis did not give two distinct clusters as expected but rather one main cluster and two outliers. The main cluster was composed of accessions with high or low lignin content, whereas both outliers (V and R) contained low lignin concentrations. It is possible that the presence of a single main cluster instead of the expected two clusters could indicate that there is not enough of a difference in lignin composition to separate the two groups even though the lignin region has been highlighted as variable when observing the spectra data.
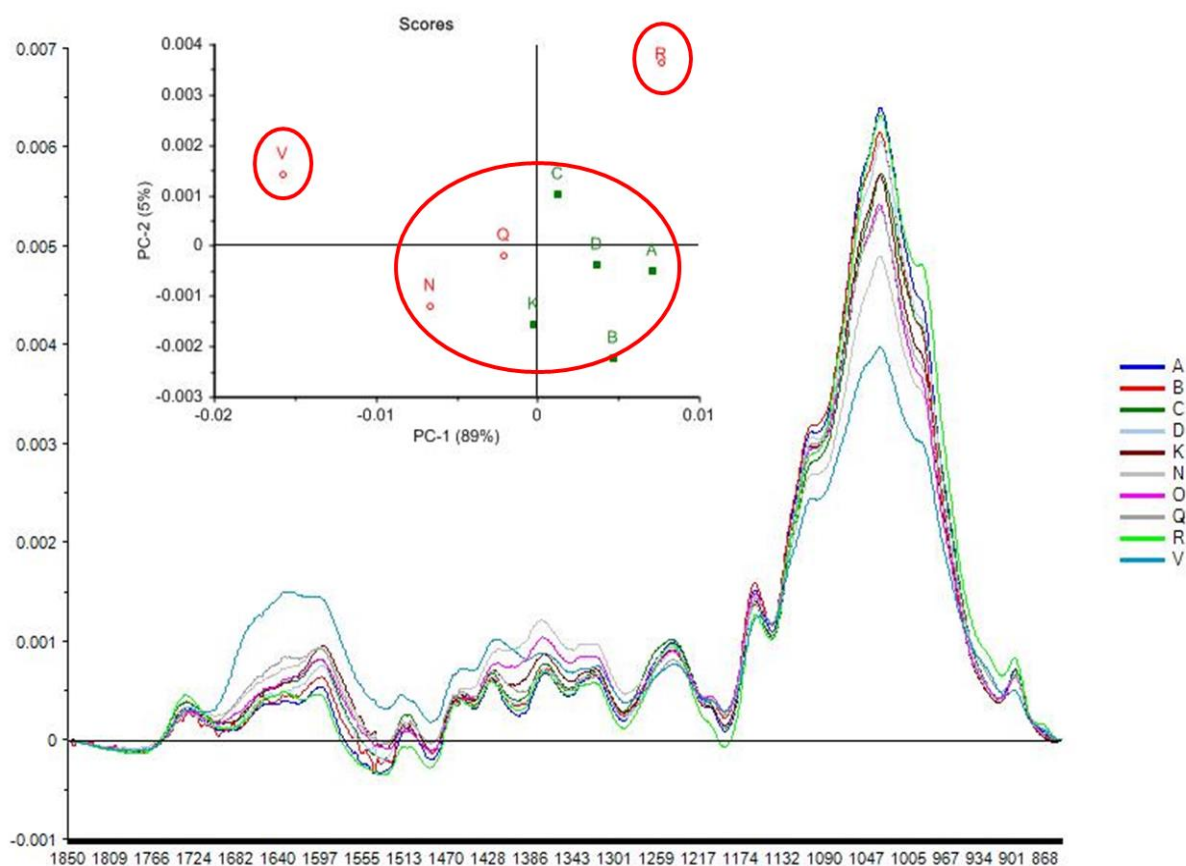
Figure 2.6: ATR-FTIR analysis of the spectra from the Brachypodium accessions that have either high (green) or low (red) lignin content using PCA. The spectrum is an average of three replicates per sample analysed over the region 1850 – 850 cm^{-1} at a resolution of 4 cm^{-1} for 256 scans. All data was peak normalised before analysis using UnscramblerX. The red circles indicate the groupings of samples identified during PCA analysis. Sample identification: (A) Est0, (B) Esp1, (C) Esp0, (D) Cre0, (K) BdTr11.1c, (N) Bd3-1, (O) Bd21-3, (Q) Bd18-1, (R) Abrc8d, (V) Abrc1d.

It was decided to take a closer look at the two low lignin outliers (V and R) to determine the reason as to why they were not clustering together. These accessions were analysed together with two high lignin lines (A and D). From this analysis it was observed that accession R clustered together with the high lignin lines (Figure 2.7). When the spectra were analysed it was determined that accession V was different from the other accessions in the lignin region but also throughout the spectrum whereas accession R was very similar to the high lignin lines, specifically accession A.
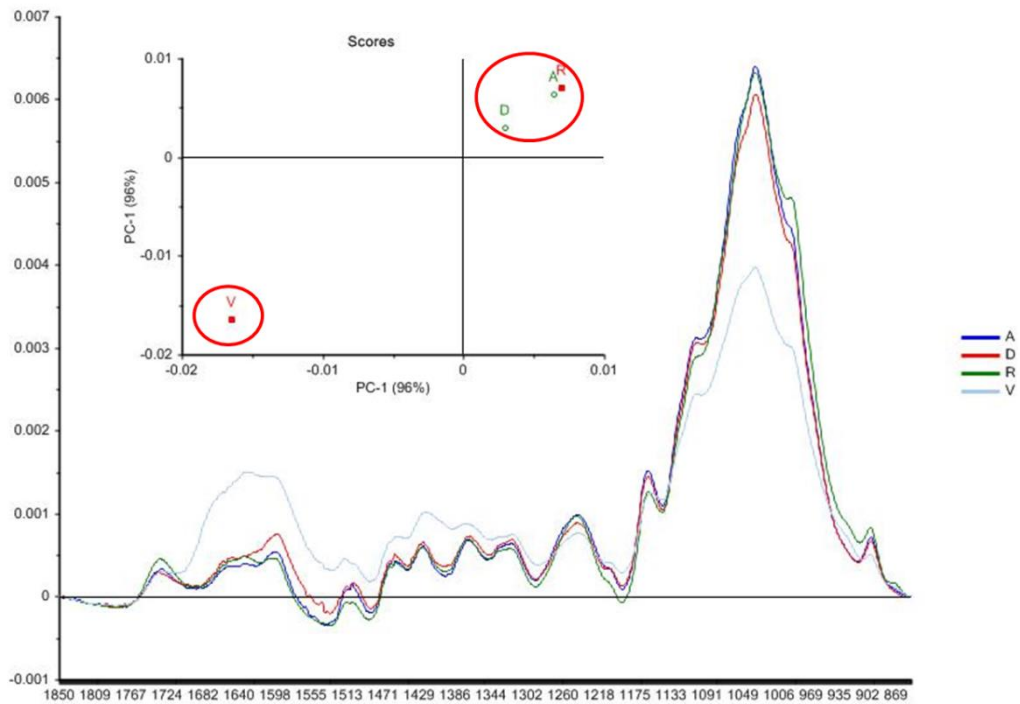
Figure 2.7: ATR-FTIR analysis of the spectra from selected outlier Brachypodium accessions that have either high (green) or low (red) lignin content using PCA The spectrum is an average of three replicates per sample analysed over the region 1850 – 850 cm^{-1} at a resolution of 4 cm^{-1} for 256 scans. All data was peak normalised before analysis using UnscramblerX. The red circles indicate the groupings of samples identified during PCA analysis. Sample identification: (A) Est0, (D) Cre0, (R) Abrc8d, (V) Abrc1d.

The ATR-FTIR results therefore confirm what was observed in the initial analysis. There is some difference in lignin content but this is not necessarily significant to affect digestibility. The main region of interest for change in cell wall composition appears to be within the sugars that make up the different components of the lignocellulosic material. ATR-FTIR is a quick and easy method to use to obtain information about the cell wall composition. However, it is not always easy to attribute a clear influence of specific polysaccharides so any observations have to be backed up by further evidence obtained from chemical experiments (Hori *et al* 2003).

2.3.3 Monosaccharide analysis of saccharification hydrolysate

Monosaccharide analysis was conducted on the 22 natural accessions to determine which sugars were released during saccharification and if there was a difference between the plant lines. The hydrolysate after saccharification analysis involving the hot water pretreatment was used to determine which sugar monomers were released during the enzymatic digestion.

From Figure 2.8 it can be noted that the accession BdTr3.1a has the most significant difference in sugars that were released from the cell wall. In this case is has a higher proportion of xylose and mannose compared with the other accessions. The main building block of hemicellulose in grasses is xylan, which is composed of a backbone of β-(1→4)-linked xylose residues. Another type of hemicellulose is mannans, which have a β-(1→4)-linked backbone of mannose units. (Scheller *et al*, 2010). This increase in xylose and mannose monomers in the hydrolysate could indicate that the hemicellulose component of the cell wall is more accessible or easier to break down by the enzymes or even that there is an increase in the amount of these monomers present.

Another group of accessions that have a different sugar composition include the following plants; BdTr4.1d, BdTr6.1b, BdTr7.1a, BdTr9.1a, Cre0, Esp0, Esp1 and Est0. These accessions all have an increased amount of glucose and xylose monomers in the hydrolysate. This could indicate an increase in the cellulose and/or hemicellulose digestibility due to improved accessibility to enzyme degradation or an increase in content as cellulose is constructed from glucose subunits whereas xyloglucan contains glucose and xylose. However, it must be taken into account that xyloglucan only makes up a small fraction of grass secondary cell walls (Scheller *et al* 2010). Another thing to note about these accessions is that the majority of them are tetraploid (4n), which could explain the difference in composition in comparison to the diploid lines. It has been noted that polyploidy can create diverse phenotypes in terms of plant physiology and influence developmental processes such as tolerance to stress and differences in flowering time (De Storme *et al* 2014). It is unclear how it specifically affects cell wall composition.
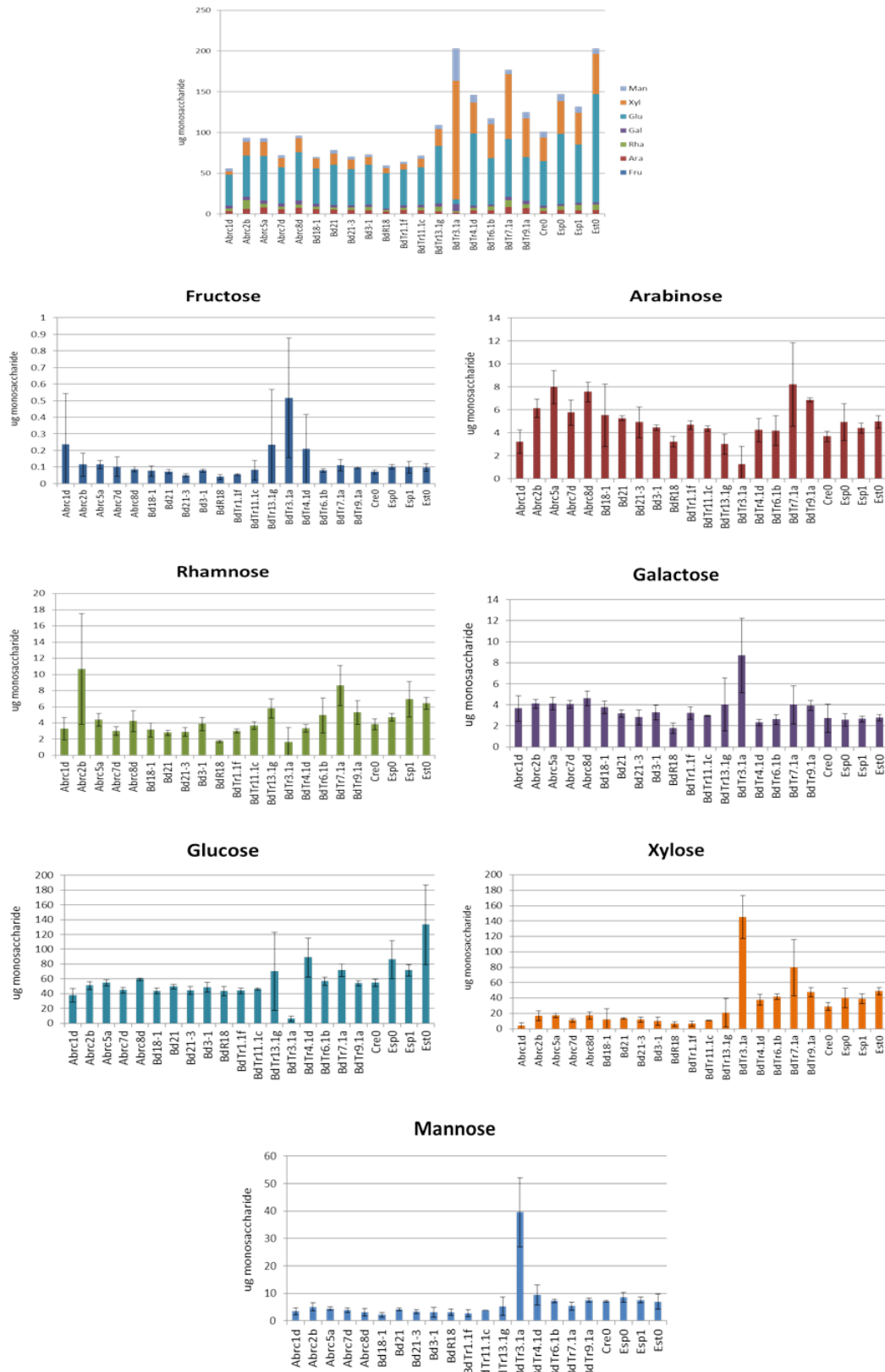
Figure 2.8: Monosaccharide analysis of the neutral sugars released in the hydrolysate of Brachypodium accessions after saccharification with a hot water pretreatment. The samples were prepared using a 2M TFA protocol and they were run using HPAEC on a Dionex together with quantifiable standards. The results are the means and standard deviations of three replicates.

2.3.4 Silica analysis of natural accessions

The Brachypodium accessions have varying amounts of silica within the cell wall (Figure 2.9). Accession Bd18-1 has the lowest silica content of 0.13% whereas BdTr6.1b has the highest amount of silica, 1.42%. Therefore the average silica content for the accessions tested was 0.69%. This is comparable to published data concerning silica content in other plants. It has been reported that typically $0.1 - 5\%$ of the dry weight of plants is silicon (Reidiner *et al* 2012). Some members of the grass family are considered high silica accumulators such as rice, which has been reported to contain between 3% and 10% silica per plant dry weight (Jin *et al* 2006, Currie *et al* 2007). However, from this data it appears that Brachypodium is a low accumulator of silica.
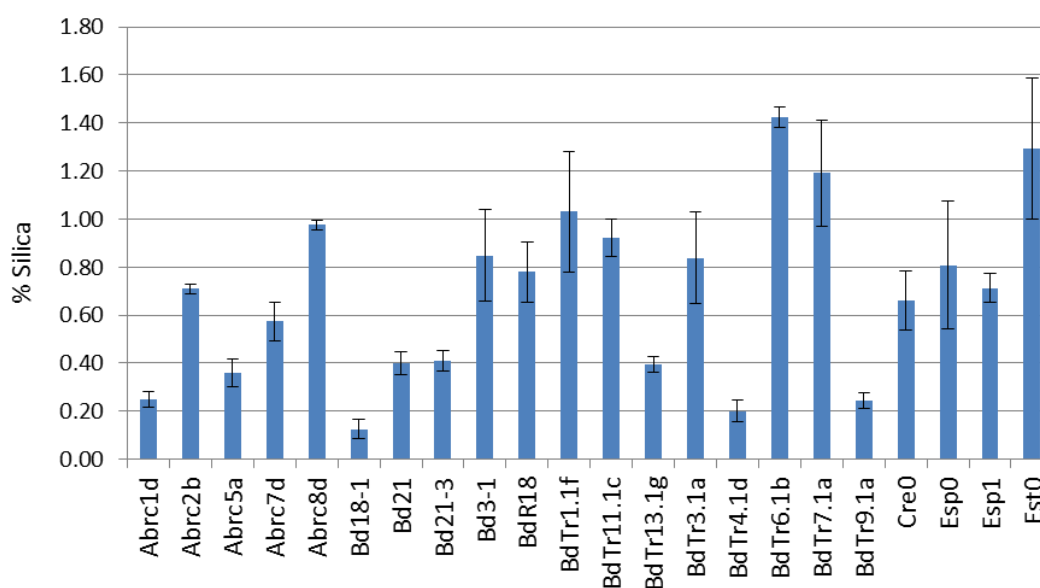
Figure 2.9: Silica content of ground stem internodes from the Brachypodium accessions. The biomass pellets were analysed using a portable x-ray fluorescence spectrometer and a standard curve was produced using certified reference materials. The results are the means and standard deviations of three replicates.

The amount of silica was then compared with the amount of sugar released during saccharification to determine if silica content has an effect on digestibility. From the correlation of the two traits (Figure 2.10) it can be determined that silica does not significantly influence digestibility when the samples undergo a sodium hydroxide (*t = 0.871, df = 20, p = 0.394, R² = 0.037*) or a hot water (*t = 1.901, df = 20, p = 0.07,*

58

$R^2 = 0.153$) pretreatment. This result is different to what has been reported before as previous studies on ruminants have noted that silica has a negative effect on the digestibility of forage grasses. This is possibly due to the silica forming a physical barrier that prevents enzymes produced by stomach microorganisms from digesting the grass (Agbagla-Dohnani *et al* 2003). Further studies in rice straw have confirmed that both silica and lignin have a negative effect on digestibility and the saccharification potential of the lignocellulosic material is also dependent on the part of the plant being analysed (Jin *et al* 2006). However, in both of these cases plants with high silica content were studied, namely rice, whereas the studied Brachypodium accessions only contain a relatively small amount of silica.
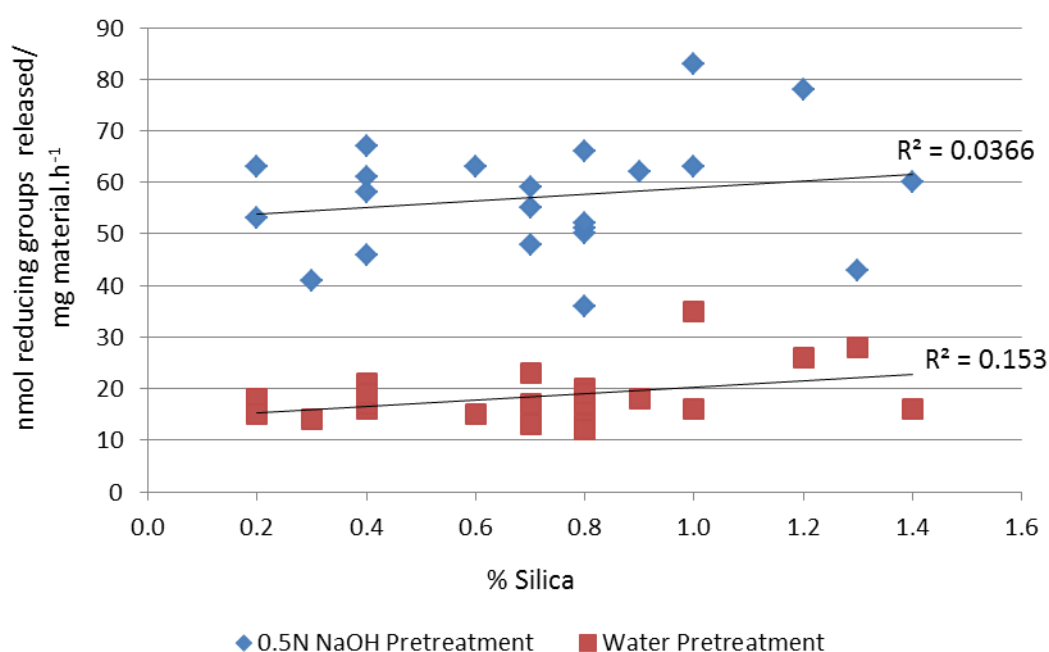
Figure 2.10: Correlation between the amount of reducing groups released during saccharification with either the 0.5N NaOH or water pretreatment and the silica content of the Brachypodium accessions.

2.3.5 Saccharification analysis of Bd21 x BdTr1-1f

A number of natural accessions were suggested as possible lines for creating crosses. Abrc8d and Bd18.1 were identified due to having the highest and lowest digestibility respectively. BdTr3.1a was also included due to the large amount of xylose released during saccharification as was observed in the data from the monosaccharide

analysis of the hydrolysate. However, in the end BdTr1-1f was selected due to it being the diploid accession with the highest amount of glucose released during enzyme digestion. BdTr1-1f was crossed with Bd21 by INRA-Versailles, France to produce a RIL population. Bd21 was used as it is considered to represent the wild type accession and it has been used in developing a number of genetic tools as well as being fully sequenced. Bd21 and BdTr1-1f appear to have differing cell wall properties (Table 2.2). In the case of Bd21 it appears to be more recalcitrant to digestion as it releases only 45.62 nmol reducing groups released/mg material in an hour, which is lower than BdTr1-1f (62.87 nmol reducing groups released/mg material.h^{-1}) as well as below the average release by the all the accessions (57.02 nmol reducing groups released/mg material.h^{-1}) screened. BdTr1-1f contains more silica (1.04%) than Bd21 (0.4%). However, Bd21 and BdTr1-1f contain similar lignin content.

Table 2.2: Properties of natural accessions Bd21 and BdTr1.1f.

Accession	Origin	Ploidy	Lignin (%)	Saccharification - 0.5N NaOH (nmol reducing groups released/ mg material.h-^{1})	Silica (%)
Bd21	Iraq	2n	22	45.62	0.4
BdTr1.1f	Turkey	2n	20	62.87	1.03

The F$_2$ plants from the Bd21 x BdTr1-1f Brachypodium population were grown by INRA-Versailles, France and then sent to the University of York for analysis. The F$_2$ plants were screened twice using saccharification analysis with a 0.5N NaOH pretreatment to determine the variation in digestibility within the population. Both screens gave very similar distributions with the average of Screen 1 being 31.75 nmol reducing groups released/mg material.h^{-1} and the average of Screen 2 is 31.58 nmol reducing groups released/mg material.h^{-1} (Figure 2.11). The overall digestibility of the plant lines is relatively low however the conditions used allows for minor alterations in digestion can be identified when screening a population. This is achieved by using a mild pretreatment which only allows a limited release of the sugars from the cell wall this is in contrast with what is observed in a biorefinery where the maximum amount of sugars are released by using much harsher pretreatment conditions. Therefore, the aim in this experiment was to determine if

there was a variation in digestibility present within the Bd21 x BdTr1-1f population. This was confirmed by plant line 462 showing the lowest digestion (19.39 nmol reducing groups released/mg material.h^{-1}) and line 429 the highest (44.02 nmol reducing groups released/mg material.h^{-1}). This population will continue to be developed by INRA-Versailles, France to produce a RIL population to be used for studying recalcitrance in lignocellulose.
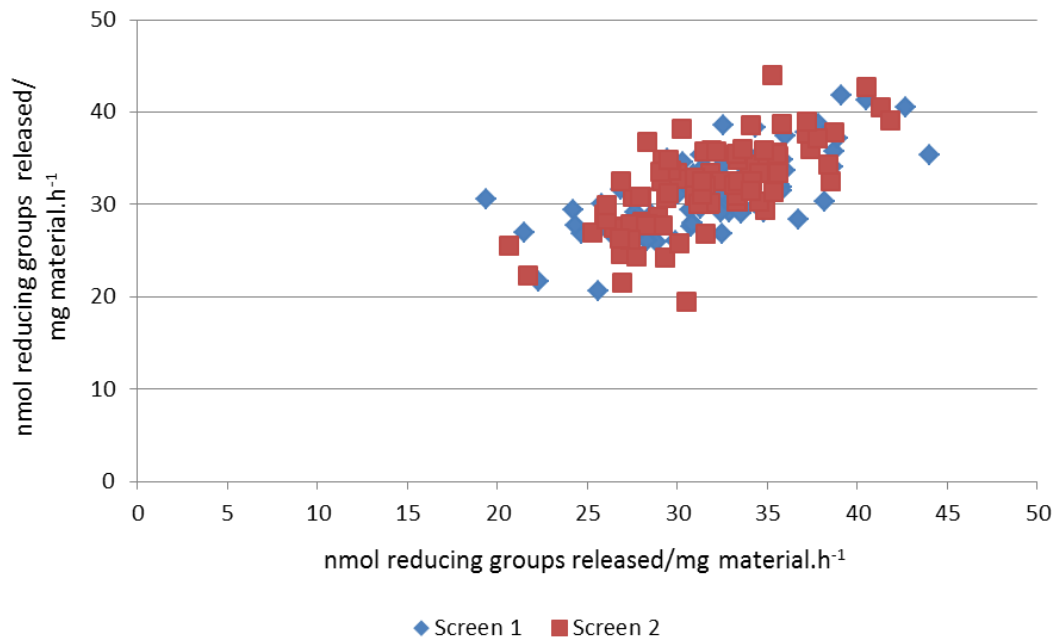
Figure 2.11: The mean distribution of the saccharification results of the Brachypodium population (Bd21 x BdTr1-1f) which was grown at INRA-Versailles, France. The ground stem internodes underwent digestion with a commercial cellulase for 8 hours at 50 ºC following a pretreatment of 0.5 N NaOH for 30 minutes at 90º C. that was screened twice using a 0.5N NaOH pretreatment. The results are the means of two replicates.

2.4 Conclusions

The aim this study was to analyse natural Brachypodium accessions to determine if there is a difference in their lignocellulosic digestibility due to a variation in cell wall composition. This information could then be used to develop specific populations to be implemented in future studies to further develop our current knowledge of secondary cell wall biosynthesis. This knowledge could lead to improvements in biofuel production. Brachypodium was chosen as the model plant for grasses instead

of using Arabidopsis as monocots have very different cell walls compared with dicots.

Natural variation within Brachypodium has been study by Pacheco-Villalobos *et al* (2012) in terms of root system architecture. They noted that there was variation present within some accessions and it would be useful to develop this research further to understand the genetic causes. Lou *et al* (2011) successfully showed that drought tolerance in Brachypodium is variable within a group of natural accessions and this was due to genotypic differences and not due to geographic location. Finally, natural accessions have also been used to study flowering time and vernalisation in Brachypodium (Schwartz *et al* 2010). Again, differences were seen amongst the accessions in both the vernalized and non-vernalised samples.

Even with these reports on natural accession variation in Brachypodium there doesn't appear to be any studies on whether there are significant differences in terms of secondary cell wall composition. However, some work has been conducted in Miscanthus looking at different cell wall components. In the first case, natural accessions were studied to understand the influence of hemicellulose composition has on digestibility (Li *et al* 2013) and the other cases focused on either cellulose (Zhang *et al* 2013) or lignin (Li *et al* 2014). All three cases were able to determine a possible target for further research to improve digestibility by possible cell wall modification such as changing the degree of arabinose substitution of xylan or the S/G ratio in lignin. It was therefore hoped that if there were difference in the cell wall composition of the Brachypodium natural accessions that they too could be exploited to gain a better understanding of cell wall recalcitrance in grasses and therefore add to the current knowledge of cell wall biosynthesis.

Lignocellulose digestibility of the natural accessions was tested using a saccharification analysis protocol involving either a hot water or a mild alkaline pretreatment of sodium hydroxide. From this analysis it was determined that both treatments indicated a variation in digestibility potential between the accessions. As expected the sodium hydroxide pretreatment resulted in more reducing groups being released than simple hot water pretreatment as this is seen in other studies in Brachypodium (Gomez *et al* 2008) as well as other lignocellulosic material, such as

sugarcane bagasse (Yu *et al* 2013). This is probably due to alkaline pretreatments being able to remove all the lignin present as well as the acetyl and uronic acid substitutions found in hemicellulose whereas hot water pretreatments only remove hemicellulose and a small amount of lignin (Kumar *et al* 2009, Harmsen *et al* 2010, Chaturvedi *et al* 2013). It is notable that some of the natural accessions (Est0 and Esp1) had a higher digestion potential when the hot water pretreatment was used instead of the sodium hydroxide pretreatment, therefore these accessions may be worth studying further to understand the possible reason for this. The increased digestion of Est0 and Esp1 with a hot water pretreatment does not appear to be related to the amount of lignin present as they both contain relatively high levels of lignin. However, the lignin composition may have an effect on digestion as Li *et al* (2010) has reported that Arabidopsis plants that have a high concentration of S lignin showed an improvement in digestibility when a hot water pretreatment was implemented.

Initially ATR-FTIR was used to undertake cell wall composition analysis of the natural accessions. It is a crude method which can highlight differences and thereby indicate the direction in which to take for more in-depth analysis. When all the data was compared using PCA it was observed that there was one main cluster and three smaller clusters therefore indicating that the majority of the accessions were similar in composition. When the outliers were analysed together and the spectra were analysed it was determined that there were differences in lignin, hemicellulose and sugar composition. Therefore not a single cell wall component stood out as the main influencer of saccharification.

Lignin plays an important role in keeping the structure of tissue, helping maintain plant hydrophobicity as well as preventing fungal and bacterial degradation (Thomas *et al* 2010). It is able to do provide protective function due to its complex structure as well as forming cross linkages with other carbohydrates, which limits the accessibility of hydrolytic enzymes (Timpano *et al* 2014, Ho-Yue-Kuang *et al* 2015).

Initially, lignin's role in recalcitrance was studied due to its effect on digestibility of corn, which is used for animal feed. It was shown that lignin has a negative correlation with digestibility (Krakowsky *et al* 2005). Since then there have been a

number of articles involving Brachypodium as well as other grass species reporting that a reduction in lignin results in an increase in saccharification potential. For example, the *bm* mutants and specifically the BdCOMT6 mutant shown a decrease in lignin, which results in an increase in digestibility (Dalmais *et al* 2013, Ho-Yue-Kuang *et al* 2015). In the case of maize stover and *Medicago sativa* (Alfalfa) a number of lines have been created each containing a decrease in a different polyphenolic pathway enzyme. Studies of these lines have all linked a decrease in enzyme to the presence of less lignin and therefore and increase in saccharification (Vanholme *et al* 2008, Chen *et al* 2009). However, more recently research is starting to show that lignin is not always the major limiting factor in digestibility. For example, Timpano *et al* (2014) showed that a mutant Brachypodium plant contained an increase in lignin content however it still had an increase in saccharification.

Due to these reports it was decided to compare the Klason lignin content of the Brachypodium accessions to the saccharification data for each of the lines. It was determined that there was no correlation between the two. Therefore, lignin content does not appear to be the main contributing factor to the variation in digestibility that is observed in the natural accessions. It is possible that only when the polyphenolic pathway genes are significantly disrupted they then have more of an effect on lignin accumulation and therefore affect digestibility more extensively. It has been reported that changes in S/G monomer ratios can increase saccharification as can the presence of more H units (Vega-Sanchez *et al* 2010, Marriott *et al* 2016). Therefore, future work could look at whether there is a variation in lignin composition within the natural accessions to determine if this has more of an influence on digestibility than lignin content.

Another important area to focus on during cell wall studies are the sugars that are found as either cellulose or hemicellulose. Both these fractions of the cell wall can influence recalcitrance. Cellulose is found in the cell wall as a crystallised structure which can prevent enzymes as well as water, which is important for hydrolytic reactions, from entering and thereby reducing digestibility (Marriott *et al* 2016). It has been suggested that an increase in total cellulose content can lead to an increase in digestibility in the case of tension wood of *Populus* (Andersson-Gunneras *et al* 2006) however this is not always the case as seen in Xu *et al* 2012 who compared

cellulose levels and biomass digestibility in Miscanthus. In the same study, Xu *et al* (2012) reported that cellulose crystallinity was a significant negative factor that affects saccharification. This has also been reported by Wu *et al* (2013) in a study on wheat accessions and rice mutants. They went further to say that it was possible that the arabinose substitutions of xylan caused this disruption to the crystal structure of cellulose. It is thought that this disruption occurs due to the arabinose sidechains of xylan interlinking with the β-1,-glucan chains of the cellulose via hydrogen bonds, which then reduce the cellulose crystallinity (Li *et al* 2013). Therefore, increasing hemicellulose content can increase the saccharification potential of biomass. Digestibility can also be improved by changing the composition of hemicellulose as seen in *irx* mutants of rice, which exhibit altered xylan biosynthesis (Chen *et al* 2013). Manipulating the hexose/pentose ratio of hemicellulose can affect the amount of usable sugars available for fermentation as pentoses cannot be fermented as efficiently as hexoses (Loqué *et al* 2015).

The sugars released during saccharification were analysed to determine if there were differences between the natural accessions. It was determined that there was some variation between the accessions with BdTr3.1a showing a relatively high amount of xylose and mannose released. This may indicate an improved accessibility to the hemicellulose fraction or an increase in the amount of these monomers found within the cell wall as the main building block of hemicellulose in grasses is xylan, which consists of a backbone of xylose residues and another component of hemicellulose is mannan, which consist of a mannose backbone (Scheller *et al* 2010).

There was also a group of accessions that showed a relative increase in the amount of glucose and xylose released when related to the other accessions. Further work needs to be conducted to determine the reason for this but it is possibly due to differences in hemicellulose composition. Another interesting observation is that the group of accessions with the highest release of glucose and xylose were predominantly tetraploid accessions. How this affects cell wall composition is unclear however it has been reported that polyploidy can influence developmental process and cause changes in plant physiology (De Storme *et al* 2014) therefore this could be interesting to follow up in the future.

Finally, the silica content of the various accessions was analysed and compared with the saccharification data to determine if there was a correlation due to previous work in ruminants reporting the negative effect silica has on digestibility of forage grasses (Agbagla-Dohnani *et al* 2003). However, in this study no correlation was detected even though there was a variation in the amount of silica present within the accessions. This could be due to the fact that the Brachypodium accessions contained relatively low levels of silica compared to some grasses. This data can still be used in future work to select those lines with differing silica content to conduct further studies focused specifically on silica in Brachypodium.

Due to the variation observed in cell wall composition amongst the natural accessions it was possible to select specific lines to create a RIL population for the future analysis of digestibility. The lines selected were Bd21 and BdTr1.1-1F, which were crossed to create a F_2 population. This population then underwent saccharification analysis and it was determined that there was sufficient variation within it for continued development into a RIL population. This population can be used in future studies such as QTL analysis to detect genes responsible for the recalcitrant nature of lignocellulosic material once it is at the F_{6-8} generation.

This study shows that there is a natural variation present, in terms of cell wall composition and digestibility, amongst the different natural accessions of Brachypodium. This knowledge could therefore be exploited in future research targeting cell wall biosynthesis and the recalcitrant nature of lignocellulosic material.

Chapter 3: QTL analysis of a *Brachypodium distachyon* RIL population (Bd3.1 x Bd21) to identify genes involved in cell wall digestibility.

3.1   Introduction

Plant cell wall digestibility has been an area of increased interest due to the growth in demand for sustainable biofuels. Currently, the cost of second generation biofuels is not competitive with fossil fuels because of the high demand of energy and expensive enzymes needed during production (Gomez *et al* 2008, Carroll *et al* 2009). It is thought that improvements can be made to biomass to make it more digestible in the future by developing our understanding of the biosynthesis of lignocellulosic material and the genes involved in creating this recalcitrant material.

There are two general methods to identifying and understanding genes involved in particular processes, namely forward and reverse genetics approaches. Forward genetics is based on trying to determine the genes responsible for a specific phenotypic effect. This is achieved by screening a population for natural or induced variation in a trait and mapping this to a specific locus underpinning a particular change in phenotype. In contrast, reverse genetics employs targeted gene disruption in order to test a hypothetical role for a gene using transgenesis (such as the use of T-DNA insertional mutant plant lines), genome editing or molecular screening. In this study it was decided to use Quantitative Trait Loci (QTL) analysis as a forward genetic approach to identify genes that have a measurable impact on the digestibility of lignocellulosic material. The advantages of this method of analysis are that it is not limited by current knowledge about plant cell wall digestibility; it can identify a number of diverse genes; and it is able to deal with complex traits that don't follow typical Mendelian inheritance as they are the result of the interaction of multiple genes (Doerge 2002). Digestibility is an example of a complex trait.

QTL analysis is used to study traits that are controlled by a number of genes, which tend to be influenced by the environment (Doerge 2002). This type of analysis involves the identification of regions within the genome responsible for the

differences observed within quantitative phenotypes (Doerge 2002, Borevitz *et al* 2004). Phenotypes are the observed physical attributes of an individual which are due to the interaction between the genetic composition, or genotype, of the individual and the environment. This is achieved by measuring the link between the inheritance of genetic markers and phenotypic variation within a segregating population. A marker will only be linked to a QTL if it is inherited together with the phenotype otherwise the marker will show independent segregation within the population (Collard *et al* 2005). It is important to understand the contributions of genetic and environmental components to observed phenotypic variation, and to have the capability of measuring the trait of interest with sufficient sensitivity to achieve this. The genetic cause of variation will not be identified if there is too much variation caused by non-genetic factors (Borevitz *et al* 2004, Murry *et al* 2008, Oakey *et al* 2013). The most commonly used approach for QTL analysis by association mapping employs populations of RILs derived from a cross between parental lines showing differences in the trait of interest. To improve the chances of successfully identifying QTLs during analysis there are a number of considerations to take into account when designing an experiment. Most importantly, the trait that is being studied should show measurable differences in the parental lines, secondly the same plant lines that are used to construct the genotype data must be used for the phenotypic analysis (Collard *et al* 2005). For QTL analysis to be successful the non-genetic variation observed needs to be as small as possible so that the genetic variation is not swamped by it (Murray *et al* 2008, Oakey *et al* 2013). Environmental effects can be reduced by increasing the number of population replicates studied as well as repeating them over time. The size of the population can also have an effect as it has been shown that screening larger populations is more successful at detecting phenotypes with smaller effects (Doerge 2002, Collard *et al* 2005). The outcome of the study can also be improved by using markers that are closely linked to genes as well as having a map densely packed with markers (Collard *et al* 2005, Semagn *et al* 2006).

In this study a RIL population of Brachypodium was used as it is has been shown to be a successful model plant for agriculturally important grass crops (Bevan *et al* 2010, The International Brachypodium Initiative 2010). An advantage is that its genome has been fully sequenced and assembled (The International Brachypodium

Initiative 2010). This can be used to help identify possible candidate genes that are found within the region of detected QTLs (Borevitz *et al* 2004). A mapping population of plants is needed for QTL analysis. In this case the RIL population Bd3.1 x Bd21 was chosen instead of producing one specifically for the project as it takes a number of years to create. RIL populations are created from homozygous, inbred parental lines that are crossed to generate heterozygous F_1 lines. These lines are allowed to self-pollinate over $6 - 8$ generations to produce a F_n population that consists of fully homozygous lines, which consist of approximately 50% of each parent in different combinations (Doerge 2002, Fazio *et al* 2003). The Bd3.1 x Bd21 population was generated by David Garvin's laboratory to identify the barley stripe mosaic virus resistant gene *bsr1* (Cui *et al* 2012. The population contains 165 plant lines and a genetic map of 768 SNP markers over five linkage groups. This equates to a marker density of 3.0 cM/marker on the genetic map or 475.4 kb/marker on the physical map. The recombination rate is therefore 6.2 cM/Mb (Cui *et al* 2012). The advantage of using a RIL population is that it is homozygous true breeding so it can be repeatedly grown without a change occurring within the population's genotypic makeup. RIL populations can be used together with both dominant and co-dominant markers for genotyping. Co-dominant markers can distinguish between both homozygous and heterozygous genotypes whereas dominant markers can only detect homozygous genotypes. However, because RIL populations are homozygous true-breeding both dominant and co-dominant markers are found only in the homozygous state thereby providing the same amount of information. This homozygous only state of marker alleles leads to a reduction of experimental error as well as improved accuracy in determining environmental and genetic relationships. Finally, the absence of heterozygous alleles means that there are only two genotype classes thereby simplifying genetic segregation but this does mean that the effect caused by dominance cannot be assessed (Van der Schaar *et al* 1997, Fazio *et al* 2003).

The aim of this study was to look at the variation in saccharification within the RIL Brachypodium population (Bd3.1 x Bd21) so that QTLs linked to digestibility could be detected, which may lead to the identification of candidate genes that impact on straw digestibility.

3.2 Material and methods

3.2.1 Preparation of plant material

Seeds from a RIL population (Bd3.1 x Bd21) were obtained from David Garvin (Cui *et al* 2012) and grown at INRA-Versailles, France as two randomised replicates (Block D and E). Three seeds per plant line were sown in a 14 cm pot before placing in the glasshouse. The plants were not vernalized. A total of 146 lines were successfully harvested from Block D whereas only 127 lines were harvested from Block E (Appendix A). The dried plant material, together with the seeds, was sent to the University of York for saccharification analysis.

The seeds from Block D were sown at the University of York as three randomised replicates (Blocks 1 – 3). There were four seeds per well of a P15 tray, thereby producing a total of 12 plants per line. The 30 trays per block were placed on a single bench in the glasshouse. The trays were laid out so that Block 2 was closest to the door and Block 3 was closest to the back windows, which left Block 1 in the middle. Before putting the trays in the glasshouse they spent three weeks at 4 ºC to vernalize. After about three weeks in the glasshouse the plants were staked to help support the stems. Watering was stopped when the plants started to senesce. After approximately two weeks, when they were completely dry, they were harvested (Appendix B).

3.2.2 Plant morphology

During germination the number of seeds that developed for each plant line within each replicate was noted so that the germination frequency could be calculated. The amount of biomass found in each plant was determined after harvesting by weighing the material as a whole plant (total biomass) or just the stem when the inflorescence and leaves had been removed. The height, excluding the inflorescence, of each plant was also measured.

3.2.3 Saccharification analysis

Plant material from INRA-Versailles, France (Block D and E) and the University of York (Block 1 – 3) was screened twice to determine the saccharification potential of the various plant lines within the RIL population. The stem material was prepared for analysis by removing the top and bottom internodes as well as all nodes. In the case of the plants grown at INRA-Versailles all the stems were used but only the main stem was selected from the plants grown at the University of York. The stem material was prepared and screened using the grinding and weighing robot as well as the liquid handling robot with a sodium hydroxide pretreatment as described in Chapter 2 (Gomez *et al* 2010, Gomez *et al* 2011).

3.2.4 Quantitative trait loci analysis

The QTL analysis was conducted using the method described in "A Guide to QTL Mapping with R/qtl" (Broman and Sen 2009). The genotype data for the RIL Bd3.1 x Bd21 population was supplied by David Garvin (Cui *et al* 2012).

The saccharification data was analysed using ANOVA to determine if there were any environmental effects such as plate position, block position, run day and well position. It was determined that well position had the most effect and it was corrected for using a correction co-efficient that was calculated from running four plates containing only filter paper discs. A correction coefficient was also included for the weight of the sample to remove any from the analysis that contained too little or too much material. Once these corrections were completed the saccharification data and the genotype date was assembled into a large Microsoft Excel table in the format required by the program. This table was then imported into R/qtl as a csv (comma-delimited) file.

The R/qtl program was installed using the command install.packages("qtl") and the R/qtl package was loaded using the command library(qtl) before the csv data file was imported. Next the algorithm used for the analysis of a RIL population was added using the command data=convert2riself(mydata).

Standard interval mapping was performed to identify loci with important marginal effects using a genome-wide scan. This was achieved by calculating the conditional genotype probabilities which were used to produce a plot. The significant threshold was then determined using a permutation test for 1000 replicates and was displayed as a 5% LOD score which was used to pick out peaks from the plot which reached this threshold. The LOD score determines the probability that the effect being detected is in fact a QTL. A 5% LOD threshold relates to a 95% probability that a QTL has been detected at that position on the chromosome. The effect of the detected QTLs was obtained by creating an effect plot and the output from the plot was recovered. The function fitqtl was used to determine the fit of the model using 128 imputations with a 1 cM grid and thereby calculating the genetic variance of the detected QTL.

Finally, epistasis was analysed by running the calculation which produces a plot of interactions. The data from this can be observed as well as effect plots can be created.

The proportion of the phenotypic variance that is described by the QTL is known as heritability. The heritability of the QTL analysis of Blocks $1 - 3$ was conducted as broad-sense (H^2) heritability from the value of the mean squares of the RILs (Parker *et al* 1998, Broman and Sen 2009). The H^2 was calculated as follows:

$$H^2 = V_G/V_T$$
$$= V_G/(V_G+V_E)$$

Were, V_G is the genotype variance, V_E is the environmental variance and V_T is the total variance of the trait of interest.

3.3   Results and discussion

3.3.1 Preliminary experiment

A preliminary experiment was conducted to determine if the planned experimental methodology was optimal or if further optimisation of the system was necessary. The RIL population Bd3.1 x Bd21 was grown at INRA-Versailles as two randomly replicated blocks (D and E) and the dry material was harvested and sent to the University of York for saccharification analysis. The parental lines were analysed first to determine if there was a difference in their digestibility (Figure 3.1). Initially, it appeared that parental line Bd3.1 was more recalcitrant because slightly less sugar was released when compared with Bd21. However, when the lines were compared statistically this was shown to not be the case. The Mann-Whitney test, the non-parametric version of the two-sample t-test, indicated that there was no significant difference in digestibility between the two parents ($W = 5$, $p = 0.151$).
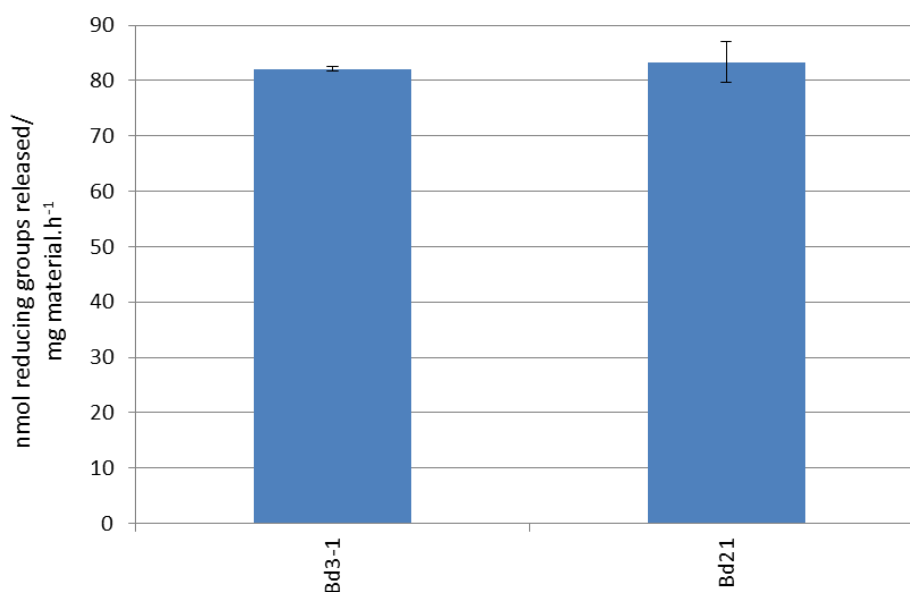
Figure 3.1: Saccharification analysis of the ground biomass from the two parental lines used to create the RIL population, which were grown at INRA-Versailles, France. The ground material from the stem internodes underwent digestion with a commercial cellulase for 8 hours at 50 ℃ following a pretreatment of 0.5 N NaOH for 30 minutes at 90 ℃. The results are the means and standard deviations of five replicates.

The 146 plant lines from Block D and 127 lines from Block E that were successfully harvested (Appendix A) were each analysed twice. The average amount of sugar

released from the plants grown in Block D was 62.73 nmol reducing groups released/mg material.h^{-1}, which is nearly double the amount, released from the plants grown in Block E (36.54 nmol reducing groups released/mg material.h^{-1}). The distribution of the results obtained is shown in Figure 3.2, which indicated that there is some variation between the results obtained from the two blocks. This was statistically tested using the non-parametric form of the two-sample t-test. The Mann-Whitney test indicated that there was a significant difference between Block D and E (*W = 13689, p < 2.2 x 10^{-16}*). This difference between the two blocks could be due to their positioning within the glasshouse thereby leading to the environment having an effect on their growth and subsequent digestibility (Oakey *et al* 2013). The most digestible RIL in Block D was L098 (79.26 nmol reducing groups released/mg material.h^{-1}) whereas in Block E it was L047 (49.97 nmol reducing groups released/mg material.h^{-1}). It was also noted that the ranking of the top 10 and bottom 10 lines according to saccharification potential did not remain consistent between the two blocks. From these results it was noted that further optimisation in growth and sample collection conditions would be necessary to decrease the variation observed between the replicates.
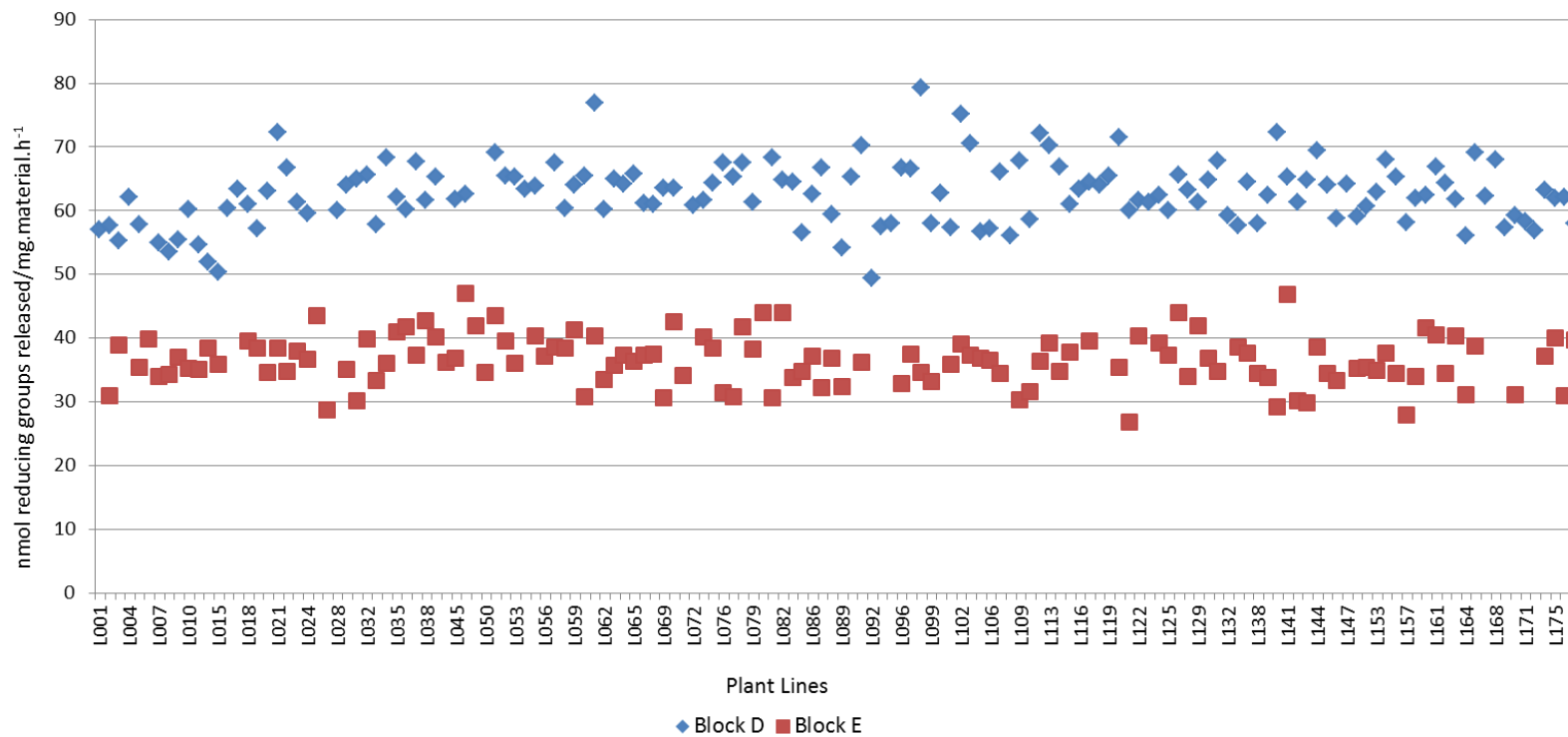
Figure 3.2: The mean saccharification distribution and ranking of the Brachypodium RIL plant lines that were grown at INRA-Versailles, France. The ground material from the stem internodes underwent digestion with a commercial cellulase for 8 hours at 50 °C following a pretreatment of 0.5 N NaOH for 30 minutes at 90 °C. The results are the means and standard deviation of two replicates, Block D and E, which were analysed twice.

QTL analysis was undertaken using the saccharification data from Block D for both rounds of analysis (Screen 1 and 2) was used together with the genotype data for the RIL population. It was decided to analyse the replicated runs (Run 1 and 2) individually because of the large variation between them. Figure 3.3 displays the results obtained from the analysis after the data had been corrected to minimise for any environmental variation caused by the position of the sample within the 96-well plates. However, even with this correction there was still too much environmental variation. This variation led to high background noise being present, which prevented any of the peaks from reaching the LOD 5% threshold of 3.2. Therefore, we were not successful in identifying any possible QTLs linked to digestibility (Oakey *et al* 2013).
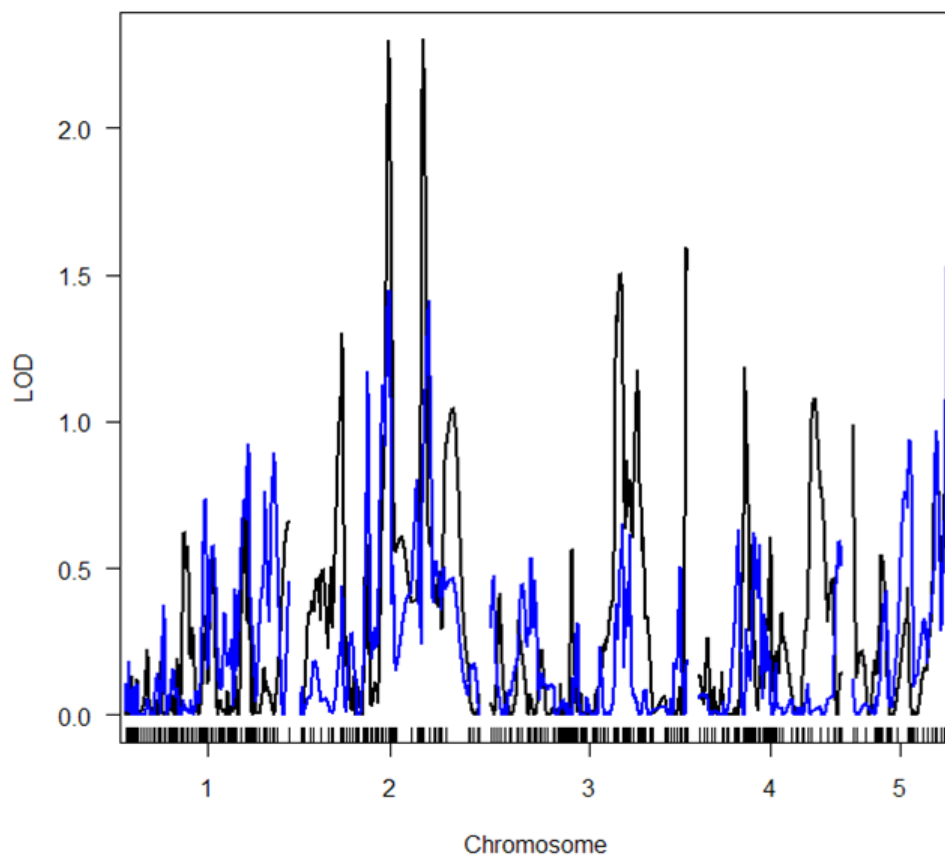
Figure 3.3: QTL analysis of the Brachypodium RILs grown in Block D at INRA-Versailles, France. The analysis was undertaken using the average saccharification data obtained from the screening of the material in duplicate (Black = Run 1, Blue = Run 2). The LOD 5% threshold was calculated as 3.2.

3.3.2 Repeating the experiment at York

The experiment was repeated at the University of York with improved conditions to decrease the environmental variation. In the growing phase it was decided to increase the number of replicates grown, therefore increasing from two to three blocks of plants. This would help to identify any blocks that were outliers. A record was also kept of where each plant line was grown within a randomised block.

Plant morphology data (germination frequency, plant height and biomass) would also be kept so that it could be noted if any environmental influences affected the replicates individually during this stage. It was hoped that if there was any significant difference between the blocks in terms of saccharification potential it could be accounted for by any environmental changes that could affect growth and be seen in other morphological traits.

During sample collection only the main stem of each plant to be screened was selected for saccharification analysis thereby reducing variation caused by including multiple stems.

Before QTL analysis took place the data was treated with correction coefficients for the position of the sample within the plate as well as for the weight of the sample within each well. These corrections therefore take into account any variation that may have been introduced into the system during the laboratory phase of the experiment.

3.3.2.1 Plant morphology

Twelve seeds were sown for each plant line in each block. Therefore, germination was determined by counting the number of seeds that developed into full sized, healthy plants.

When comparing the germination frequency of the two parental lines it was observed that Bd3.1 had a better germination rate at 99% than Bd21 at 63% (Figure 3.4). This

was shown to be statistically significant when using the Mann-Whitney test ($W = 3.25, p = 0.017$). However, there is a lot of variation within Bd21 as indicated by the standard error.
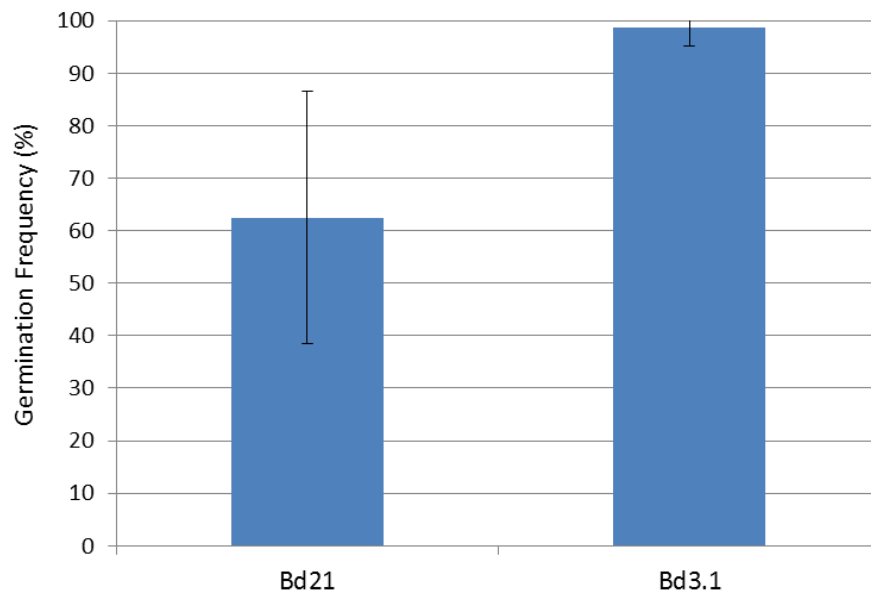
Figure 3.4: Comparison of the germination frequency between the two parental lines used to produce the RIL population. The results are the means and standard deviations of 72 replicates grown over the three experimentally replicated blocks.

The distribution of the germination frequency of the plant lines within the separate blocks was analysed (Figure 3.5). It appeared to be very similar for all three blocks with the majority of plant lines having a germination frequency of $80 - 100\%$. The only difference in the distribution is the small peak at 40% within Block 1 that was caused by 10 lines. There were also a few outliers that had a consistently low germination frequency across all three blocks, L001 (40%), L032 ($< 25\%$), L074 ($< 50\%$) and L119 ($< 25\%$).
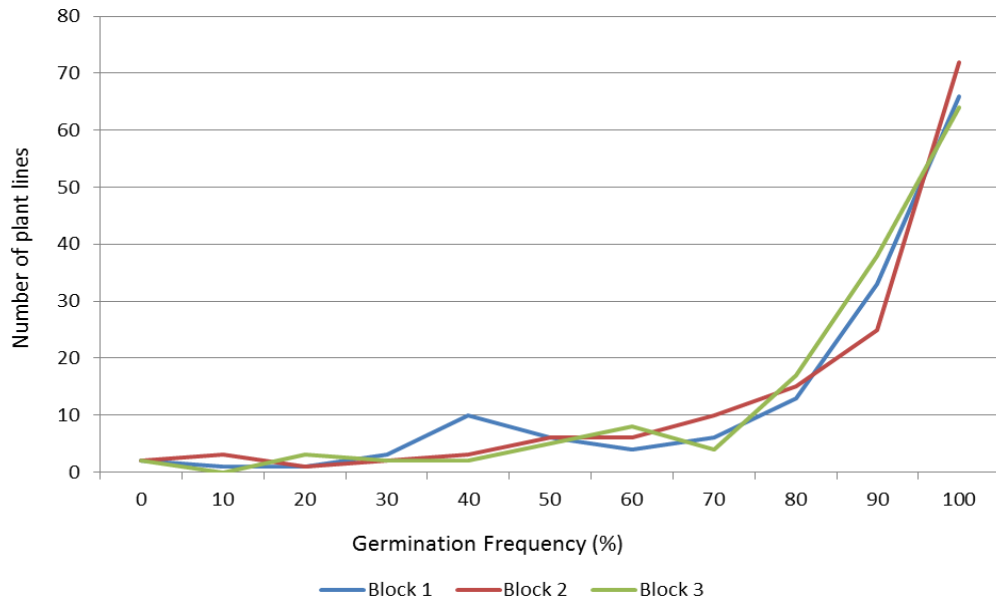
Figure 3.5: The distribution in germination frequency of the RILs grown in the three replicated blocks. The results are the means and standard deviations of 12 replicates grown per block.

The height of the plants was measured after harvesting once the inflorescence had been removed. In Figure 3.6 the parental lines were compared and it was observed that the parental line Bd3.1 was taller (33.4 cm) than Bd21 (26.03 cm). This difference in height was statistically significant (two-sample t-test: *t = -4.538, df = 10, p = 0.001*). Tyler *et al* 2014 reported similar heights for Bd3.1 (40 cm) and Bd21 (28 cm) though their measurement included the seed head.
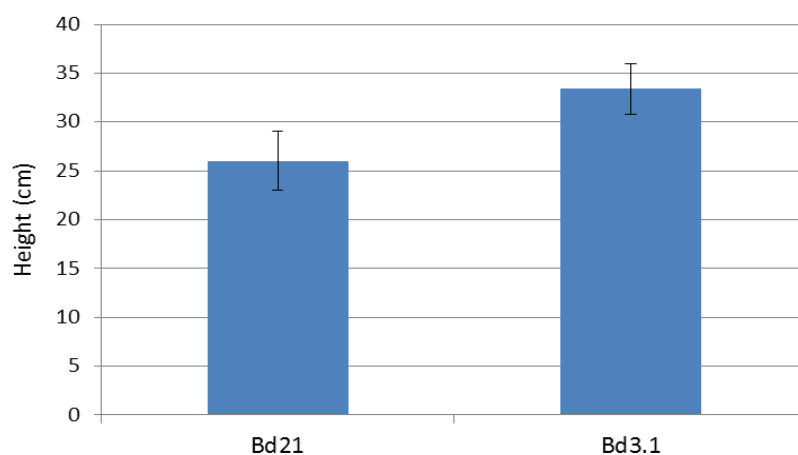
Figure 3.6: Comparison of the height of the plant between the two parental lines used to produce the RIL population. The height of the main stem excluding the seed head was measured. The results are the means and standard deviations of 72 replicates grown over the three experimentally replicated blocks.

79

The distribution of the RIL population within the three blocks with regards to the height of the plants was compared (Figure 3.7). The three blocks had a similar distribution with only Block 2 being slightly different as its distribution was shifted higher with its peak being at around 30 cm whereas Block 1 and 3 had the majority of its plants reaching a height of approximately 25 cm. This is confirmed statistically when reviewing the mean height for each of the blocks (Block 1 = 28.06 cm, Block 2 = 31.66 cm and Block 3 = 28.25 cm). Block 2 also contained a couple of outliers that were unusually tall in just this case (L031 = 55 cm and L078 = 57 cm).
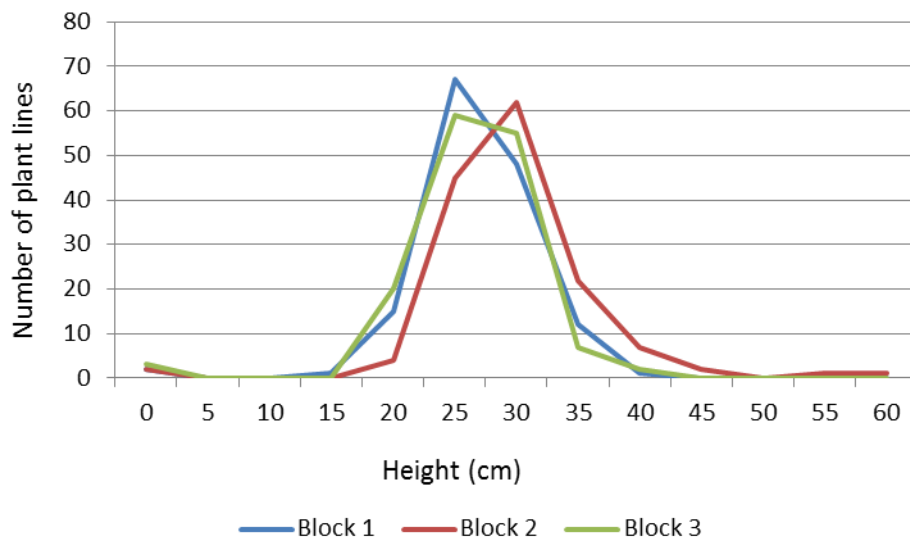
Figure 3.7: The distribution in height of the RILs grown in the three replicated blocks. The height of the main stem excluding the seed head was measured. The results are the means and standard deviations of 12 replicates grown per block.

Two measurements were taken to determine the biomass of the plant line. This first was the total biomass, which included the inflorescence whereas the second measurement was of the stem only. The parental lines were compared to determine if there was a difference between the two under the different conditions (Figure 3.8A and B). Parental line Bd3.1 had slightly more biomass with or without the inflorescence being present (0.436 g and 0.165 g respectively) when compared with Bd21 (0.380 g and 0.144 g). Under both conditions this difference in biomass proved to not be statistically significant (two-sample t-test: total biomass $t = 0.707$, $df = 10$, $p = 0.496$ and stem only biomass $t = 1.200$, $df = 10$, $p = 0.257$).
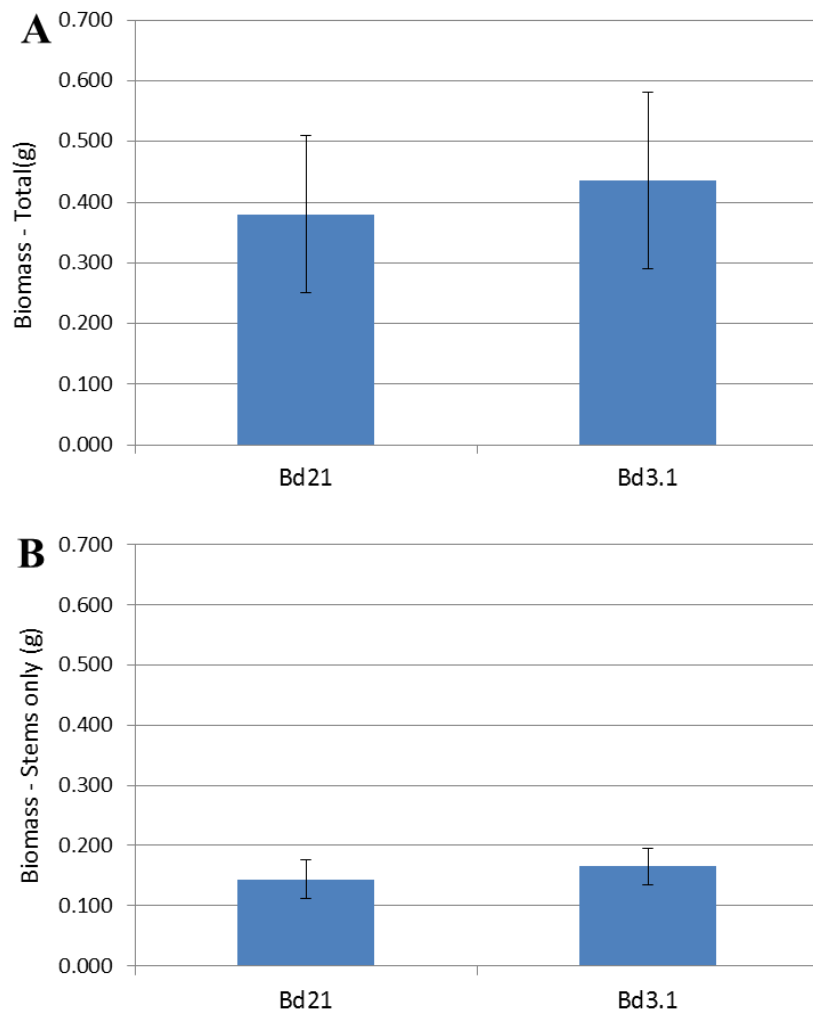
Figure 3.8: Comparison of biomass present within the parental plants as either (A) total or (B) stem only biomass. The entire aerial portion of the plant was weighed for the total biomass measurement whereas only the main stem without the inflorescence was weighed for stem only. The results are the means and standard deviations of 72 replicates grown over the three experimentally replicated blocks.

The distribution of the total (Figure 3.9A) and stem only (Figure 3.9B) biomass of the RIL population for each of the blocks was analysed to determine if there were any differences between the replicates. When the total biomass was analysed it was observed that Block 1 had less biomass (mean = 0.291 g) than Block 2 (mean = 0.532 g) and Block 3 (mean = 0.482 g). The differences between the blocks were not as noticeable when the stem only biomass was analysed (mean: Block 1 = 0.127 g, Block 2 = 0.176 g and Block 3 = 0.153 g). This difference in biomass within the Block 1 replicate could be due to the plants undergoing early senescence due to a heatwave that was experienced while the plants were growing. Blocks 2 and 3 were

kept cooler as they were positioned directly under the cooling system in the glasshouse. This early senescence may have resulted in there being less time to fully produce seeds, thereby resulting in less seed biomass. It has been reported in the past that high temperatures after anthesis negatively affect the development of seeds because it results in a shorter grain growth period, which causes low grain weight as well as other poor physical and physiological properties (Elgersma *et al* 1993, Grass *et al* 1995).
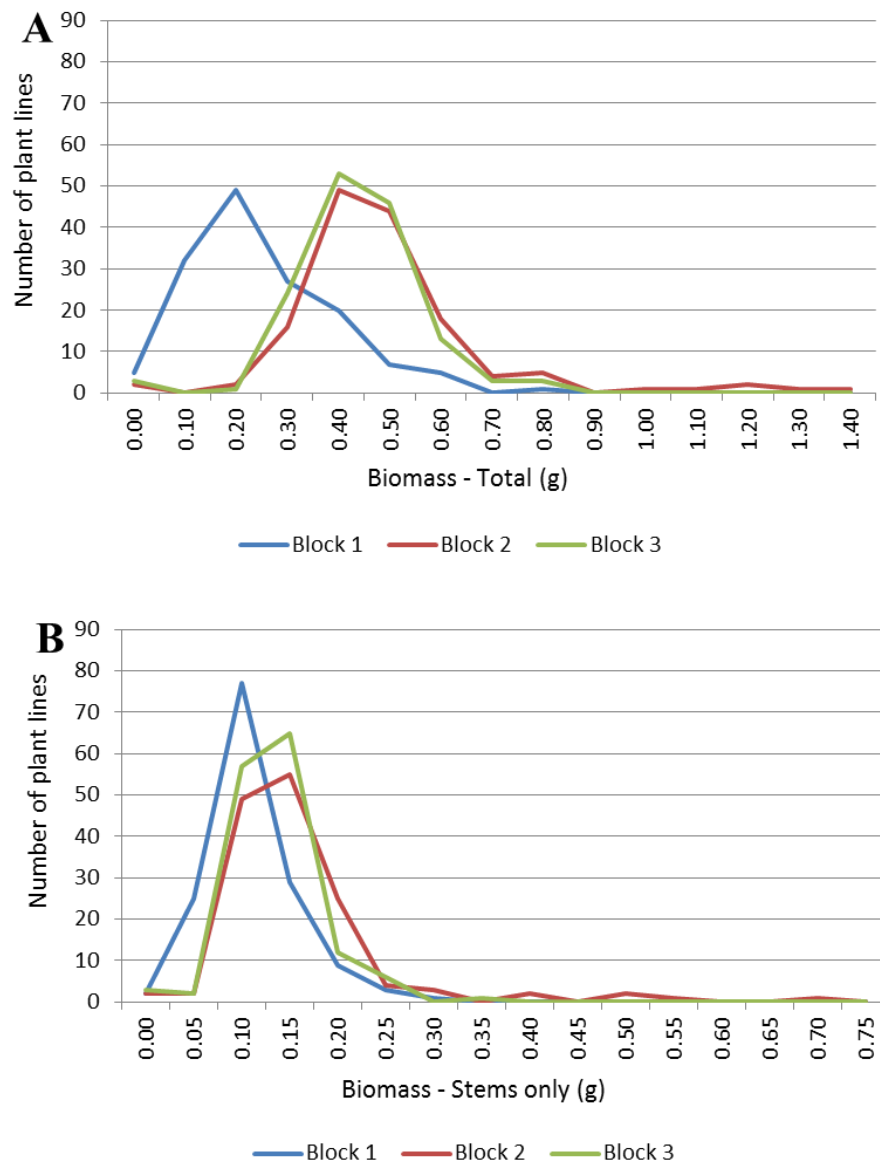
Figure 3.9: The distribution in the amount of biomass (A) total and (B) stem only of the RILs grown in the three replicated blocks. The entire aerial portion of the plant was weighed for the total biomass measurement whereas only the main stem without the inflorescence was weighed for stem only. The results are the means and standard deviations of 12 replicates grown per block.

From the analysis of the results obtained from the morphological study focusing on germination frequency, height and biomass it can be determined that only biomass was significantly affected by the environment.

3.3.2.2   Saccharification analysis

The RIL population that was regrown in triplicate (Blocks 1 – 3) was screened for digestibility using a saccharification method involving a mild alkaline pretreatment (Gomez *et al* 2010, Gomez *et al* 2011). A total of 144 plant lines were successfully harvested from Blocks 1 and 2 as well as 142 from Block 3 (Appendix B).

The parental lines were analysed to determine if there was a difference between their saccharification potentials. In Figure 3.10 it appears that parental line Bd21 is more digestible as more sugar was released (37 nmol reducing groups released/mg material.h^{-1}) than from parent Bd3.1 (31 nmol reducing groups released/mg material.h^{-1}). This observation was statistically supported by a two-sample t-test, which is used to compare two independent means. Statistically it was shown that the parental lines are significantly different from each other when comparing digestibility (*t = -2.992, df = 14, p = 0.01*). These results concerning saccharification potentials of the parental lines is different from what was observed originally involving the material produced in France. This change in significance is due to improved experimental conditions in terms of sampling, which means that the variation caused by the environment has been reduced. This significant difference in saccharification potential between the parental lines is also reassuring as the success of QTL analysis depends on the parental lines being divergent from each other in terms of the trait being analysed (Collard *et al* 2005).
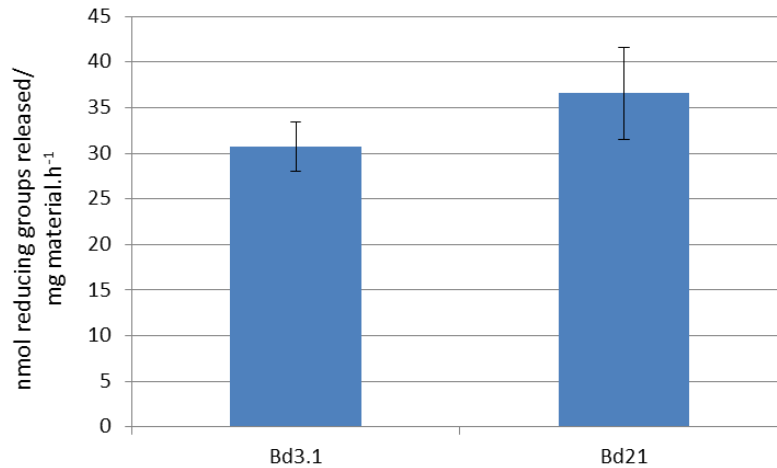
Figure 3.10: Saccharification analysis of the ground biomass from the two parental lines used to create the RIL population, which were grown at the University of York. The ground material from the main stem internodes underwent digestion with a commercial cellulase for 8 hours at 50 ℃ following a pretreatment of 0.5 N NaOH for 30 minutes at 90 ℃. The results are the means and standard deviations of 48 replicates.

Each replicate was analysed twice and the distribution of the results were plotted (Figure 3.11). The amount of sugar released from Blocks 2 and 3 were similar however Block 1 showed a lot of variation within the block as well as compared with the other two blocks. The average amount of sugar released from the plants was 46.67 nmol reducing groups released/mg material.h^{-1} for Block 1, 29.83 nmol reducing groups released/mg material.h^{-1} for Block 2 and 35 nmol reducing groups released/mg material.h^{-1} for Block 3. It is thought that the weather and the position of Block 1 within the glasshouse played a role in the observed variation. There was a heatwave when the plants were growing and this replicate was not directly under the cooling system whereas Block 2 and Block 3 were. The unusually high temperatures possibly led to this block of plants senescing earlier than the other two blocks. It was also determined statistically that Blocks 2 and 3 are significantly different (Mann-Whitney: *W = 16364, p < 2.2 x 10^{-16}*) when compared with each other.
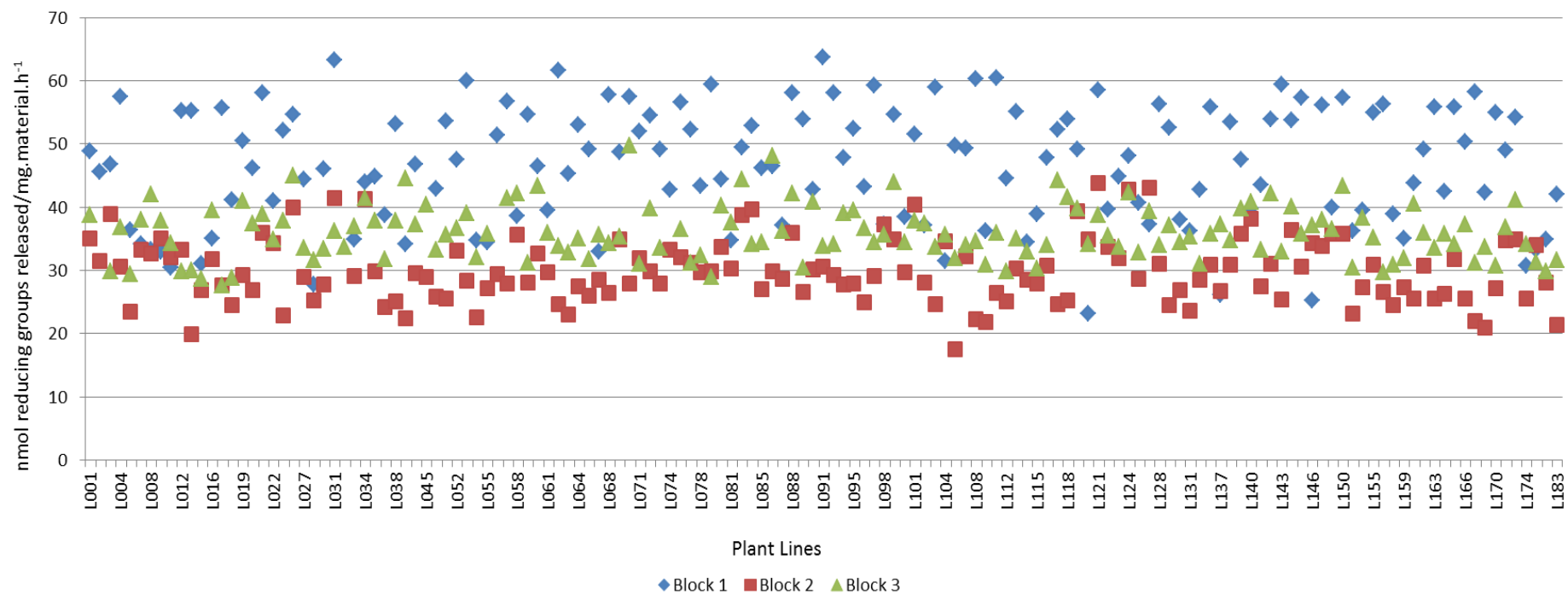
Figure 3.11: The mean saccharification distribution and ranking of the Brachypodium RIL plant lines that were grown at the University of York, UK. The ground material from the main stem internodes underwent digestion with a commercial cellulase for 8 hours at 50 ℃ following a pretreatment of 0.5 N NaOH for 30 minutes at 90 ℃. The results are the means and standard deviation of 3 replicates, Block 1, 2 and 3 containing 12 plants per line, which were analysed twice.

The rank order of all the lines for each replicate was determined using the saccharification data (Appendix C). The plant line that was the most digestible for each replicate was identified as L091 for Block 1, L121 for Block 2 and L070 for Block 3.

3.3.2.3 QTL analysis

The saccharification data was reviewed to determine the effect of environmental variation and whether it could be reduced before the QTL analysis was conducted. It was concluded that the possible areas that environmental variation could occur during saccharification analysis included:

- The day the samples were run.
- The 96-well plate the sample was on.
- The well position within the 96-well plates.
- The weight of the sample.

The data for the day the samples were run, the plate they were on and the position on the plate were all tested using ANOVA. It was determined that the day the samples were run had no significant effect and neither did the plate they were on. This was probably due to including filter paper discs on each of the plates to act as internal controls.

The ANOVA of the well position effect was determined to be highly significant ($p = 0.634$) because the well position is affected by the small variations in temperature across the plate that occurs during the heating steps. However, this environmental effect was then corrected for by using a co-efficient that was pre-determined by running three 96-well plates containing only filter paper discs and comparing the amount of reducing sugars released within each well. This error was a systematic error that occurs due to the position of the well on the liquid handling robot (Gomez *et al* 2010).

The effect of the weight of the sample in each well has to be taken into account because wells containing more than 4 mg of the sample will result in more reducing groups being released thereby causing a false indication of increased saccharification potential to occur whereas those containing less than 4 mg will result in a decrease in reducing groups released and a false decrease in saccharification potential. Therefore a correction was made to account for this environmental effect caused by sample weight, which was achieved by adjusting the saccharification value measured based on the difference between the desired sample weight of 4 mg and the actual measured weight of the sample.

Once the saccharification data had been reviewed and adjusted it was then analysed together with the genotypic data supplied by David Garvin (Cui *et al* 2012). Initially, the saccharification data from Block 1 and Block 2 were analysed together to identify any possible QTLs linked to digestibility (Figure 3.12). A single QTL was detected on chromosome 5 that surpassed the LOD 5% threshold of 3.01 (Figure 3.12A). This QTL was found to be linked to marker BD1676_1 at genetic map position 159 cM. Figure 3.12B shows the effect the alleles of this marker has on the plants, namely those containing allele AA are more digestible than those with allele BB. Further analysis was conducted and it was determined that there was a further QTL on chromosome 3 that acted in epistasis with the QTL on chromosome 5. Doerge (2002) classifies epistasis as the genetic interaction between two or more loci that have an effect on the phenotype of the trait of interest and this effect is more than the sum of the effects of the individual loci. It is important to test for epistasis for complex traits because if it is not included then bias may be introduced concerning the effect the genetic component of the trait of interest plays. In the long term, epistasis has an important influence on breeding methods. However, epistasis is not always easy to detect because genotype x environment interactions as well as linkage equilibrium can have a restraining effect. There are different types of epistasis but the simplest forms of interaction are complementary and duplicate (Tan *et al* 2001, Shiringani *et al* 2010, Shiringani *et al* 2011). In this case the QTL on chromosome 5 was found to be linked to marker BD1415_1 at position 82.5 cM on chromosome 3. Allele AA of marker BD1415_1 added to the digestibility of plants with allele AA of marker BD1676_1 (Figure 3.12C).
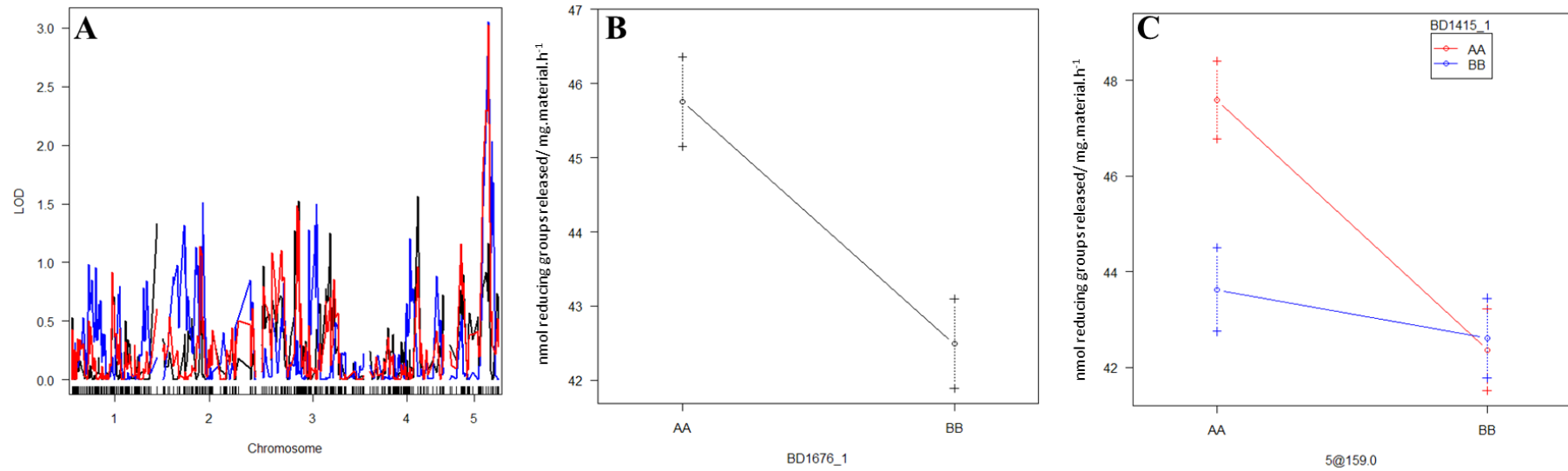
Figure 3.12: (A) QTL analysis of the Brachypodium RILs grown in Block 1 and 2 at the University of York, UK. (Black = Block 1, Blue = Block 2, Red = Average) (B) The effect of the alleles on the QTL found linked to marker BD1676_1 and (C) the effect of the alleles of the epistatic QTL linked with marker BD1415_1. The analysis was undertaken using the average saccharification data. The LOD 5% threshold was calculated as 3.01.

The analysis of Blocks 2 and 3 was then conducted to see if the same QTLs were detected as well as to determine if there were any new QTLs (Figure 3.13A). Two QTLs were detected that exceeded the LOD 5% threshold of 3.0. The first QTL detected was linked to marker BD2181_2 on chromosome 2 at 302 cM. This QTL had not been observed before and was only visible in the data from Block 3. It appeared to increase the digestibility of those plants with allele AA (Figure 3.13B). The second QTL detected was the QTL linked to marker BD1676_1 on chromosome 5, which had been observed before (Figure 3.13C). However, the epistatic QTL on chromosome 3, linked to marker BD1415_1, could not be detected in either Block 2 or Block 3.

Finally, the data from all the blocks were used together for QTL analysis. However, this time only the QTL on chromosome 5, linked to marker BD1676_1, was found to exceeded the LOD 5% threshold of 3.03 (Figure 3.14).

A possible reason for only detecting one QTL is that the population size is relatively small and therefore only major QTLs will be detected and not those that have a low effect (Parker *et al* 1999). Another possible reason is that the parental lines didn't show a large enough difference in digestibility even though it was a significant difference (Tan *et al* 2001).
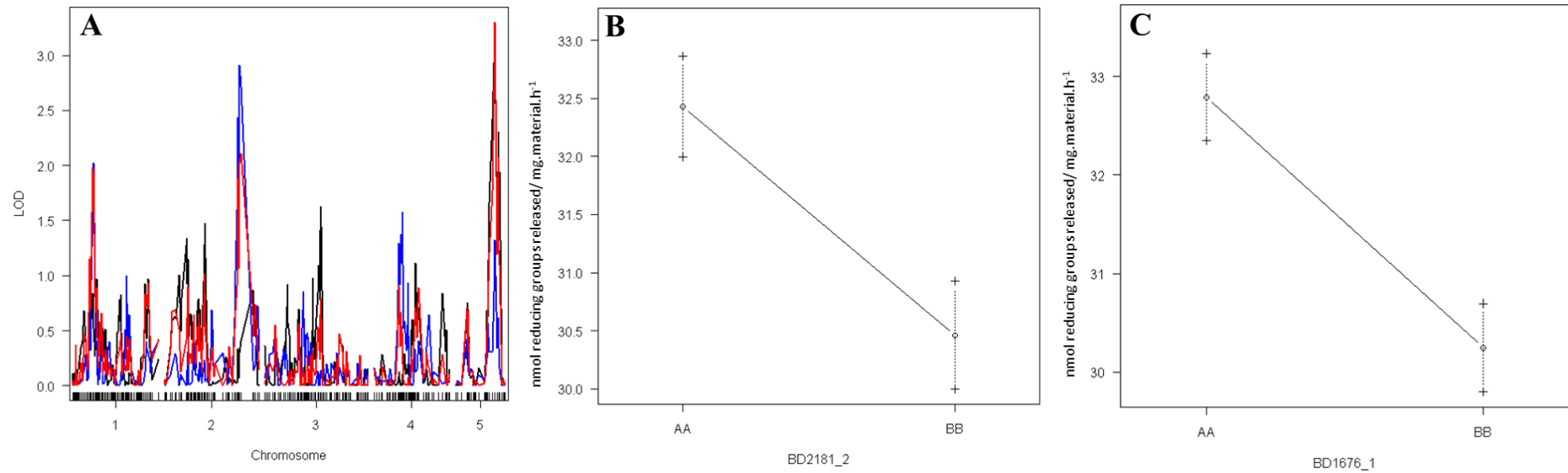
Figure 3.13: (A) QTL analysis of the Brachypodium RILs grown in Block 2 and 3 at the University of York, UK. (Black = Block 2, Blue = Block 3, Red = Average) (B) The effect of the alleles on the QTL linked to marker BD2181_2 and (C) BD1676_1. The analysis was undertaken using the average saccharification data. The LOD 5% threshold was calculated as 3.0.
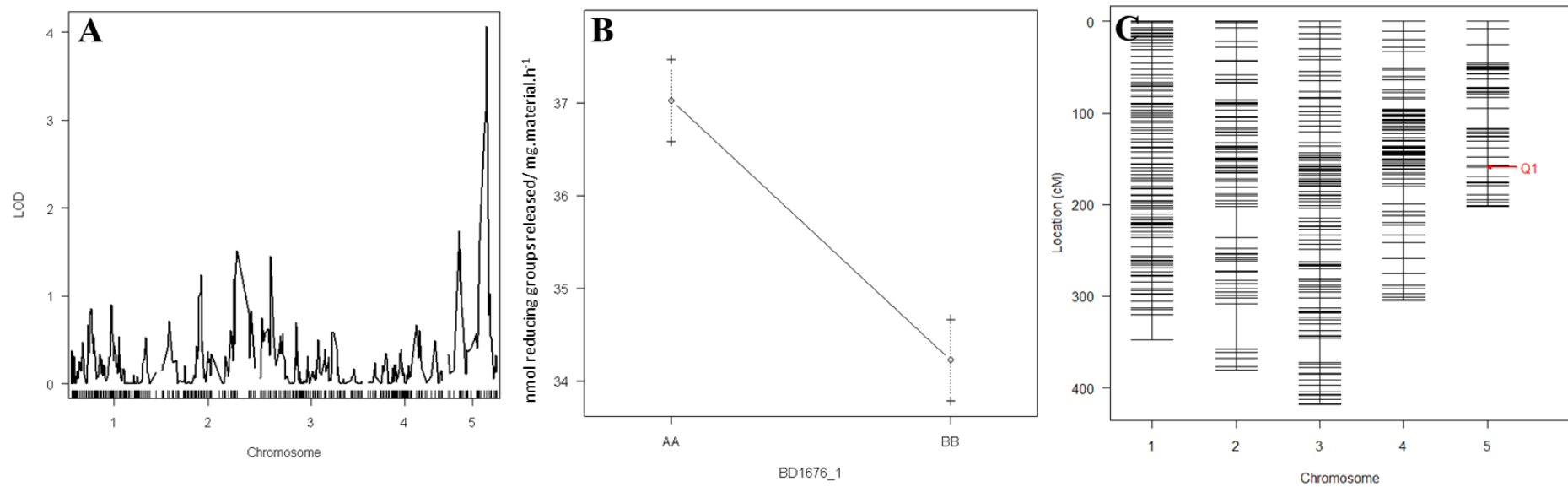
Figure 3.14: (A) QTL analysis of the Brachypodium RILs grown in Block 1, 2 and 3 at the University of York, UK. (Black = Average). (B) The effect of the alleles on the QTL found linked to marker BD1676_1 and (C) the position of the QTL on the genetic map. The analysis was undertaken using the average saccharification data. The LOD 5% threshold was calculated as 3.03.

3.4   Conclusion

A preliminary study was initially undertaken whereby the Brachypodium RIL population Bd3.1 x Bd21 was grown in duplicate at INRA-Versailles and stem samples were sent for analysis at the University of York. Saccharification analysis highlighted the fact that there was too much variation within the experiment, in terms of the amount of sugar released. This observed environmental variation within the saccharification data resulted in too much background noise during analysis, which overshadowed any genetic causes of the phenotypic variation. The importance of taking into account non-genetic sources of variation during large multiphase experiments was investigated by Oakey *et al* 2013. In this study the saccharification potential of elite barley cultivars was analysed before the data was included in a genome wide association study. It was determined that there was a substantial increase in heritability recorded when field and laboratory variation is included within the experimental model. The results obtained in the study by Oakey *et al* 2013 correspond with those reported by Smith *et al* 2005 who suggested that the largest proportion of non-genetic variation within a study is due to variation occurring in the laboratory.

A prerequisite for successful association analysis is that the parental lines of the mapping population must differ for the trait of interest (Collard *et al* 2005). However, during the saccharification analysis of the samples from INRA-Versailles it was noted that the parental lines did not show any significant difference. This lack of difference was probably also due to the high incidence of non-genetic variation observed within the preliminary study. Collard *et al* 2005 previously reported that a number of factors can influence the detection of QTLs, namely genetic properties, environmental effects, population size and experimental error. They suggested that to overcome these factors an experiment should contain a dense coverage of markers, be replicated across sites and over time, contain a large population and remove errors due to mistakes in genotyping or phenotypic evaluation.

The experiment was then repeated due to the preliminary experiment being unsuccessful in detecting QTLs. However, due to a number of observations being

made during the preliminary experiment it was decided to implement various measures to decrease the amount of environmental variation present within the experiment. This included the use of more replicates (Collard *et al* 2005) and the addition of correction coefficients to take into account the effect of plate position and sample weight (Oakey *et al* 2013).

Plant morphology measurements that included germination frequency, plant height and biomass were conducted to determine if environmental changes due to the position of the replicates within the glasshouse was having an effect. It was determined that there was a similar germination rate across the replicates but there was a difference in height and total stem biomass. When analysing the plant lines it was discovered that one of the replicates had reduced total biomass when compared with the others however this reduction was less pronounced when considering the stem only biomass. It appears that this replicate had less biomass in its seeds. The possible reason for this is that the replicate was possibly affected by a heatwave as it wasn't positioned directly under the cooling system. The environmental condition possibly caused this replicate to undergo senescence early, which could have influenced seed development leading to seeds with lower biomass. In 1993, Elgersam *et al* reported in a study on perennial ryegrass (*Lolium perenne* L.) that temperature had no effect on seed set in their study but they did state this does not give any indication on the effect of seed yield in terms of seed weight and size. The also stated that the temperature was only increased at anthesis whereas it is possible that increased temperatures during grain development may reduce the duration of seed development in terms of the weight of the seed. In 1995, Grass *et al* conducted a study in wheat (*Triticum durum*) that appears to confirm this hypothesis because plants that were exposed to high temperatures 10 days after anthesis showed changes in seed mass due to the early maturation of the ears and senescence of the plants. This change results in a reduced period for grain filling therefore resulting in a reduction in grain weight.

The results obtained during the phenotypic analysis of plant morphology can also be used in future work to identify QTLs related to the traits of germination frequency and plant height as a significant difference between the parental lines were observed. QTLs related to germination have been successfully identified in various species

including rice (Lee *et al* 2015) and maize (Wang *et al* 2016) as shown in recent studies. Lee *et al* identified five QTLs related to germination rate in rice whereas Wang *et al* reported 28 QTLs in a maize RIL population linked to four seed vigour traits, which included germination percentage. QTL analysis of plant height has often been studied in important agricultural crops such as maize (Zhang *et al* 2006) and now more recently in biomass crops such as switchgrass (Serba *et al* 2014). Zhang *et al* identified nine QTLs that explained 78.27% of the phenotypic variance observed in a maize F_2 population whereas Serba *et al* reported five QTLs in a heterozygous pseudo-F_1 population, which explained between 4.3 and 17.4% of the phenotypic variation seen in plant height.

Saccharification analysis of the plants grown at the University of York indicated that the parental lines were in fact significantly different in terms of digestibility, which is different from what was observed during the preliminary experiment. The difference in result is probably due to the removal of the variation observed during the preliminary experiment by improved experimental design. The saccharification data also highlighted that the replicate affected by the heat also contained more variation in the amount of sugar released when compared with the other two replicates. It appears that the early senescence due to the hot weather may have caused a change in cell wall formation by either making it more digestible or resulting in more sugar being present within the wall. The effect of abiotic stress on plant cell walls is not clear because a number of other factors can have an influence, such as the plant species or genotype, the intensity of the stress and the stage of development the plant is at. However, Le Gall *et al* 2015 reported that there are two main responses by the cell wall to heat stress. Firstly, there can be an increase in the amount of expansin proteins and xyloglucan endotransglucosylase or hydrolase (XTH). Secondly, the cell wall can be thicker due to the increase in deposition of lignin and hemicellulose.

During QTL analysis three possible QTLs were detected. The first QTL was found on chromosome 5 and is linked to marker BD1676.1. This QTL was detected during the analysis involving the data from all three replicates as well as when different combinations were analysed. The results indicated that those plant lines containing the AA allele were more digestible than those with BB. The second QTL detected

was found to be in epistasis with the QTL on chromosome 5. The epistatic QTL is situated on chromosome 3 and is linked to marker BD1415.1. However, this QTL interaction was only detected once. Finally, a QTL linked to marker BD2181.2 on chromosome 2 was detected. This QTL on chromosome 2 was also only found during the analysis of a single replicate. Therefore, the strongest QTL is the one on chromosome 5 followed by its' epistatic partner on chromosome 3. The QTL on chromosome 5 can be classed as a major QTL as it accounts for 11.83% of the total variance that is explained by genetic factors (Collard $et\ al$ 2005, Prioul $et\ al$ 2009). The H^2 for this QTL was also calculated as 0.45 when using the value of the mean squares method (Parker $et\ al$ 1998, Broman and Sen 2009). This value is lower than other published H^2 values related to various lignocellulose traits such as neutral detergent fibre (NDF), acid detergent lignin (ADL) and acid detergent fibre (ADF). These values were reported as 0.92, 0.74 and 0.92 respectively within a maize population by Krakowsy $et\ al$ 2005. Shiringani $et\ al$ 2011 reported values of 0.70, 0.67 and 0.84 respectively for a sorghum population, which were lower values than those achieved with the maize population. Shiringani $et\ al$ 2011 also calculated the H^2 values specifically for cellulose (0.82) and hemicellulose (0.24). A different sorghum population was analysed by Murray $et\ al$ 2008 and they determined the H^2 values for cellulose and hemicellulose to be 0.60 and 0.50 respectively. These values are lower than those obtained by Shiringani $et\ al$ 2011 but this is possible due to Murry $et\ al$ 2008 only looking specifically at stem cellulose and hemicellulose instead of the whole plant. These different values obtained in the various studies indicate that there is a wide variability in the heritability of the trait under consideration. However, in all cases there is a moderate to large genetic factor controlling the trait.

This study has successfully identified one main QTL for stem digestibility as well as an epistatic QTL in a Brachypodium RIL population. The next step is to verify these QTLs in a further generation to determine that they are not false positives. Candidate genes linked to the QTL on chromosome 5 and the epistatic QTL on chromosome 3 will also be identified.

Chapter 4: Confirmation of QTLs and the validation of candidate genes using selected *Brachypodium distachyon* RILs.

4.1   Introduction

Once QTLs have been identified it is important to validate the causal relationship for candidate genes found in the QTL region in order to obtain scientific value from association genetic studies. In the past this needed a substantial amount of time and resources as it involved either growing the material in additional locations or over a number of years. In some cases the marker is even analysed in different germplasm, such as near isogenic lines (NILs) or heterogeneous inbred lines (HIFs) (Flint-Garcia *et al* 2003, Borevitz *et al* 2004, Collard *et al* 2005, Thomas *et al* 2010) that require many plant generations to obtain. These approaches involve a further round of funding and time and therefore QTLs are rarely confirmed (Collard *et al* 2005).

In this study it was decided to take a single-marker analysis approach, otherwise known as single-point analysis. This approach uses simple statistical methods, such as t-test and ANOVA to determine if there is a link between a single marker and the trait of interest (Doerge 2002, Collard *et al* 2005). The initial step involves selecting plant lines from the original RIL population that contained differing alleles at the marker identified during QTL analysis. This creates a smaller subpopulation, thereby allowing more replicates of each selected plant line to be included in the analysis. The increase in replication allows for a more robust QTL confirmation.

There is no clear and easy method of identifying the genes underlying the observed QTLs. Traditionally; positional cloning was used to refine the QTL region by fine mapping thereby reducing the list of possible candidate genes. However, this method is time consuming as it involves creating more genetic markers within the QTL region to produce a denser map (Prioul *et al* 1999, Borevitz *et al* 2004, Thomas *et al* 2010). Another route to identifying candidate genes relies on having the genome sequence as well as prior knowledge of the pathway that is under investigation. In this case, the genes within the QTL region are collated and those that relate to the

pathway are selected as candidate genes. The selected candidate genes then have to be validated by determining if there is a correlation between the allelic polymorphism of the gene and the trait under investigation (Prioul *et al* 1999, Borevitz *et al* 2004, Thomas *et al* 2010, Shiringani *et al* 2011). It is this second approach that is followed in this study as the genome for Brachypodium has been fully sequenced (The International Brachypodium Initiative 2010) and the genetic map of this population is relatively dense with markers already (Huo *et al* 2011).

Further candidate gene confirmation can be undertaken by studying the expression levels of the candidate genes. This can be achieved by using a next generation sequencing method involving RNA sequencing (RNAseq). This method has be shown to be successful in identifying candidate genes linked with mouse skeletal muscle QTLs (Lionikas *et al* 2012) as well as identifying a candidate gene responsible for vitamin C concentration in apple fruit (Mellidou *et al* 2012). Recently, Venu *et al* 2014 have used this method to find genes associated with heterosis in rice.

RNAseq is used to measure the proportional representation of all the transcripts within a specific tissue at a particular time point. It can also be used to quantify the change in expression of the genes identified (Wang *et al* 2009, Costa *et al* 2010). The advantage of using this method to study the transcriptome is that it is not limited to only known transcripts. Other advantages include that it can be used to determine the annotation of the genes identified, it has low background noise, it produces a large amount of data within a single run and it is highly reproducible. However, there are a number of challenges still to face with this technology. Firstly, the construction of the library involves fragmenting the RNA. This can be achieved by using a number of methods but these each lead to differences in outcome bias. Secondly, due to the large amount of data produced there is the challenge of storing, retrieving and processing it all efficiently (Morozava *et al* 2009, Wang *et al* 2009, Costa *et al* 2010).

The aim of this chapter was to confirm the previously identified QTLs linked to chromosome 3 and 5 as well as to validate possible candidate genes.

4.2   Material and methods

4.2.1 Plant line selection and preparation of plant material

The plant lines were selected for QTL validation using the Graphical Genotypes (GGT v2.0) software (http://www.plantbreeding.wur.nl/UK/software_ggt.html). The lines were initially selected for those containing the allele AA at marker BD1415_1 on chromosome 3 and either allele AA or BB at marker BD1676_1 on chromosome 5. A further round of selection was undertaken using a pairwise comparison of the GGTv2.0 selected lines to identify those lines that had a genotypic background of more than 65% similarity.

A total of 24 plant lines were selected and sown as six randomised replicates (Block 4 – 9). Each replicate consisted of five trays containing 12 seeds per plant line. The seeds were initially placed at 4ºC for three weeks for vernalisation after which they were moved to the glasshouse. After approximately, three weeks the plants were staked and watering was stopped once senesce began. It took approximately two weeks for the plants to be completely dry so that harvesting could take place.

A further replicate, Block 10, was also sown at the same time but after approximately four weeks in the glasshouse the green stems were harvested for RNA extraction. The nodes as well as first and last internodes were discarded from the main stem before it was flash frozen in liquid nitrogen. The samples were stored at -80ºC until needed.

4.2.2 Saccharification analysis and QTL confirmation

The main stem of the harvested plant material was selected for saccharification analysis. The biomass was prepared and analysed using the semi-automatic robotic platform as described in Chapter 2 (Gomez *et al* 2010, Gomez *et al* 2011).

The QTL confirmation was undertaken using a one-way ANOVA to determine if there was a significant difference in saccharification between the plant lines that

contained the allele AA or BB at marker BD1676_1 on chromosome 5. Each block was analysed individually as well as all the blocks together.

4.2.3 Identification of candidate genes

The genomic region around the QTL was studied using GBrowser (http://mips.helmholtz-muenchen.de/gbrowse/plant/cgi-bin/gbrowse/brachy/) to identify possible candidate genes. The region covered for each QTL was from markers found on either side of the one detected during the analysis thereby including any possible genes that may have co-segregated together with the identified marker. The region from marker BD2599.1 (map position: 4723098) to BD1188.3 (map position: 5459948) was covered for marker BD1415.1 (map position: 5337220) on chromosome 3 (Appendix D). In the case of marker BD1676.1 (map position: 25970456) on chromosome 5, the region covered was from marker BD4088.6 (map position: 25889793) to marker BD3488.1 (map position: 26478751). The genes found within this region are listed in Appendix E.

4.2.4 Polymorphism detection within candidate genes

Parental lines, Bd21 and Bd3-1, were sown so that green stems could be harvested after four weeks. 100 mg of the whole stem, excluding the first and last internode, was collected from each parental line and placed in 1.5 ml tubes before flash freezing in liquid nitrogen to store at -80 ºC until RNA extraction could take place.

RNA extraction was conducted using the Qiagen RNeasy Mini kit (Qiagen, UK) on samples that had been ground in liquid nitrogen using a pestle and mortar. The quality and quantity of the RNA was checked using a 1% agarose gel as well as the NanoDrop spectrophotometer (ThermoScientific, UK) from the Technology Facility (University of York). The samples from each of the parental lines were diluted to 2 ug in 10 ul.

The cDNA was created from the RNA by first incubating at 65 ºC for 5 minutes together with 1 ul 10 mM dNTPs and 1 ul Oligo dT. Once this had cooled the samples had 4 ul 5X buffer, 2 ul dTT, 1 ul RiboLock RNase inhibitor and 1 ul

SuperScript II reverse transcriptase (Thermofisher, UK) added before further incubation at 42 ℃ for 50 minutes. This was followed by an inactivation step of 15 minutes at 70 ℃ before 180 ul dH$_2$O was added and the samples were stored at -20 ℃.

The Brachypodium target gene, Bradi5g25290.1, was amplified from the cDNA using primers designed according to the specification of the cloning kit. The sequence of the primers used was as follows:

- GT43_F: 5' – CAC CAT GAA GCT CCC GCT – 3'
- GT43_R: 5' – CTA GTG ACC ATC TTC AGT ATT TAC TAC G – 3'

The PCR product was extracted from a 1% Agarose gel using the QIAquick Gel Extraction kit (Qiagen, UK). The concentration was checked again using the NanoDrop spectrophotometer.

The extracted PCR product was cloned using the StrataClone Blunt PCR cloning kit together with the StataClone SoloPack competent cells according to manufacturer's protocol (Agilent, UK). The white colonies were selected from the LB-kanamycin plates, which contained 2% X-gal. DNA from the white colonies was extracted using the Wizard Plus SV miniprep DNA purification system (Promega, UK) and M13 forward and reverse primers were used for DNA sequencing (Dundee University).

BioEdit v7.2.5 software was used to determine the presence of any SNPs by comparing the sequences of the cloned parents to each other as well as to the mRNA sequence from NCBI (Accession: XM_010242235).

Sorting Intolerant From Tolerant (SIFT) analysis was conducted (http://sift.jcvi.org/www/SIFT_seq_submit2.html) to determine if any amino acid substitutions had an effect on the function of the protein. This analysis is based on determining the degree of conservation of the amino acid when it is aligned with other closely related sequences (Ng *et al* 2001).

4.2.5 Transcriptomics

RNAseq on an Illumina platform was used to determine the expression levels of genes in the stem cell walls. The saccharification rankings of the selected RILs used during QTL validation were considered when determining which lines to select for RNAseq. It was determined that the two highest and two lowest plant lines with differing alleles at marker BD1676_1 on chromosome 5 would be used, namely L163 and L176 for decreased saccharification (allele BB) as well as L166 and L149 for increased saccharification (allele AA).

The plant lines L149, L163, L166 and L176, which were grown in Block 10 were harvested as four week old plants. 100 mg of main stem, excluding the first and last internode, from each plant was collected in 1.5 ml tubes. Three replicates were collected for each plant line. They were all flash frozen in liquid nitrogen before storing at -80 ℃ until RNA extraction.

The RNA was extracted using the Qiagen RNeasy Mini kit (Qiagen, UK). The RNA concentration was determined using a NanoDrop spectrophotometer (ThermoScientific, UK). The concentration and quality of the RNA was further analysed by the Technology Facility at the University of York using the Agilent 2100 BioAnalyser (Agilent, UK). The following criteria had to be met before the samples could be sent to The Genome Analysis Centre (TGAC, Norwich, UK) for the RNAseq on their Illumina HiSeq 2000/2500:

- RNA concentration of 20 ng/ul
- RNA integrity number (RIN) of 8 or higher.

The raw data was initially analysed by Zhesi He (Department of Biology, University of York) who conducted a pairwise comparison using a 5% discovery rate threshold to identify genes with either an increase or decrease in expression.

4.3 Results and discussion

4.3.1 QTL confirmation

4.3.1.1 Selection of plant lines

The plant lines from the RIL population were selected for QTL confirmation on the basis of their genotype at the markers linked to the QTLs. Plants with the allele AA at marker BD1415_1 that is linked to the QTL on chromosome 3 were selected. These lines were then selected according to the allele on chromosome 5 linked to marker BD1676_1. In this case lines that were either AA or BB were selected. This resulted in 80 lines being selected (41 lines with allele AA and 39 lines with allele BB).

However, 80 lines was still too many for an accurate determination. To narrow the number of plant lines down further another round of selection was conducted whereby a pairwise comparison was done using the 80 selected lines to determine which lines had the greatest similarity in their background genotype (Table 4.1).

Table 4.1: Extract of the table containing the results obtained from the pairwise comparison analysis of the Brachypodium RIL lines to determine those lines with similar background genotypes. Scores given for each pair as a percentage with those higher than 65% similarity are highlighted in red.

		Genotype AA								
		L003	L008	L011	L012	L018	L019	L020	L021	L023
Genotype BB	L002	52	53	53	27	48	38	56	59	49
	L005	41	32	52	53	56	58	52	45	55
	L006	53	31	55	49	57	51	54	49	56
	L010	36	45	43	52	62	54	59	44	67
	L017	52	71	54	43	42	36	36	56	49
	L024	50	50	59	38	46	45	46	55	37
	L026	63	42	51	37	56	46	49	49	64
	L033	58	51	36	72	43	48	45	53	43
	L035	39	52	57	37	44	40	56	57	47
	L037	44	40	51	45	56	52	51	45	58
	L044	41	50	45	45	38	48	45	53	54
	L045	39	43	37	54	44	44	46	44	45
	L049	40	33	46	44	58	52	57	41	52
	L056	50	47	45	63	39	41	27	48	27
	L059	37	44	49	40	51	52	54	40	42
	L063	54	28	49	56	58	54	50	46	55
	L065	43	40	42	50	47	50	40	43	38
	L079	35	61	40	55	35	33	45	54	45
	L084	35	56	44	45	35	44	45	53	40

In total 14 pairs of lines were identified that have a background genotype similarity of 65% or more (Table 4.2). This resulted in 24 individual plant lines being selected and grown as six randomised replicates (Block 4 – 9) for the validation of the QTLs.

Table 4.2: The 14 pairs of plant lines that have a background genotype similarity of 65% or larger according the results obtained from conducting a pairwise analysis.

	AA	BB	Score %
1	L012	L033	72
2	L008	L017	71
3	L122	L085	70
4	L140	L079	70
5	L069	L026	69
6	L020	L176	68
7	L092	L163	68
8	L023	L010	67
9	L156	L079	67
10	L166	L026	67
11	L125	L035	66
12	L149	L098	66
13	L104	L079	65
14	L127	L163	65

4.3.1.2 Saccharification analysis

The selected plant lines were grown as six randomised replicates (Block 4 – 9) and were each screened once on the saccharification platform using a mild alkaline pretreatment. The data was first analysed to determine the distribution of the results for each of the replicates (Figure 4.1). It was determined that the results for each replicate did overlap however there was some variation in the distribution. Block 4 had the narrowest distribution of results (*variance = 6.309*), therefore the least variation in digestibility between the plant lines. Whereas, Block 9 had the most variation (*variance = 79.190*) with the widest distribution of results. Blocks 6, 7 and 8 had a similar distribution of results as their means were approximately 40 nmol reducing groups released/mg material.h^{-1}. In the case of Block 5 the distribution had shifted to the left resulting in the lowest mean (34 nmol reducing groups released/mg material.h^{-1}) but with the second highest distribution (*variance = 66.761*).
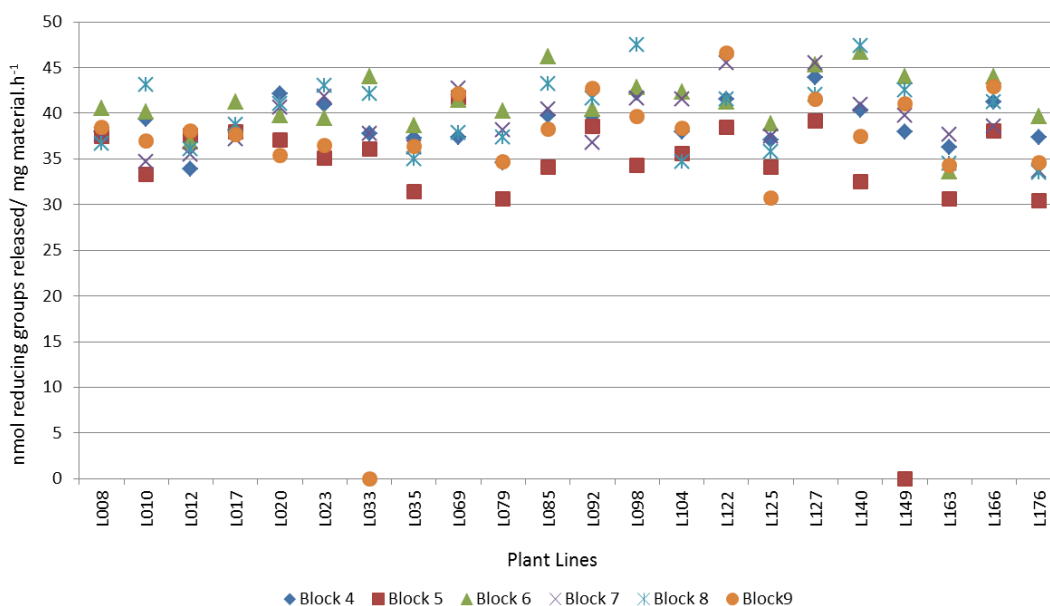
Figure 4.1: The distribution and ranking of the selected Brachypodium RIL plant lines that were grown as six randomised replicates. The ground material from the stem internodes underwent digestion with a commercial cellulase for 8 hours at 50 ℃ following a pretreatment of 0.5 N NaOH for 30 minutes at 90 ℃. The results are the means of six replicates, Blocks 4 - 9, containing 12 plants per line. ,

The plant lines within each replicate were ranked from least to most digestible using standard competition ranking to determine if their rank stayed the same even if the amount of sugar released changed across replicates. The mean rank for each plant line across each replicate was compared to determine if there was a significant change (Table 4.3). This was achieved using the Kruskall-Wallis test, which is the nonparametric version of the one way ANOVA. It was determined that six plant lines (L012, L020, L069, L085, L122 and L127) showed a significant difference ($p < 0.05$) in rank position across the replicates. However, it was observed that of these six plant lines only L020 and L069 showed a deviation in rank position across all the replicates, whereas the remaining lines (L012, L085, L122 and L127) only contained a single outlying replicate. It can therefore be established that for the majority of plant lines their digestibility potential remains constant even if the actual amount of sugar released fluctuates due to possible environmental factors.

Table 4.3: The mean rank of each selected Brachypodium RIL together with the *p*-value indicating if there is a significant difference (highlighted in red) between the ranks obtained for each of the replicates.

Plant Line	Mean Rank	*p*-Value
L008	10	0.375
L010	9	0.06
L012	6	0.04
L017	11	0.56
L020	15	0.01
L023	13	0.05
L033	13	0.06
L035	17	0.45
L069	16	0.01
L079	7	0.05
L085	16	0.04
L092	14	0.65
L098	17	0.07
L104	10	0.06
L122	17	0.04
L125	7	0.07
L127	19	0.04
L140	15	0.05
L149	16	0.15
L163	3	0.07
L166	17	0.58
L176	4	0.27

4.3.1.3   Confirmation

A one-way ANOVA statistical test was used to determine if the QTLs were true QTLs and not false positives. This was achieved by determining if there was a significant difference between those plant lines that contained allele AA or allele BB at marker BD1676_1 on chromosome 5. It was hypothesised that if the QTL was a false positive the mean of those plants with genotype AA would be the same as the

mean of the plants with genotype BB. However, if the F-value was greater than 1 it would indicate that this was not true and there was in fact a difference between the alleles and the QTLs were true.

Initially, Block 4 – 9 were analysed individually (Table 4.4) and it was determined that Block 6 and 8 showed no difference between genotypes. However, Block 5 and 9 showed a significant difference between alleles. In the case of Block 9 the digestibility of the AA plant lines was more ($F = 3.7, Pr = 0.02$) than the BB plant lines. There was an even greater difference ($F = 9.2, Pr = 0.007$) between AA and BB plant lines in Block 5. Finally, all the blocks were analysed together to determine if the difference in digestibility was still true. It was shown to be significantly different with a F-value of 6.0 ($Pr = 0.023$). Therefore, it can be deduced that the QTLs detected in the initial experiment are true QTLs and not false positive results.

Table 4.4: The results obtained during the analysis of QTL confirmation using a one-way ANOVA statistical test. Red highlighted values indicate a significant difference between plant lines containing the AA or BB allele of marker BD1676_1.

Replicate	Mean		One-way ANOVA	
	AA	BB	F-value	Pr(>F)
Block 4	38.12	36.56	2.38	0.139
Block 5	35.35	31.72	9.169	0.0069
Block 6	40.21	39.61	0.229	0.638
Block 7	39.06	36.62	3.699	0.06
Block 8	38.93	38.75	0.009	0.97
Block 9	37.94	35.29	3.74	0.02
All	38.26	36.13	6.04	0.023

4.3.2 Candidate genes

The regions spanning the markers identified during QTL analysis were studied to identify candidate genes, which might have roles in affecting cell wall digestibility. The markers BD2599_1 and BD1188_3 on chromosome 3 were identified as the closest markers found on either side of BD1515_1, which is the marker linked to the

QTL on this chromosome. This region of Brachypodium genome was investigated *in silico* and a total of 108 genes were identified within this region (Appendix D).

The same process was used to collate a list of possible candidate genes for the QTL on chromosome 5. In this case, the region between markers BD4088_6 and BD3488_1 found on either side of the QTL linked marker BD1676_1 was investigated. This exploration resulted in a total of 104 possible genes being identified (Appendix E).

4.3.2.1   Chromosome 3

A short list of three possible candidate genes from the *in silico* investigation of the selected genomic region on chromosome 3 were highlighted as they are known to play a role in cell wall biosynthesis.
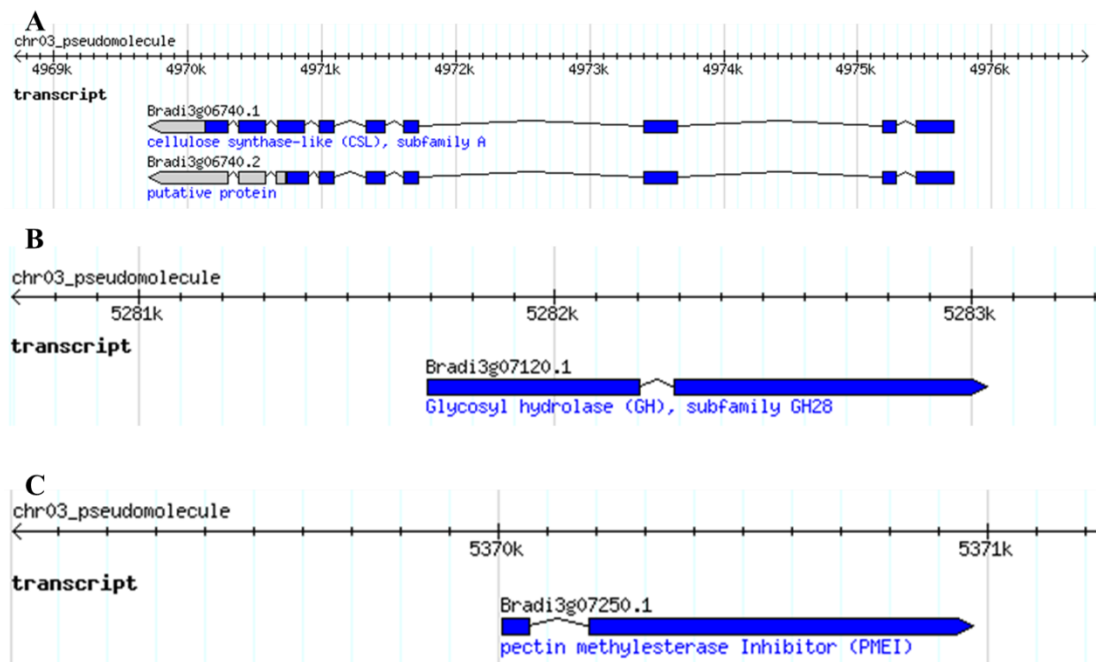
Figure 4.2: The structure of possible candidate gene selected according to *in silico* analysis of the Brachypodium genomic region found between markers BD2599_1 and BD1158_3 on chromosome 3 (A) Bradi3g06740 (B) Bradi3g07120.1 and (C) Bradi3g07250.1. Figure from Gbrowser (http://mips.helmholtz-muenchen.de/gbrowse/plant/cgi-bin/gbrowse/brachy/).

The first gene identified was Bradi3g06740.1 (map position 4981314) and it has a genomic size of 6008bp containing nine exons that are spliced to form a transcript of 1982bp. There also appears to be two possible open reading frames at this locus, however Bradi3g06740.2 contains an insertion of approximately 30bp, which results in an earlier termination (Figure 4.2A). This gene has been classified as belonging to the glycosyl transferase family group 2 proteins and having a glucomannan 4-beta-mannosyltransferase function according to the Pfam database. Cellulose synthase-like (Csl) subfamily A proteins have been shown to have a role in hemicellulose biosynthesis, specifically in terms of mannan and glucomannan synthesis within the cell wall (Scheller *et al* 2010).

Bradi3g07120.1 (map position 5294447) was the next gene identified as it is a glycosyl hydrolase (GH28) with a predicted polygalacturonase function according to the Pfam database. The GH28 gene has a genomic size of 1344bp containing two exons that form a transcript of 1257bp (Figure 4.2B). Glycosyl hydrolases (GH) are known to play a role in cell wall biosynthesis as their function is to cleave bonds found between carbohydrates. GH28 specifically acts on pectin, which is found as a major component of the middle lamella and plays a role in keeping adjacent cells connected. When GH28 disrupts pectin it results in a reduction in cell to cell adhesion thereby allowing for cell wall remodelling during plant development (Tyler *et al* 2010).

Finally, Bradi3g07250.1 (map position 5383310) was identified as a possible candidate gene. This gene contains two exons that form a transcript of 846bp from a genomic DNA sequence of 967bp (Figure 4.2C). Bradi3g07250.1 has been classified according to the Pfam database as a pectin methylesterase inhibitor (PMEI), which is a regulator of pectin methylesterases. These methylesterases are involved in cell wall biosynthesis thereby resulting in pectins with different structures and functions (Giovane *et al* 2004).

Gene expression data for these three possible candidate genes was collated from the Phytozome database to determine if the genes are expressed in the relevant tissue and stages of Brachypodium (Table 4.5).

Table 4.5: The expression levels of possible chromosome 3 candidate genes according to data collected from selected *Brachypodium distachyon* v3.1 expression libraries within the Phytozome database (https://phytozome.jgi.doe.gov).

	3g06740.1 (CslA)		3g07120.1 (GH28)		3g07250.1 (PMEI)	
	FPKM	Locus DE	FPKM	Locus DE	FPKM	Locus DE
Flag leaf 47d 18lgt 6dk	0.295	LOW	0.015		0.268	
Flower 47d 18lgt 6dk	11.082		1.608	HIGH	0.056	LOW
Leaf mature 47d 18lgt 6dk	0.014	LOW	0.054		0.047	LOW
Leaf young 23d 18lgt 6dk	0.005	LOW	0.021		0.035	LOW
Shoot 24d 18lgt 6dk	6.797		0.05		0.105	
Stem base 47d 18lgt 6dk	22.219		Not Expressed		0.224	
Stem tip 47d 18lgt 6dk	13.251		0.064		0.496	HIGH
Stem 47d 18lgt 6dk	17.237		0.041		0.54	HIGH

FPKM = Fragments per kilobase of transcript per million mapped reads.

Locus DE = The expression level for the gene in this library is more than 1 std dev above/below the average across all libraries.

From the expression data it appears that CslA is the most highly expressed in the stems making it a good candidate to have a role in stem digestibility, with PMEI as also a possible candidate as it shows expression in the stem tip and total stem tissue libraries. GH28 has the highest expression in flowers.

Finally, SNP *in silico* data was collated for the possible candidate genes to determine if any variations within the coding sequence have been identified when comparing the parental sequences Bd21 and Bd3.1 (Figure 4.3). A difference in the sequence could possibly explain the allelic variation observed during saccharification analysis. Both GH28 and PMEI contain a number of sequence differences in the two plant lines, suggesting potential for variations in protein activity, whereas no SNPs are apparent in the CslA gene. In this case GH28 and PMEI could be taken forward as candidate genes based on this information but together with the expression data it appears that CslA and PMEI are the most probable candidate genes.
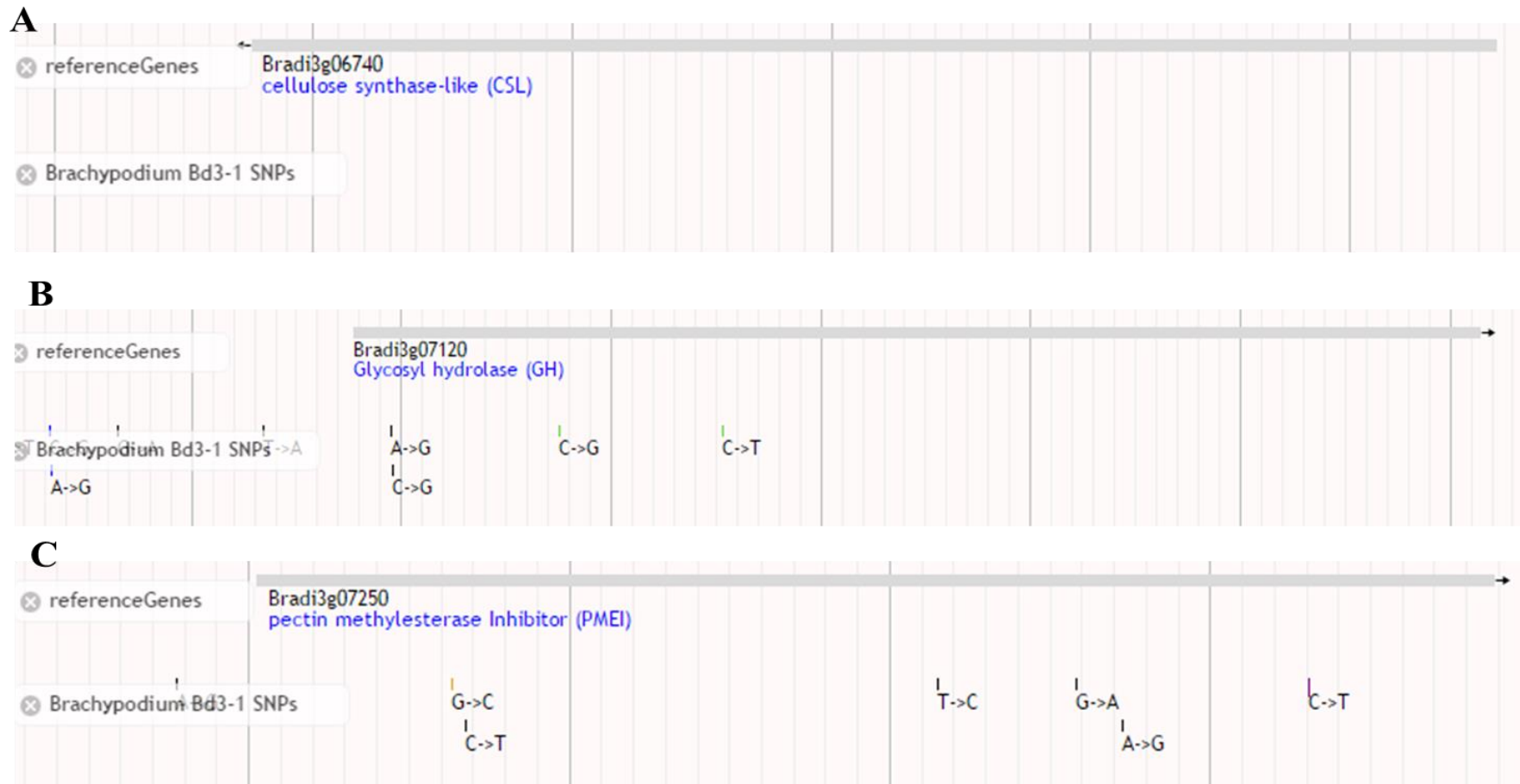
Figure 4.3: *In silico* analysis of SNPs identified from the comparison of Brachypodium Bd21 and Bd3.1 genomic sequences of the possible chromosome 3 candidate genes (A) Bradi3g06740 (B) Bradi3g07120.1 and (C) Bradi3g07250.1. Figure from Gbrowser (http://jbrowser.brachypodium.org).

From the expression data and the SNP analysis results it could be hypothesised that that Bradi3g07250 (PMEI) is a possible target for further analysis as it shows a high expression within stem tissue as well as containing a known genomic sequence variation.

4.3.2.2   Chromosome 5

Six possible candidate genes on chromosome 5 were selected for further *in silico* investigation as they have been reported to play a role in cell wall biosynthesis. The first three genes Bradi5g24180.2 (map position 26084177), Bradi5g24190.1 (map position 26089868) and Bradi5g24310.1 (map position 26180320) belong to the WAK receptor-like protein kinase subfamily b group of proteins (Figure 4.4). These proteins are found to span the plasma membrane and have a direct link with the cell wall as they are covalently linked to pectin. It has been shown that a reduction in WAKb expression relates to a loss in cell expansion (Kohorn 2001). All three genes contain three exons each and range in size from 1521bp to 2306bp when transcribed.
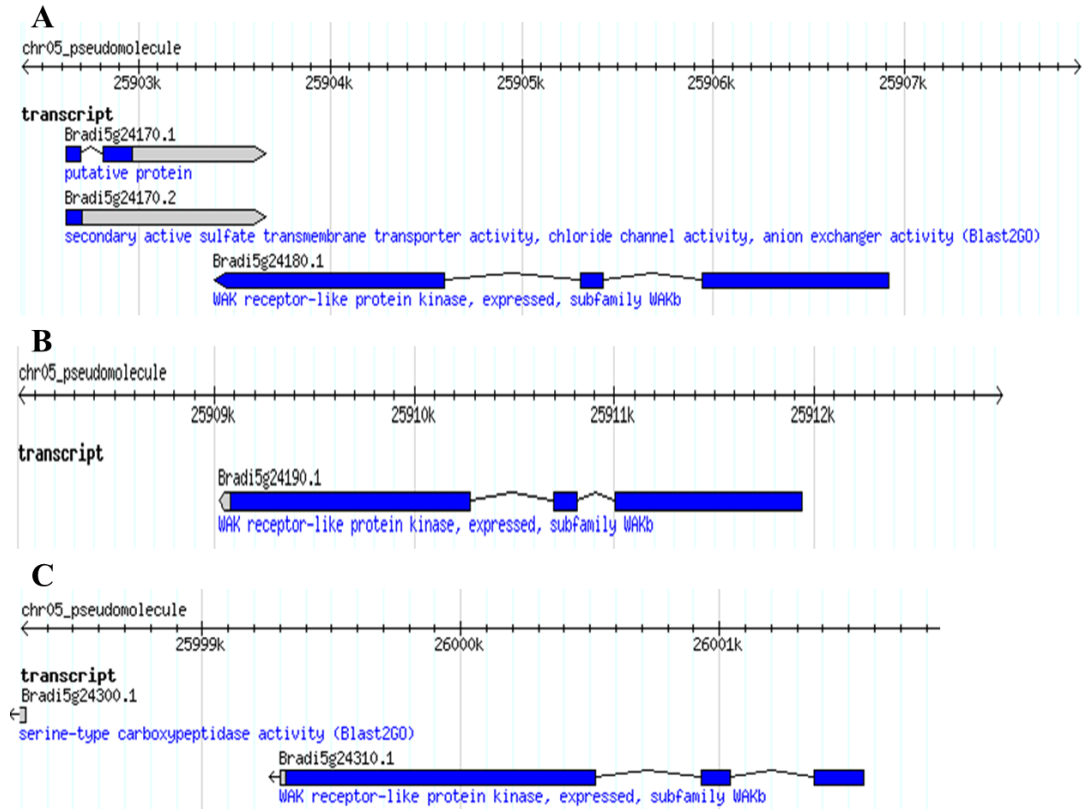
Figure 4.4: The structure of three WAKb genes (A) Bradi5g24180.2 (B) Bradi5g24190.1 and (C) Bradi5g24310.1 that have been identified as possible candidate gene according to the *in silico* analysis of the Brachypodium genomic region found between markers BD4088_6 and BD3488_1 on chromosome 5 (Figure from Gbrowser (http://mips.helmholtz-muenchen.de/gbrowse/plant/cgi-bin/gbrowse/brachy/).

The next gene, Bradi5g24280.1 (map position 26163724), encodes a putative UDP-galactosyltransferase activity (Figure 4.5A). It is thought that the galactosyltransferase enzymes are probably membrane bound within the Golgi apparatus and are responsible for the incorporation of galactose into the non-cellulosic polysaccharides and glycoproteins found within the cell wall (Norambuena *et al* 2002). The Brachypodium galactosyltransferase gene has a genomic size of 2207bp containing three exons which form a 1605bp transcript.
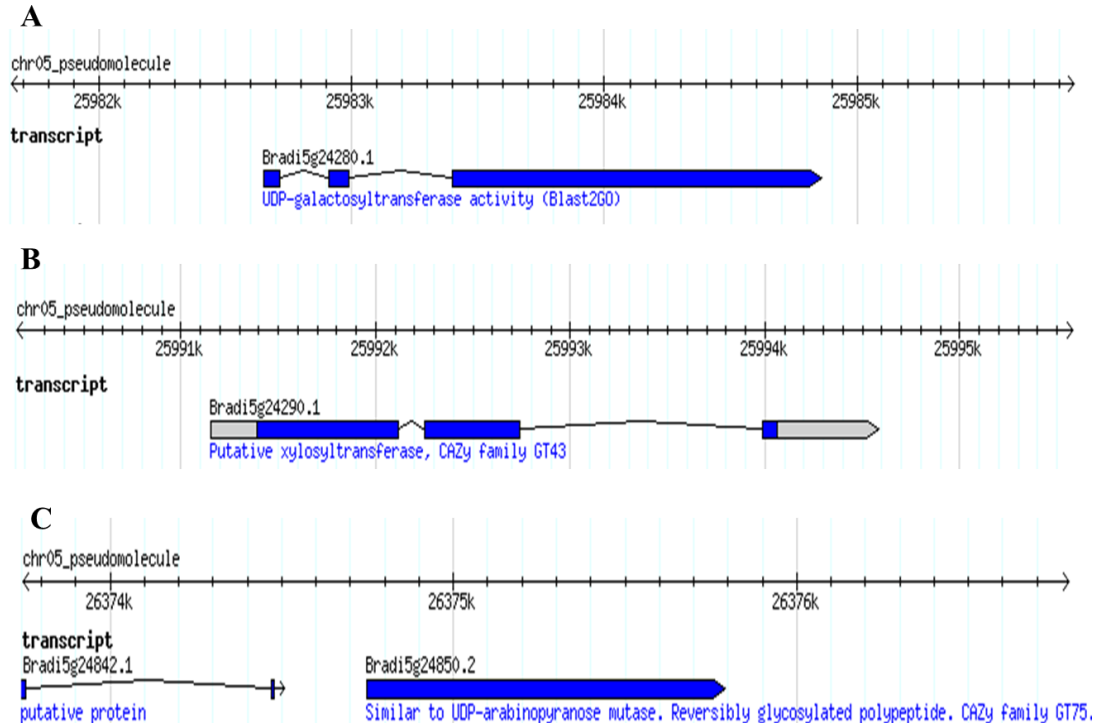
Figure 4.5: The structure of a further three possible candidate genes (A) Bradi5g24280.1 (B) Bradi5g24290.1 and (C) Bradi5g24850.2 found on chromosome 5 between markers BD4088_6 and BD3488_1, which have been selected during an *in silico* analysis of this region. (Figure from Gbrowser (http://mips.helmholtz-muenchen.de/gbrowse/plant/cgi-bin/gbrowse/brachy/).

Another possible candidate gene is Bradi5g24290.1 (map position 26172185). This gene contains three exons and has a transcript size of 2045bp (Figure 4.5B). It has been classified as a xylosyltransferase from the GT43 family. It has been reported that GT43 together with other glycosyltransferases (GT8 and GT47) are involved in the biosynthesis of the xylan backbone found in hemicellulose (Carpita *et al* 2012, Dhugga 2012).

The final possible candidate gene found on chromosome 5 is Bradi5g24850.1 (map position 26374746) and consists of a single exon of 1047bp (Figure 4.5C). According to Pfam this gene encodes a putative UDP-arabinopyranose mutase and belongs to the GT75 family of enzymes. It has been reported that arabinopyranose mutases are involved in the biosynthesis of cell wall polysaccharides by interconverting UDP-arabinofuranosyl and UDP-arabinopyranose as the furanose form of arabinose is more common than the pyranose form found in

114

rhamnogalacturonan I, glucuronoarabinoxylans and other glycoproteins (Sumiyoshi *et al* 2015).

The gene expression data was collated from the Phytozome database in regards to the six chromosome 5 candidate genes to determine if they are substantially expressed in the relevant Brachypodium plant tissues (Table 4.6). From this data it appears that all three WAKb genes show low levels of expression in stem tissues whereas GT43 and GT75 both show substantial expression in stem tissues. This data therefore points to the possibility of both GT43 and GT75 being good candidate genes.

Table 4.6: The expression levels of possible chromosome 5 candidate genes according to data collected from selected *Brachypodium distachyon* v3.1 expression libraries within the Phytozome database (https://phytozome.jgi.doe.gov).

	5g24180.2 (WAKb)		5g24190.1 (WAKb)		5g24310.1 (WAKb)		5g24280.1 (Galactosyltransferase)		5g24290.1 (GT43)		5g24850.1 (GT75)	
	FPKM	Locus DE	FPKM	Locus DE	FPKM	Locus DE	FPKM	Locus DE	FPKM	Locus DE	FPKM	Locus DE
Flag leaf 47d 18lgt 6dk	0.181		0.16	LOW	0.713		0.19		2.859	LOW	47.744	LOW
Flower 47d 18lgt 6dk	0.325	HIGH	1.125		0.655		1.116		31.991	HIGH	151.81	HIGH
Leaf mature 47d 18lgt 6dk	0.501	HIGH	0.606		0.983		0.085		3.451		53.802	LOW
Leaf young 23d 18lgt 6dk	0.192		0.701		0.821		0.181		1.559	LOW	39.207	LOW
Shoot 24d 18lgt 6dk	0.166		0.954		0.836		0.044		4.058		78.597	
Stem base 47d 18lgt 6dk	0.184		1.874		0.776		0.052		9.252		165.28	HIGH
Stem tip 47d 18lgt 6dk	0.158		1.446		0.785		0.042		14.591	HIGH	136.98	HIGH
Stem 47d 18lgt 6dk	0.096	LOW	1.074		0.77		0.085		25.954	HIGH	162.18	HIGH

FPKM = Fragments per kilobase of transcript per million mapped reads.
Locus DE = For the gene, the expression level in this library is more than 1 std dev above/below
the average across all libraries.

The SNP *in silico* data for the six possible candidate genes on chromosome 5 were assessed to determine if any variations within the genomic sequence occurs when comparing the parental sequences Bd21 and Bd3.1 (Figure 4.6). This data shows that no SNPs have been observed in Bradi5g24310.1 and Bradi5g24850.1 however a lot of variation was seen in genes Bradi5g24190.1, Bradi5g280.1 and especially Bradi5g24180.2. Gene Bradi5g24290.1 seems to only contain a single polymorphism.

From the expression data and the SNP analysis results it could be hypothesised that that Bradi5g24290.1 (GT43) is a possible target for further analysis as it shows an increase in expression within stem tissue as well as containing a known genomic sequence variation, which possibly could lead to an allelic variation in expression resulting in differences in saccharification potentials of the stem biomass.

Figure 4.6: *In silico* analysis of SNPs identified from the comparison of Brachypodium Bd21 and Bd3.1 genomic sequences of the possible chromosome 5 candidate genes (A) Bradi5g24180.2 (B) Bradi5g24190.1 (C) Bradi5g24310.1 (D) Bradi5g24280.1 (E) Bradi5g24290.1 and (F) Bradi5g24850.1. Figure from Gbrowser (http://jbrowser.brachypodium.org).

4.3.3 Polymorphism detection within the candidate genes

The candidate gene, Bradi5g24290.1 on chromosome 5, was cloned from each of the parental lines and sequenced to determine if there was a SNP present to indicate any allelic difference between the two lines to explain the difference in digestibility. The parental lines were aligned to the wild type, Bd21, sequence that was obtained from the NCBI database (Figure 4.7). From this alignment it was determined that there were two SNPs present. The first SNP was identified at a position of approximately 111bp from the start of the coding region and involved a change from a cytosine (C) to an adenine (A) nucleotide whereas, the second SNP at approximately 238bp from the start of the coding region resulted in an alteration from a guanine (G) to an A nucleotide.
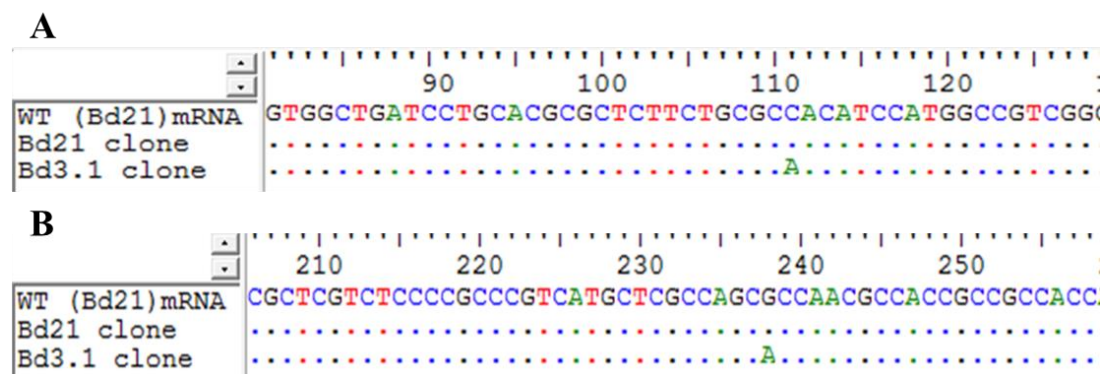
Figure 4.7: Sequence alignment of Bradi5g24290.1 cloned from the parent lines Bd21 and Bd3.1 compared with WT mRNA sequence (Accession number: XM_010242235) from the NCBI database indicating a (A) silent SNP and a (B) missense SNP.

The predicted codon sequence was evaluated to determine if the SNPs resulted in any change to the amino acids (Figure 4.8). The first SNP found at amino acid position 37 resulted in a nucleotide substitution of GCC to GCA, which both encode for alanine. Therefore, this SNP is a silent allelic variation and would not have an effect on protein function.

ATG AAG CTC CCG CTG TCG GCG GGG TCG CCC GAC GCG GCG GCG CTG CCG GAG TCG GCC AGC AAG

M   K   L   P   L   S   A   G   S   P   D   A   A   A   L   P   E   S   A   S   K

CCG TCG CTG CCG GCC ACG TGG CTG ATC CTG CAC GCG CTC TTC TGC GCA/C ACA TCC ATG GCC GTC

P   S   L   P   A   T   W   L   I   L   H   A   L   F   C   A/A   T   S   M   A   V

GGG TTC CGC TTC TCG CGC CTC GTC GTC TTC CTC CTC TTC CTC CCG ACC CCG CCC ATG GAC CCC GCC

G   F   R   F   S   R   L   V   V   F   L   L   F   L   P   T   P   P   M   D   P   A

GCG CAC CTC GTG TCG CTC GTC TCC CCG CCC GTC ATG CTC GCC AGC A/G CC AAC GCC ACC GCC GCC

A   H   L   V   S   L   V   S   P   P   V   M   L   A   S   T/A   N   A   T   A   A

ACC ATC ACC ACC ACC ACG ACC ACC ACC ACC ACC GTC ACC ACC ACC ACG GTC GCC GAG GCG

T   I   T   T   T   T   T   T   T   T   T   V   T   T   T   T   V   A   E   A

Figure 4.8: Part of the mRNA sequence (black) and protein (blue) sequence (Accession number: XP_010240537.1) of Bradi5g24290.1 indicating the change in nucleotide and amino acid (red) for both SNPs.

The second SNP results in a GCC to ACC nucleotide substitution at amino acid position 80 causing a change from alanine to threonine at this position. This represents a missense variation because it results in a change in amino acid but doesn't result in a premature stop in translation which would produce a truncated protein. This change in amino acid could possibly have an influence on the structure and function of the protein as alanine and threonine are very different amino acids (Figure 4.9). Alanine is a small (molecular weight: 71.09 Da) and simple amino acid whereas threonine is larger (molecular weight: 101.11 Da) and contains a hydroxyl group. Alanine is also classified as nonpolar and has a neutral pH because its methyl group is non-reactive and it can't be phosphorylated. On the other hand, threonine also has a neutral pH but it is classified as a polar amino acid. Threonine is also prone to various posttranscriptional modifications such as phosphorylation.
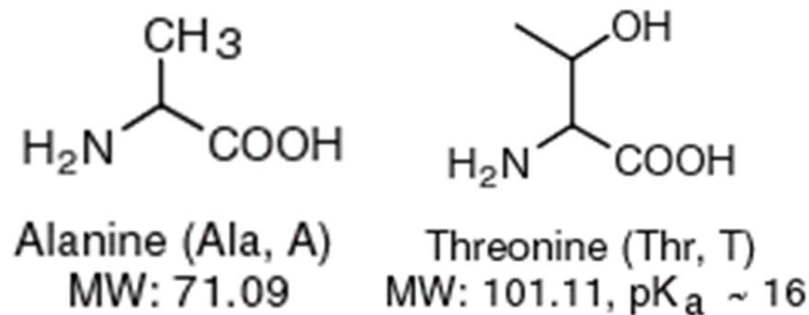
Figure 4.9: Structure and molecular weight of the amino acids Alanine and Threonine.

SIFT analysis was conducted to determine if there was a possibility that the change in amino acid at the second SNP could result in a change in protein structure or function. SIFT is one of a number of similar tools that are used for the characterisation of missense SNPs. SIFT is based on the analysis of both experimental mutagenesis data as well as human polymorphism and disease data. This tool has been predominantly used in human genetic research but has been used successfully in determining the possible effect of amino acid variations in agricultural crops such as maize as well as in model organisms including Arabidopsis (Till *et al* 2004; Sim *et al* 2012). SIFT analysis involves investigating a number of databases to determine if the region within the protein is conserved. It produces a score for each of the databases and the lower the score the more likely there will be an effect on the protein function as it is more likely that the region being analysed is conserved amongst various species. The results from the SIFT analysis are shown in Table 4.7 and it appears that the region is possibly conserved across species as the SIFT score for each of the databases searched was below 1.0. The only amino acid substitutions that appear to be tolerated at amino acid position 80 are arginine, glycine and serine as they are the only ones that have been reported within the databases.

Table 4.7: The results from SIFT analysis indicating the changed amino acid highlighted in red.

Database	Predicted amino acid not tolerated	Number of sequences compared	SIFT Score	Predicted amino acid tolerated
SwissPort	ywvtsrqpnmlkihgfedc	15	0.06	a
TrEMBL	whyfimqrndelckvtp	54	0.16	gsa
uniRef90	whyfmiqrndelckvtpg	61	0.1	sa
NCBI	whyfimqrndelkcvtp	72	0.15	gsa

The same method was also followed for the candidate gene Bradi3g06740 but unfortunately it was not cloned successfully. This was possibly due to there being two possible open reading frames for the gene within the same region (refer to Chapter 3 Figure 3.15), which results in two very similar PCR products being amplified.

4.3.4 Transcriptomics

The transcriptome of the Brachypodium lines that had been identified as having either a high (L149 and L166) or low (L176 and L163) digestibility potential during saccharification analysis were investigated. The transcriptomic data was used to determine which genes had a change in expression as well as to determine if transcripts of the candidate genes, Bradi5g24290.1 and Bradi3g06740, were present in developing stems as is needed if they are to have an effect on straw digestibility.

The data from RNAseq was analysed by using a pairwise comparison between the lines involving a 5% threshold discovery rate to identify genes that had either an increase or decrease in expression (Table 4.8). In each comparison Bradi5g24290.1 was identified as having a significant differential expression, however Bradi3g06740 was not identified in the dataset significantly weakening any plausible causative role in determining the differences in digestibility in our analyses.

Table 4.8: Summary of results obtained from RNAseq of the Brachypodium transcriptome of selected lines with either high (↑) or low (↓) digestibility.

		Genotype AA	
		L149 ↑	**L166 ↑↑**
Genotype BB	**L176 ↓**	Total = 269 genes GT43 = # 48 (p = 1.63x10^{-11})	Total = 1846 genes GT43 = # 128 (p = 1.10x10^{-11})
	L163 ↓↓	Total = 653 genes GT43 = # 53 (p = 4.39x10^{-14})	Total = 1398 genes GT43 = # 233 (p = 5.27x10^{-09})

The genes that showed a change in expression during the pairwise comparison were then compared to determine which ones showed a variation within all four comparisons as well as how many were novel within each comparison (Figure 4.10). A total of 2720 genes showed a change in expression levels, of these only 65 genes showed a significant variation within all four comparisons (Appendix F). The candidate gene Bradi5g24290.1 was found within this list at position 19 (p-value = 1.10×10^{-11}) when the genes are ranked from most to least significant using the data produced during the comparison of L166 and L163 (the most and least digestible lines). No other known genes involved in cell wall biosynthesis were isolated besides Bradi5g24290.1, however 27 genes with unknown function were detected that had a change in expression levels.
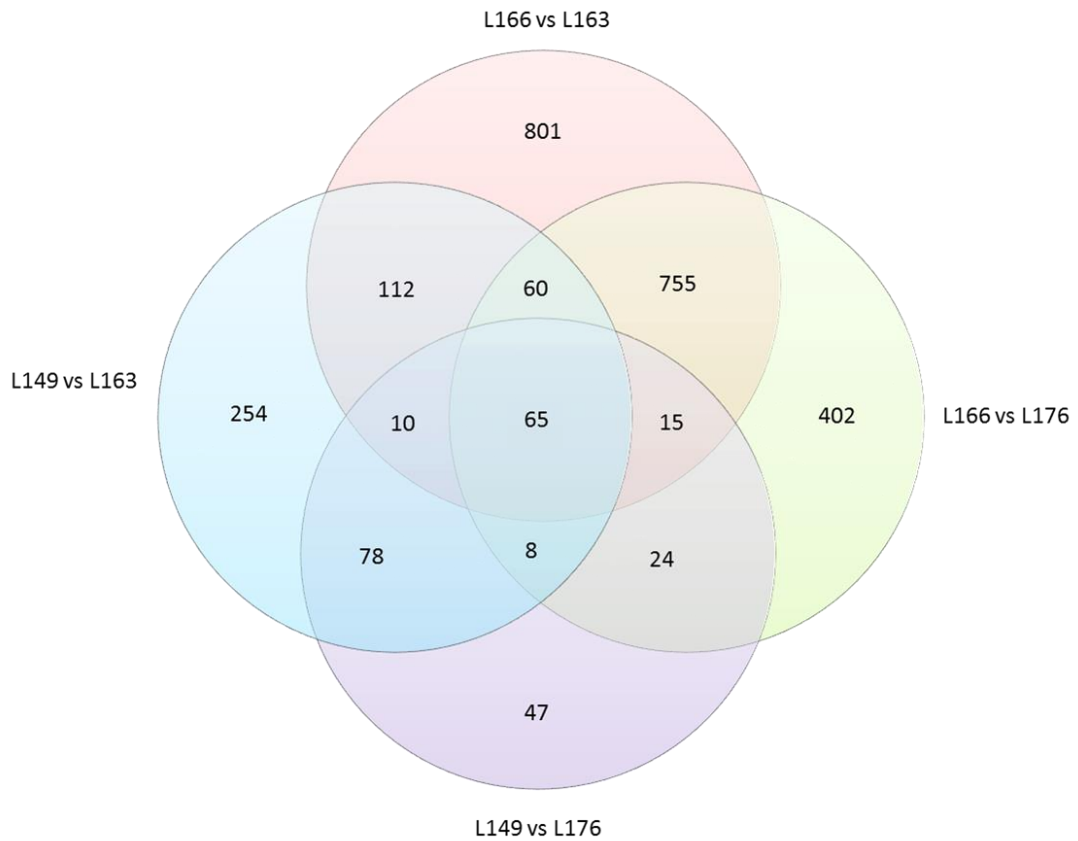
Figure 4.10: The number of genes identified during the pairwise comparison of plant lines with high and low digestibility potential that showed a significant change in expression levels.

4.4   Conclusion

In this study the aim was to confirm the QTLs identified previously linked to chromosome 3 and 5 of Brachypodium. This is an important step in QTL analysis in terms of identifying markers for marker assisted breeding programs (Borevitz *et al* 2004). It is also important in QTL research where the aim is understanding gene function. The reason for confirming the QTLs identified, is to ensure that they are not false positives, which can occur if any statistical errors or environmental anomalies have transpired during the initial analysis.

Validation of the QTLs was successfully achieved using a single-marker analysis approach. This method of analysis allows for an increased confidence in the results in a timely fashion without the necessity of increased funding which is traditionally needed when confirming QTLs using new populations, such as NILs, for analysis

(Fint-Garcia *et al* 2003, Borevitz *et al* 2004, Collard *et al* 2005, Thomas *et al* 2010). The QTLs were confirmed by using one-way ANOVA to determine if there was a significant difference in the amount of sugar released during saccharification from the selected plant lines grouped according to the alleles of marker BD1676_1. It was determined that there was a significant difference in digestibility between the alleles. This therefore supported the theory that the QTLs were in fact true QTLs and not caused by statistical errors or influenced by the environmental.

Another aim was to identify possible candidate genes linked to the identified QTLs. An *in silico* approach was used to achieve this. It was found that for this method to be successful a number of factors need to be taken into consideration. Firstly, a good genetic map was needed that was densely populated with markers so that only a small region of the genome was investigated. Secondly, comprehensive bioinformatic resources were necessary including a whole genome sequence, an expression database that included various plant tissue libraries and a SNP database that covered the parental plant lines used in the creation of the population studied. Finally, knowledge of the pathways being studied was also necessary. This method was followed instead of using positional cloning to refine the QTL area (Prioul *et al* 1999, Borevitz *et al* 2004, Thomas *et al* 2010) as this method would have involved setting up new markers which would have been too time-consuming.

Possible candidate genes for the identified QTLs were selected by exploring the Brachypodium genes found within the region around the linked marker. It was determined that the probable candidate gene for the QTL on chromosome 5 is likely to be Bradi5g24290.1, which is a possible beta-1,3-xylosyltransferase belonging to the GT43 gene family. The epistatic QTL on chromosome 3 is possibly a CslA gene, Bradi3g06740. This gene is described as having a probable glucomannan 4-beta-monnosyltransferase function. During the *in silico* analysis of chromosome 3 another possible candidate gene was identified, Bradi3g07250. This gene has been identified as a possible pectin methylesterase inhibitor (PMEI). The PME and PMEI gene family is very large and is known to have a complex expression pattern including possibly a number of redundant genes. The role of PME in pectin biosynthesis is to catalyse the demethylesterification of homogalacturonans that form the pectin polymer however PMEIs spatially regulate this demethylesterification resulting in

rigid or loose cell walls depending on the mode of demethylesterification (Pinzón-Latorre *et al* 2013). Pectin is found as a major component of primary cell walls in eudicots and noncommelinid monocots (35%) but not in grasses (2-10%). It has been reported that the PME and PMEI gene families have different expression patterns in Arabidopsis and rice which could possibly account for the different levels of pectin found within the cell walls of these plants (Wang *et al* 2013).

From the *in silico* analysis of the possible candidate genes for chromosome 3 it would appear that Bradi3g07250 (PMEI) would be the gene selected for further analysis as it has possible SNPs that could lead to allelic variation in expression as well as being expressed within the correct tissue. However, at the time of selecting a candidate gene the expression data wasn't available so it was decided to follow Bradi3g06740 (CslA) for further analysis as we knew that grasses contain only a small amount of pectin within their cell walls therefore it was more likely that a gene involved in hemicellulose biosynthesis would play a role in digestibility. This therefore highlights the need for good bioinformatic resources especially for expression data related to the plant species being studied when selecting candidate genes. Future work could now be conducted looking further at the role of PMEI as a possible candidate gene involved in cell wall digestibility even when found making up only a small component of the cell wall. De Souza *et al* 2015, have reported that pectin can have an influencing factor in recalcitrance of Miscanthus when lignin is not the main limiting component. It appears that the rhamnogalacturonan I (RG-I) and arabinogalactan (AG) polymers of pectin that are tightly bound in the cell wall negatively affect digestibility. They also reported that even when lignin in the main contributing factor it appears that some RG-I polymers, probably those bound to lignin, also have a negative influence on digestibility.

A further aim was to validate the candidate genes that had been selected during the *in silico* investigation. One method of validating candidate genes is to identify allelic variations within the genes of interest that could explain the difference in expression, which leads to the variation in trait (Prioul *et al* 1999). SNPs can be found in genomic DNA in either a region coding for a gene or in non-coding regions found between genes. SNPs found in coding regions won't necessarily affect the gene function as it depends on the type of nucleotide substitution that has occurred. Silent

SNPs happen when the substitution doesn't change the sequence of the polypeptide due to codon redundancy whereas missense and nonsense SNPs do result in a change in polypeptide sequence due to either the inclusion of a different amino acid or a premature stop codon. To identify any allelic differences in the candidate gene Bradi5g24290.1, it was cloned from each of the parental lines Bd3.1 and Bd21 and the cDNA sequences were compared to identify any possible SNPs. Two SNPs were identified within the coding region with the first of which being identified as a silent variation as the DNA substitution resulted in the amino acid remaining as an alanine. The second SNP resulted in a missense variation as the DNA substitution resulted in a change in the amino acid. In this case the non-coding region around the gene of interest was not analysed though it must be noted that even SNPs in this region could still possibly have an effect on gene splicing as well as on the binding of transcription factors.

The same method for detecting allelic variation was followed for the Bradi3g06740 candidate gene but it was not cloned successfully possibly due to the presence of two open reading frames within the region. In the future, other cloning techniques may be applied to achieve a successful clone for sequencing and SNP determination.

Another way of validating candidate genes is to analyse the expression profile of the gene of interest to determine if there is a difference in levels between the alleles (Lionikas *et al* 2012). There are a number of methods that can be used to determine candidate gene expression, which have been developed overtime as technology has progressed. Northern blots were the first method used and it involved radioactivity as well as requiring a large amount of RNA. The next method developed stemmed from the advancement of PCR techniques. RT-qPCR increased throughput as well as decreasing the amount of RNA required. However this method still only allows for a few genes to be analysed at a time. The next improvement in gene expression analysis was the development of microarray assays. This technique has the advantage of being cost-effective as well as reliable and rapid. However, it had a number of disadvantages in that there is background noise, cross-hybridization occurs and it also required previous knowledge about the genes of interest. The next development involved a number of different approaches based on sequencing, such as EST sequencing, which increased throughput further but was still labour

intensive. The most recent advancement has come in terms of next generation sequencing and specifically the ability to conduct parallel sequencing of cDNA using technologies such as Illumina (Morozova *et al* 2009, Wang *et al* 2009, Costa *et al* 2010, Lionikas *et al* 2012).

In this study it was decided to use RNAseq as not only would it provide information about the candidate genes but also any novels genes that have an influence in cell wall biosynthesis (Morozava *et al* 2009, Wang *et al* 2009, Costa *et al* 2012). RNA from plant lines that had either a high (L149 and L166) or low (L176 and L163) digestibility potential was collected and the gene expression data was analysed resulting in 2720 genes being identified as having significant changes in expression depending on digestibility potential. The number of genes identified was further reduced to 65 genes by identify those that had a significant had a significant change in expression across all four comparison conditions. The candidate gene Bradi5g24290.1 was found to be one of these conserved genes. This result therefore validates Bradi5g24290.1 as a candidate gene for the QTL on chromosome 5. Further work can now be undertaken to gain a better understanding of the function of this gene and how it relates to cell wall digestibility.

Chapter 5: Cell wall composition analysis of selected *Brachypodium distachyon* RILs and candidate gene *Arabidopsis thaliana* T-DNA lines.

5.1 Introduction

The secondary cell walls of plants consist predominantly of three main constituents; cellulose, hemicellulose and lignin. These together create a recalcitrant complex structure. Much research has gone into understanding the structure and formation of cellulose as it is an important source of energy for ruminants. Bacteria and protozoa found within the rumen are responsible for the breakdown of starch and cellulose to volatile fatty acids (Moran 2005). The biosynthesis of lignin is well characterised due to its negative influence on the digestibility of animal feed (Jung 1989, Moran 2005) however the more recent need to understand the digestibility of biomass for the biofuel industry has driven this research further. A great deal is still unknown in terms of hemicellulose even though it accounts for 20 – 50% of all polysaccharides found in lignocellulosic material (Gomez *et al* 2008). There are a number of different types of hemicellulose that have been classified according to their structure; xyloglucans, xylans, mannans or glucomannans, and β-(1-3,1-4)-glucans (Scheller *et al* 2010).

Recent research has focused on trying to understand how the different types of hemicellulose are produced. While conducting research in Arabidopsis on the cellulose synthase (CesA) genes, which are involved in the production of cellulose, it was discovered that there is a second large superfamily of genes that share sequence homology (Richmond *et al* 2000). These cellulose synthase-like (Csl) proteins appear to be integral membrane proteins that contain a number of transmembrane domains in both the carboxyl terminal region as well as the amino terminal region. They also share a number of characteristics that indicate a glycosyltransferase function such as a D, D, D, QxxRW motif (Richmond *et al* 2000). It was hypothesised that these Csl genes were involved in the production of non-cellulosic polysaccharides (Richmond *et al* 2000). Further work has revealed that there are nine

(CslA – J) subfamilies of genes within this superfamily. Four of these subfamilies (CslA, C, D and E) have been found in both monocot and dicot plants. Whereas CslF, H and J are only present in grasses, CslG and B is found only in monocots. This difference in expression is probably related to the variation in hemicellulose structure of monocots and dicots (Youngs *et al* 2007, Dhugga 2012).

The role of the Csl enzymes in hemicellulose biosynthesis has been elucidated for some of the genes. It appears that different subfamilies are involved in the synthesis of assorted sugar backbones relating to the various types of hemicellulose (Scheller *et al* 2010). A summary of the function of the different subfamilies is given bellow:

- CslA: A number are involved in the biosynthesis of mannan and glucomannan and therefore known as mannan synthases. These enzymes either function on their own as a single unit or as a homomultimer. They are expressed in the Golgi before being exported by exocytosis to the cell wall. To date, only 4 of the 9 Arabidopsis CslA genes have been reported as producing mannan or glucomannan (Liepman *et al* 2005, Liepman *et al* 2007, Scheller *et al* 2010, Dhugga 2012).

- CslB: Has been reported to be phylogenetically related to CslH even though it is found in only dicots whereas CslH is only found in monocots (Yin *et al* 2009). The role of this subfamily is not clear as only low levels of expression have been detected in roots and seedlings of Arabidopsis therefore leading to speculation that these genes have a specialised role (Youngs *et al* 2007).

- CslC: Is closely related to the CslA subfamily and are thought to originate from a common ancestral gene (Liepman *et al* 2012). It has been observed that CslC4 is involved in the synthesis of the xyloglucan backbone and is known as β-1,4-glucan synthase (Dhugga 2012). It is thought to achieve this by forming a complex with other unknown proteins (Oikawa *et al* 2013).

- CslD: Research has identified CslD2 and 3 working together in a complex to produce glucomannan found within root hair. This result is surprising as it

has previously been recorded that CslD genes are thought to be involved in the biosynthesis of β-glucan (Liepman *et al* 2012, Oikawa *et al* 2013).

- CslE: Not much is known about this group of Csl genes except that it is widely expressed (Dhugga 2012).

- CslF and CslH: Have both been reported to be involved in the synthesis of β-(1-3,1-4)-linked glucans and are known as mixed-linked glucan synthases. These two genes as well as their proteins differ in structure and don't appear to work together in a complex. They both don't need to be present for biosynthesis to take place (Scheller *et al* 2010, Dhugga 2012).

- CslG: The function of this group of Csl genes is not clear however they are found to be expressed in the flowers and leaves of Arabidopsis (Youngs et al 2007).

- CslJ: Has been found to be closely related to CslG but it has been reported to have a different gene structure (Yin *et al* 2009).

As all the other types of hemicellulose are produced by Csl genes it was originally thought that the biosynthesis of the xylan backbone would also involve this gene family. However, this is not the case as a number of other glycosyltransferases (GT8, 43 and 47) have been shown to carry out this function (Carpita *et al* 2012, Dhugga 2012). These GTs are Type II membrane proteins that have a single N-terminal membrane anchor and are found in the Golgi apparatus. They do not have the characteristic catalytic subunit found in Csl genes (Dhugga 2012). Members of the GT8 family are thought to be involved in using α-linked donor substrates to create α-glycosidic bonds whereas the members of the GT43 and 47 families appear to have an inverting enzyme activity (Brown *et al* 2007). In some plant species, xylans have been shown to contain a reducing end with a unique structure. This structure has not been reported to be present in grass xylans (Scheller *et al* 2010, Chiniquy *et al* 2013).

Work has been done to understand the functioning of the various GTs in xylan backbone synthesis. Plants containing GT mutants have been identified as having an irregular xylem because of weakened xylem vessels. This results in the vessels collapsing inward due to not being able to withstand the negative pressure that is placed on them to allow for the flow of water (Petersen *et al* 2012). The various mutated genes have therefore been classified as *irregular xylem* (*irx*) mutations known as *irx9* (GT43), *irx10* (GT47) and *irx14* (GT43). The morphology of the mutants differ from wild type plants in that their stems are thinner and shorter however *irx14* is not as severely compromised. The cell walls of *irx* mutant Arabidopsis plants contain less xylose and cellulose than wild type plants and the proportion of glucuronic acid (GlcUA) and methylated-glucuronic acid (Me-GlcUA) side chains have been affected (Brown *et al* 2007). In the case of rice *irx10* mutants the culms are shorter than in wild type plants, which lead to smaller plants. There is also a reduction in cell wall thickness because of smaller vascular bundles. Not only is there a decrease in xylose but there is a small increase in arabinose. However, there is no change in the size of the xylan polymer found in the rice *irx10* mutant unlike that which is found in the Arabidopsis *irx10* mutant. The decrease in xylan content in rice is modest when compared to Arabidopsis (Chen *et al* 2013). It appears that IRX9, 10 and 14 proteins form a complex to produce the xylan backbone. It has been hypothesised that IRX14 is responsible for passing the UDP-xylose substrate to another protein, which is likely to be IRX10, while IRX9 has a more structural role in organising and possibly assembling the complex (Oikawa *et al* 2013, Ren *et al* 2014).

The candidate genes linked to digestibility, which were identified during the QTL analysis of the Brachypodium RIL population, were identified as possible GT43 (Bradi5g24290.1) and CslA (Bradi3g06740) genes. Homozygous Arabidopsis transfer DNA (T-DNA) insertion lines for the genes of interest were selected to confirm the functionality of the candidate genes. This was achieved by studying gene function of the Arabidopsis lines and comparing them to selected Brachypodium RILs that were grouped according to the allele on chromosome 5, which is linked to marker BD1617_1. Arabidopsis T-DNA lines were chosen as there is a large collection of lines covering almost all of the genes in the genome already available thereby cutting out the need to develop our own (Borevitz *et al* 2004).

In this study a comparison of the phenotype of the Arabidopsis T-DNA lines to Brachypodium RIL lines in terms of cell wall composition was conducted to determine if they have similar traits to confirm if the Brachypodium candidate genes, Bradi3g06740 and Bradi5g24290.1 are functionally similar to that of the Arabidopsis genes and are therefore a contributor to the recalcitrant nature of lignocellulose. A number of analysis methods were employed to determine if there was a difference in cell wall properties in terms of cellulose, hemicellulose and lignin.

5.2   Material and methods

5.2.1 Phylogenetic trees

Protein sequences were collated for the Csl genes, subfamilies A – H, for *Arabidopsis thaliana* (Arabidopsis), *Oryza sativa* (rice), *Hordeum vulgare L.* (barley) and *Brachypodium distachyon* (Brachypodium). The phylogenetic tree was produced using ClustalX2 (Larkin *et al* 2007), which created a guide tree. The guide tree was bootstrapped to create a Phylip format tree, which was visualised using Dendroscope (Huson *et al* 2007). For the IRX tree the protein sequences were assembled for the subfamilies 9 and 14 from Arabidopsis, rice and Brachypodium using the same method.

5.2.2 Plant material

Powder from 24 Brachypodium RILs that were selected in Chapter 4 was used for the analysis of the cell wall. These plants had been grown from 12 seeds per plant line as and ground into a powder as described in Chapter 4.

The Arabidopsis T-DNA lines in Table 5.1 were purchased from NASC and underwent confirmation for homozygosity using PCR. These homozygous seeds were then sown in P40 plant trays as single seeds per well and were vernalized at 4ºC for four days. They were then moved to a growth room and left to grow for approximately six weeks before watering was stopped. Once the plants were

completely dry, the stems were harvested and ground to a powder in the grinding and weighing robot (Labman Automation Ltd, UK).

Table 5.1: The Arabidopsis T-DNA lines purchased from NASC corresponding to the candidate genes Bradi5g24290.1 and Bradi3g06740.

Gene	ID	SALK ID	NASC ID
At4g36890	GT43	SALK_038212	N538212
At5g22740	CslA2.1	SALK_149092C	N655198
At5g22740	CslA2.2	SALK_075579C	N666448
At5g03760	CslA9	SALK_111096C	N663662
At5g16190	CslA11.1	SALK_065682	N565682
At5g16190	CslA11.2	SALK_136121	N636121

5.2.3 Saccharification analysis

The ground powder from the Brachypodium RILs and the Arabidopsis T-DNA lines were formatted into 96-well plates before being screened using the liquid handling robot as described in previous chapters.

5.2.4 Composition analysis

5.2.4.1   Monosaccharide analysis of hemicellulose fraction

The Brachypodium RILs and the Arabidopsis T-DNA lines were analysed for monosaccharide cell wall content. The neutral sugars were analysed from the dry powder sample using HPAEC on a Dionex ICS-3000, with a Carbopac P20 column (Dionex, UK) (Jones *et al*, 2003). The samples were prepared using an acid hydrolysis protocol, which involved adding 500ul 2M trifluoroacetic acid (TFA) to the samples and flushing with Argon gas. The samples were incubated at 100ºC for four hours before evaporating overnight. The following day the samples were washed twice with 200 ul isopropyl alcohol before resuspending in 500 ul distilled water. 400 ul supernatant was removed for monosaccharide analysis and the

remaining pellet was dried for further cellulose analysis. Three replicates of each sample were analysed.

5.2.4.2   ATR-FTIR analysis

This analysis was run on the powdered Brachypodium RIL samples as well as the Arabidopsis T-DNA lines. Each plant line was analysed three times using ATR-FTIR spectroscopy (PerkinElmer, UK) and the data was analysed using Unscrambler X following the same method described in chapter 2.

5.2.4.3   Cellulose analysis

Powdered samples from the Brachypodium RILs and the Arabidopsis T-DNA lines were analysed for cellulose crystallinity by XRD. The samples were screened by Tengyao Jiang at the department of Green Chemistry at the University of York.

The insoluble cellulose content of both the Brachypodium and Arabidopsis plant lines were also determined using wet chemistry involving the Updegraff method which includes using an anthrone reagent (Foster *et al* 2010b). In this method 1 ml Updegraff reagent was added to the dry pellet produced during sample preparation for monosaccharide analysis. The sample was incubated at 100ºC for 30 minutes. The samples were then cooled to room temperature before centrifugation at 10000 rpm for 15 minutes. The supernatant was discarded and the pellet was washed with 1 m dH_20 and centrifuged for 15 minutes at 10000 rpm. The supernatant was discarded again and the pellet was washed a further three times with 1 ml acetone. The pellet was left to dry overnight at room temperature. The next day 175 ul 72% sulfuric acid was added to the pellet and left to incubate at room temperature for 30 minutes followed by vortexing and incubating for a further 15 minutes. Then 825 ul dH_2O was added and the samples were left to incubate for four hours at 120 ºC. Finally, the samples were centrifuged for 5 minutes at 10000 rpm. The amount of glucose present was quantified using a glucose standard. 40 ul of the sample was prepared by adding 360 ul dH_20 and 800 ul Anthrone reagent. The samples and glucose standards were incubated at 80 ºC for 30 minutes before transferring 200 ul to an optical plate

so that the absorption could be measured at 620 nm. All the samples were analysed in triplicate.

5.2.4.4 Hemicellulose analysis

An analysis of the molecular weight of the xylan chain of the Brachypodium lines Bd149, Bd166, Bd163 and Bd176, which represented the different alleles, was determined using a size-exclusion chromatography (SEC) method (Brown *et al* 2009). Dr Leonardo Gomez, University of York extracted the xylan and the samples were separated on a HPLC connected to a multi-angle light scattering detector and a refractive index detector system run by Dr Andrew Leech, University of York. The sample injection volume was 100 ul. Blank buffer injections were used to check for carry-over between sample runs. The data was analysed using the Astra V software and the molecular weights were estimated using the Zimm fit method with degree 1. The sample refractive index increment (dn/dc) was 0.145.

5.2.4.5 Lignin analysis

An acid soluble method involving acetyl bromide was followed to determine the total lignin content of the samples from the Brachypodium RILs and the Arabidopsis T-DNA lines (Foster *et al* 2010a). 5 mg of ground sample was weighed out in triplicate and 250 ul acetyl bromide solution (25% v/v acetyl bromide:75% glacial acetic acid) was added. The samples were incubated for three hours at 50 ºC and then cooled to room temperature before the supernatant was transferred to a 5 ml volumetric flask. 1 ml 2 M NaOH and 175 ul 0.5 M hydroxylamine HCl was added before the flasks were filled to the 5 ml mark with glacial acetic acid. The solution was mixed by inverting. 100 ul of the sample was transferred to a 1.5 ml tube and 900 ul acetic acid was added to make a 1:10 dilution. The sample was vortexed before transferral to a 1ml quartz cuvette for measurement of absorption at 280 nm (Shimadzu UV-1800 Spectrophotometer).

5.3   Results and discussion

5.3.1 Phylogenetic trees

Phylogenetic trees were produced for each candidate to determine how they group with similar proteins from different plant species. The protein sequences of the Csl superfamily were collected from Arabidopsis, rice, barley and Brachypodium. Figure 5.1 shows the groupings of all the sequences into the subfamilies A − H. The Brachypodium candidate gene Bradi3g06730 was found together with other CslA sequences as expected. It was found specifically together with CslA3, A7 and A9 from Arabidopsis and CslA1 and A9 from rice.
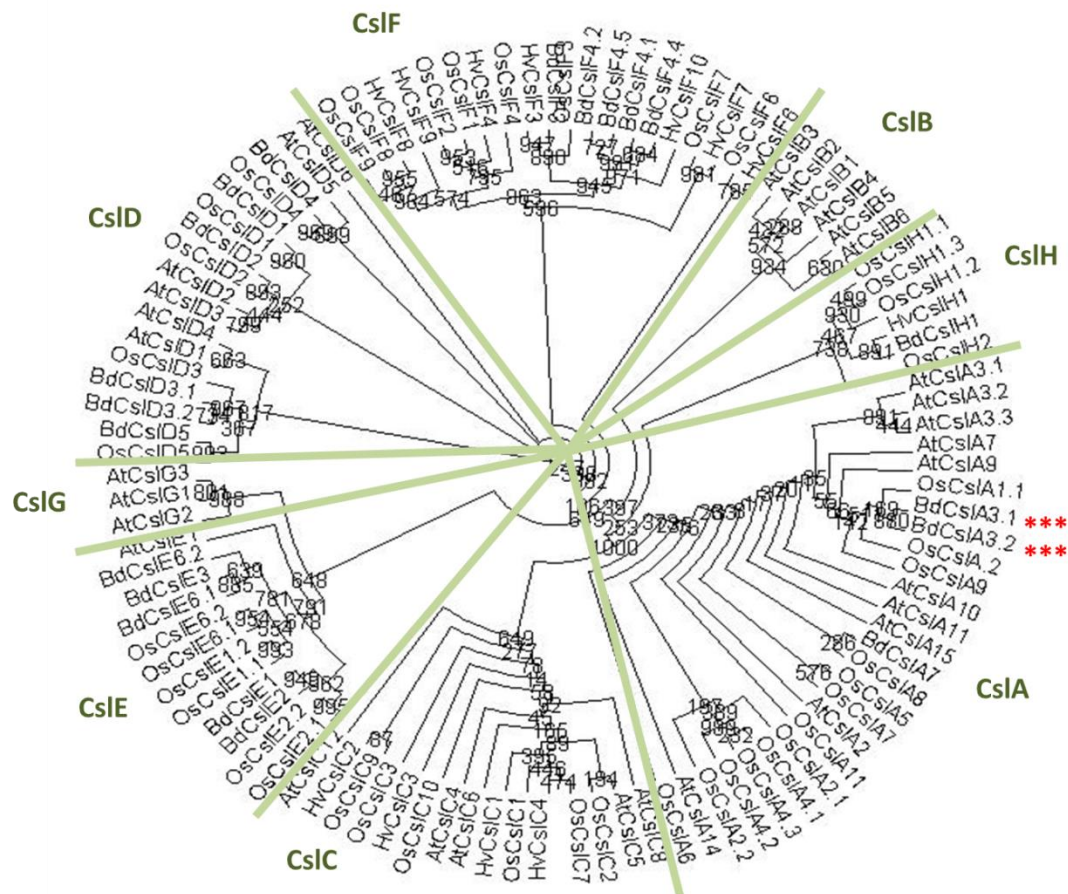
Figure 5.1: A phylogenetic tree based on protein sequence similarity, produced using ClustalX2 and visualised using Dendroscope of the Csl proteins including subfamilies A - H found in Arabidopsis, Rice, Barley and Brachypodium indicating the position of the Brachypodium candidate (highlighted in red) within the CslA cluster.

137

The protein sequences from Arabidopsis, rice and Brachypodium were collated for GT43 (IRX9 and 14). Figure 5.2 indicates that the Brachypodium candidate Bradi5g24290 groups together with other IRX14 proteins as expected.
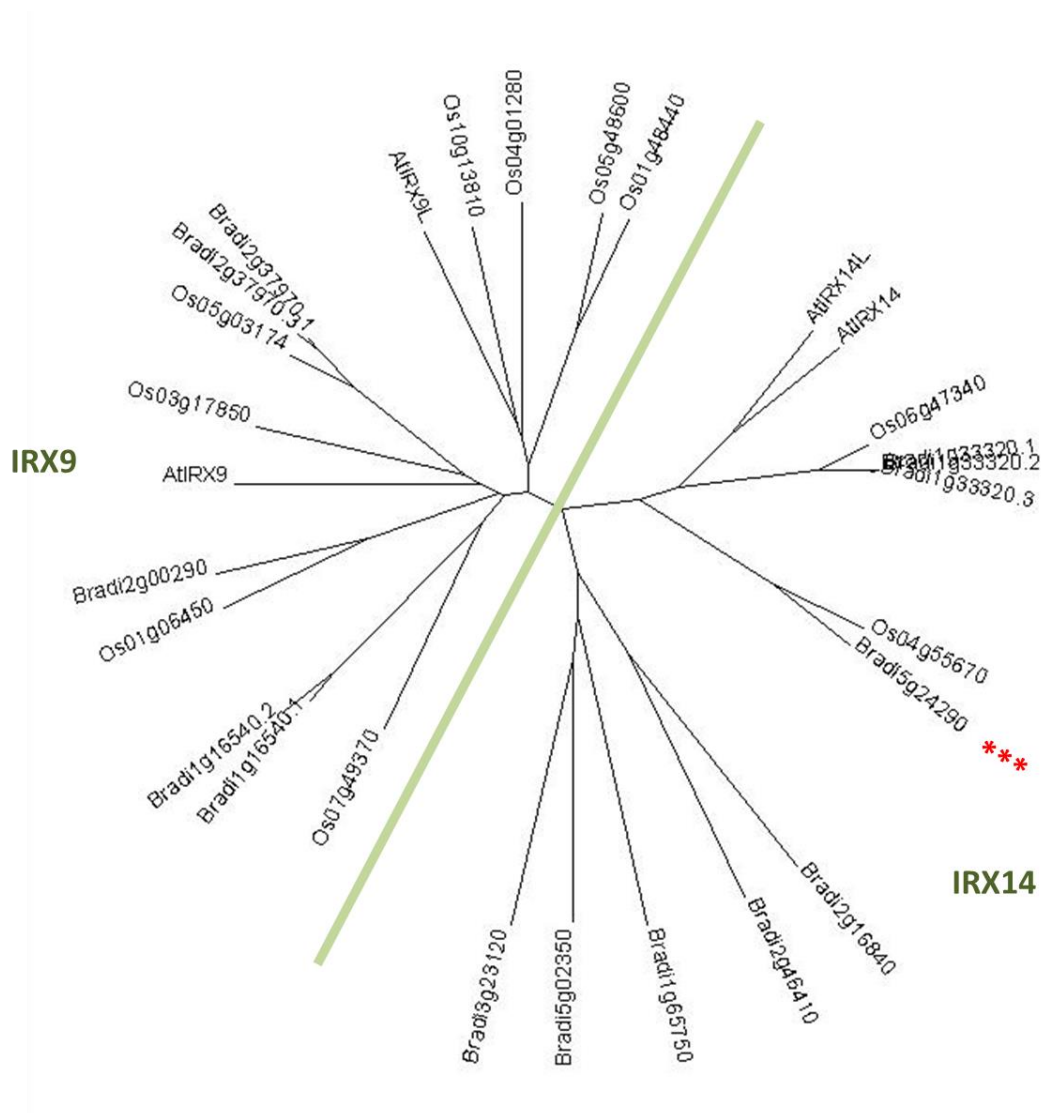
Figure 5.2: A phylogenetic tree based on protein sequence similarity, created using ClustalX2 and visualised using Dendroscope of the GT43 family consisting of IRX9 and IRX14 proteins found in Arabidopsis, rice and Brachypodium indicating the position of the Brachypodium candidate (highlighted in red) within the IRX14 cluster.

5.3.2 Saccharification analysis

The saccharification analysis of the 24 Brachypodium lines selected in chapter 4 that have similar genotypic backgrounds, allele AA at marker BD1415_1 and either allele

AA or BB at marker BD1676_1 was carried out again at the same time as the analysis of the Arabidopsis T-DNA lines to reduce any laboratory variation. As expected, the Brachypodium plant lines containing the allele AA at marker BD1676_1 were significantly more digestible than those containing allele BB (one-way ANOVA: *F = 3.56; df = 1, 19; Pr = 0.075*) (Figure 5.3A).

The digestibility of various Arabidopsis T-DNA insertion lines were analysed to determine if there was a significant difference between the digestion potential of wild type (Col.0) and the T-DNA insertion lines of At4g36890 (GT43), At5g22740 (CslA2.1 and 2.2), At5g03760 (CslA9) and At5g16190 (CslA11.1 and 11.2). Figure 5.3B indicates that there was a significant decrease in digestibility of the GT43 mutant line when compared with Col.0 (one-way ANOVA: *F = 132.7; df = 1, 6; Pr = 2.57x10^{-5}*). All the CslA lines, except for CslA11.1 showed some significant decrease in digestibility with CslA2.2 containing the most significant decrease (one-way ANOVA: *F = 8.82; df = 1, 6; Pr = 0.025*). These results mirror those that were obtained during QTL analysis in terms of marker BD1676_1, which is linked to the candidate gene Bradi5g24290.1, having the larger effect on digestibility. Whereas, marker BD1415_1 linked to the candidate gene Bradi3g06730, has a lesser effect (Figure 3.12C).
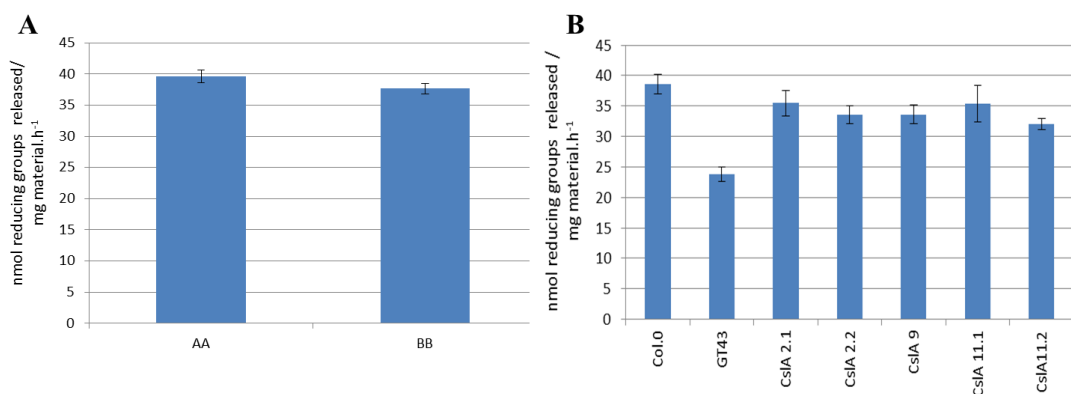
Figure 5.3: Saccharification analysis of the ground stem biomass of (A) Brachypodium RILs compared according to the allele at marker BD1676_1 and the (B) Arabidopsis T-DNA lines compared with wild type Col.0. Digestion was conducted using a commercial cellulase for 8 hours at 50ºC following a pretreatment of 0.5 N NaOH for 30 minutes at 90 ºC. The results are the means and standard deviations of eight replicates.

5.3.3 Cell wall composition analysis

It was decided to examine the cell wall composition of the Brachypodium plant lines to look for differences that might accompany the variation in digestibility. A similar approach was taken with the Arabidopsis T-DNA lines. Initially the work focused on looking at the various neutral sugars found within the cell wall to see if that would point in a specific direction. Then it went on to look at the three main components of cell walls namely; cellulose, lignin and hemicellulose.

5.3.3.1   Monosaccharide analysis of hemicellulose fraction

The neutral sugars were analysed from dry plant material of both the Brachypodium RILs and the Arabidopsis T-DNA lines.

The results obtained from the analysis of the neutral sugars from the Brachypodium RILs were compared using one-way ANOVA to determine if there was a significant difference in the amount of each sugar present when looking specifically at the genotype of the plants (Figure 5.4A). It was determined that there was no significant difference between plant lines containing allele AA or BB for any of the neutral sugars.

In the case of the Arabidopsis samples the different plant lines were compared to the wild type Col.0 to determine if there was a significant difference in the different sugars present. Only CslA2.1 shows a significant increase in glucose present (one-way ANOVA: $F = 15.98; df = 1, 4; Pr = 0.016$) when compared to Col.0. The levels of xylose present in GT43 were significantly lower than Col.0 (one-way ANOVA: $F = 12.22; df = 1, 4; Pr = 0.025$). This result was expected as it has been reported that changes in At4g36890 result in an *irx* mutant. It is thought that this gene is responsible for the formation of the xylose backbone found in hemicellulose (Brown *et al* 2007, Carpita *et al* 2012, Dhugga 2012). Finally, from the data it was noted that there was a significant increase in mannose in GT43 (one-way ANOVA: $F = 8.572; df = 1, 4; Pr = 0.043$) whereas CslA9 showed a significant decrease (one-way ANOVA: $F = 23.36; df = 1, 4; Pr = 0.008$) when compared with the wild

type. Previous research has reported the change in mannose for *cslA* mutant plants however this has not been mentioned in the case of At4g36890 (Scheller *et al* 2010).

From these results, it can be concluded that when the GT43 or CslA genes have been knocked out, as in the case of the T-DNA lines, it leads to a significant change in the composition of the cell wall in terms of sugars present. The main changes occur in the amount of glucose, xylose or mannose present. There were no clear differences in composition in the RILs carrying the alternative alleles in these genes, but this may reflect only partial differences in gene product activity rather than complete knockouts.
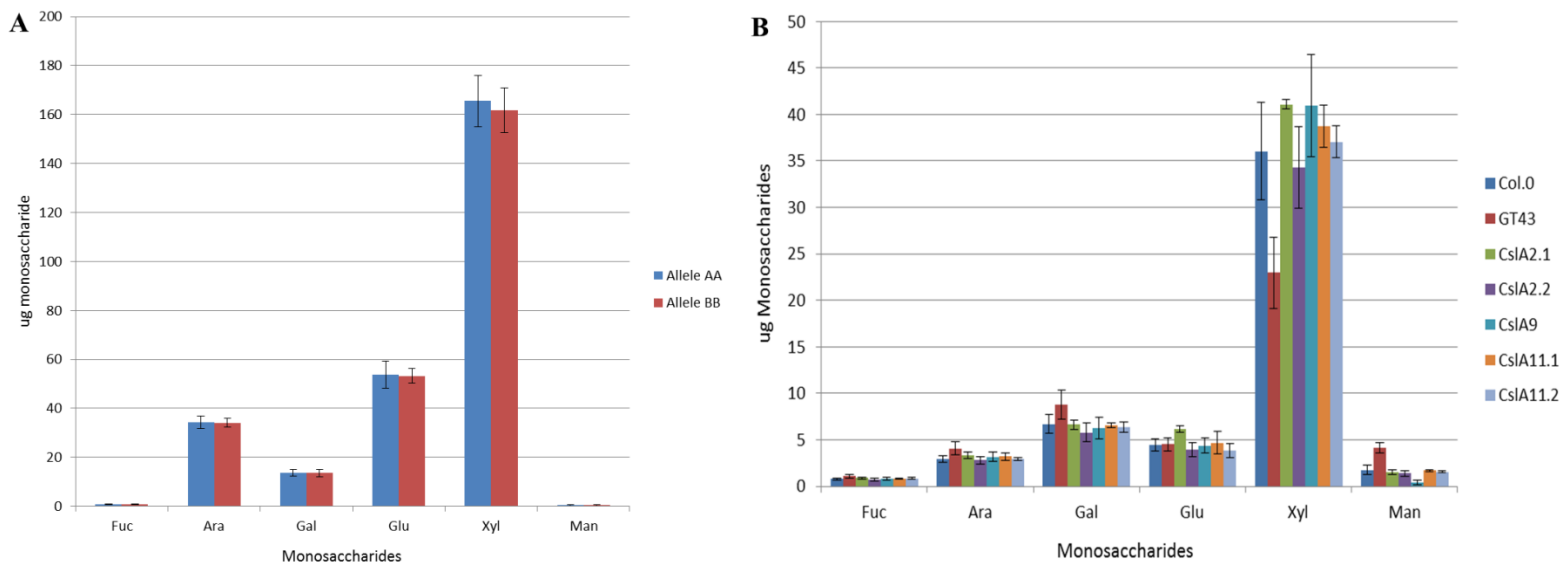
Figure 5.4: Monosaccharide analysis of the neutral sugars released from the hemicellulose fraction of ground dry biomass (A) from Brachypodium RILs compared according to the allele found at marker BD1676_1 and (B) from Arabidopsis T-DNA lines compared to wild type Col.0. The samples were prepared using a 2M TFA protocol and they were run using HPAEC on a Dionex together with quantifiable standards. The results are the means and standard deviations of three replicates.

5.3.3.2   ATR-FTIR analysis

Powdered stem material from the Brachypodium lines as well as the Arabidopsis T-DNA lines were screened using ATR-FTIR to give an indication of where in the plant cell wall any changes were present. This method has been used in a number of studies looking at the composition of plant cell walls because it is easy to use and only requires a small sample (Abidi *et al* 2014).

The spectral data from the Brachypodium lines were combined according to the allele present at marker BD1676_1 and the average data was used to produce individual spectra for those with allele AA versus those with BB (Figure 5.5). When comparing the spectra for the region $850 - 1700$ cm^{-1} it was noted that most of the variation between the two alleles was seen in the region of $1200 - 1600$ cm^{-1}. It has been reported that the peaks at 1375 and 1420 cm^{-1} are indicative to changes in cellulose structure, namely CH bending and CH_2 symmetric bending respectively (Carrillo *et al* 2004). A small difference in peak height was also noted at 1035 cm^{-1} which has been linked to a change in glucose (Kacurakova *et al* 2000). From the spectra it was concluded that the plants with allele AA have reduced cellulose compared with those that have the BB allele as the intensity of the peak at 1420 cm-1 is lower.
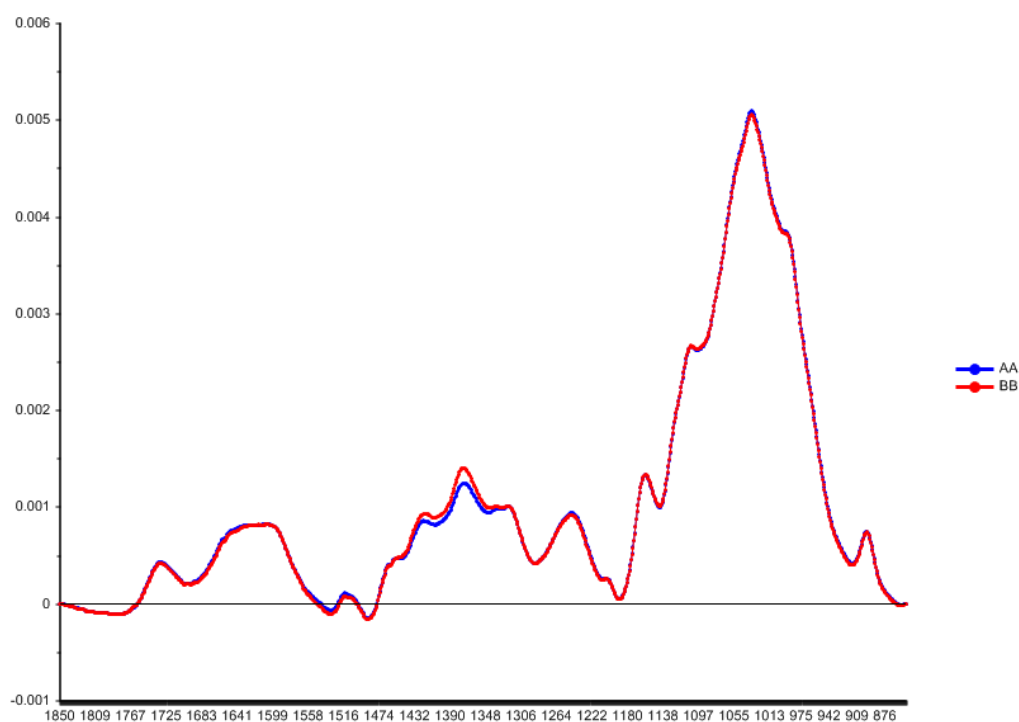
Figure 5.5: Comparison of the ATR-FTIR analysis results of the Brachypodium RILs according to their genotype at marker BD1676_1. The spectrum is an average of three replicates per sample analysed over the region 1850 – 850 cm^{-1} at a resolution of 4 cm^{-1} for 256 scans. All data was peak normalised before analysis using UnscramblerX.

For the Arabidopsis T-DNA lines, GT43 was compared to the wild type to determine if there was a difference between them in terms of cell wall composition (Figure 5.6). Again the main area of variation was those lying at wavelength 1600 – 1200 cm^{-1}, which is related to changes in cellulose. The region 1200 – 800 cm^{-1} indicating changes in polysaccharides such as hemicelluloses was also affected. However, this region is difficult to pick apart as it is influenced both by the bonds that make up the backbones of the different types of hemicelluloses as well as the individual monosaccharides that are found on the side chains of these compounds. For example, the vibration of β-(1-4)-mannan is found from 1066 – 1064 cm^{-1} and mannose is found at 1070 cm^{-1}(Kačuráková *et al* 2000 and Kačuráková *et al* 2001). Whereas, xylan has a number of vibrations found at 1240, 1128, 1082, 1045 and 978 cm^{-1} (Brown *et al* 2005) and xylose has had peaks identified at 1173, 1041, 972 and 900 cm^{-1} (Coimbra *et al* 1999).
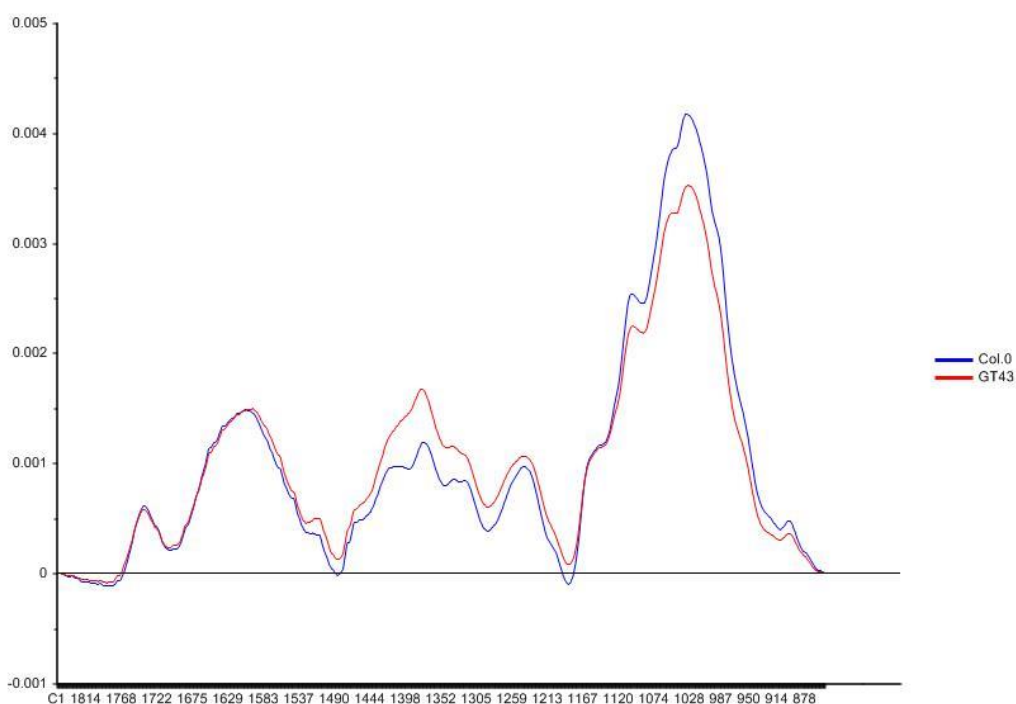
Figure 5.6: Comparison of the ATR-FTIR analysis results of the Arabidopsis T-DNA GT43 line to the wild type Col.0. The spectrum is an average of three replicates per sample analysed over the region 1850 – 850 cm^{-1} at a resolution of 4 cm^{-1} for 256 scans. All data was peak normalised before analysis using UnscramblerX.

Secondly, the various Arabidopsis T-DNA lines belonging to the CslA family were compared to the wild type and PCA was conducted to determine at what wavelengths the difference in cell walls occurred (Figure 5.7). From this analysis we could determine that there weren't any clear clusters caused by similar samples grouping together however they all did differ from the wild type. The main differences lay within the PC-2 component (Figure 5.7C), which highlighted the cellulose region at wavelength 1600 – 1200 cm^{-1} as well as the region containing various monosaccharides and polysaccharides at wavelength 1200 – 800 cm^{-1}.
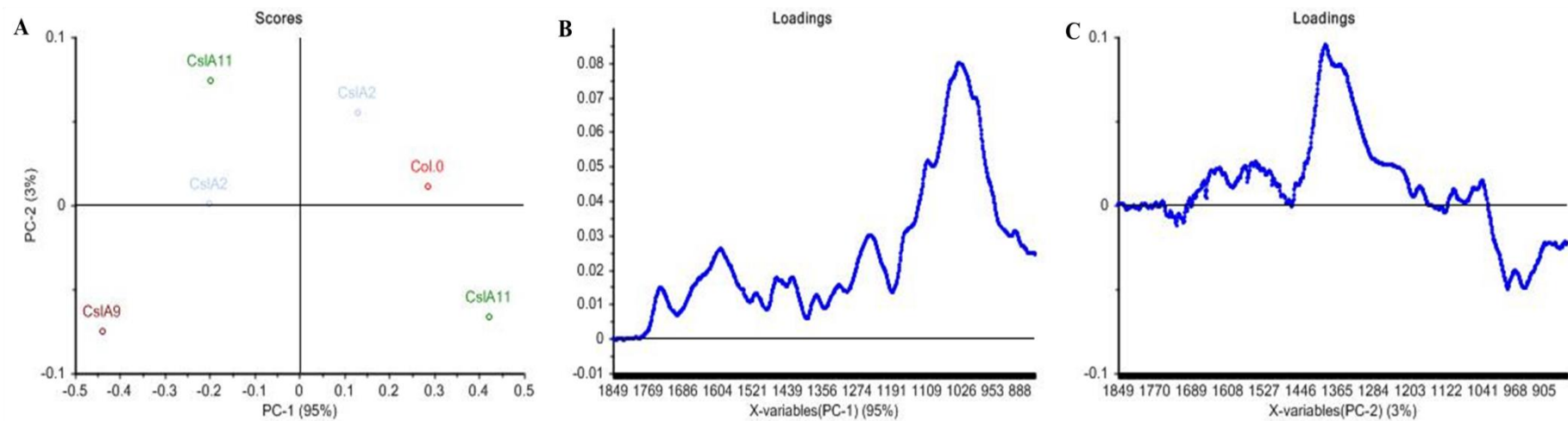
Figure 5.7: The ATR-FTIR analysis results of the various Arabidopsis T-DNA CslA lines showing the (A) PCA scores, (B) loading for PC-1 and (C) PC-2. Three replicates per sample were analysed over the region 1850 – 850cm^{-1} at a resolution of 4cm^{-1} for 256 scans. All data was peak normalised before PCA analysis using UnscramblerX.

It is possible to determine the ratio of crystallinity (Cr.R), also known as lateral order index (LOI) by using the absorbance (A) at wavelengths 1430 cm^{-1} and 893 cm^{-1} (Carrillo *et al* 2004, Ciolacu *et al* 2011). It is used to describe the relative amount of cellulose that is found in crystalline form (Liu 2013). The following formula was used to determine the degree of cellulose crystallinity for each of the alleles and T-DNA lines:

$$Cr.R = A_{1430}/A_{893}$$

The peak at A_{1430} is classified as the crystallinity band as it has been linked to the vibration caused by the symmetric bending of CH_2 in cellulose. It has been reported that a decrease in value at this point of the spectrum indicates a decrease in the degree of crystallinity of cellulose (Ciolacu *et al* 2011). It has also been reported that the Cr.R ratio increases when the crystallinity of cellulose decreases (Carrillo *et al* 2004). In the case of the GT43 Brachypodium alleles, the A1430 value for allele AA is lower than BB, which indicates that AA plants have decreased cellulose crystallinity. This finding is confirmed when referring to the Cr.R ration that is increased in the AA allele plants in comparison to the BB allele, therefore indicating again a decrease in cellulose crystallinity for AA allele plants. The change in the degree of cellulose crystallinity (ΔCr.R) between allele AA and BB was calculated as 0.3%.

In the case of the Arabidopsis T-DNA lines it is not as clear as to the change occurring in the degree of cellulose crystallinity. In the case of the GT43 line it has a lower A_{1430} value when compared to Col.0, which indicates a decrease in the degree of crystallinity however the Cr.R value is lower in the GT43 T-DNA line, which indicates an increase in crystallinity. Therefore, these results for the GT43 T-DNA lines are contradicting. Only lines CslA2.2, CslA9 and CslA11.2 appear to have a decrease in cellulose crystallinity as they a lower A_{1430} value and a higher Cr.R value when compared with Col.0. These ΔCr.R values range from 0.3 to 0.9%. The changes that were observed could possibly still explain some of the difference we see in cell wall digestibility as it is known that the crystallinity of cellulose plays a role. It is thought that amorphous cellulose regions within the crystalline cellulose structure are broken down first and then only the more recalcitrant crystalline regions

are digested. It has also been shown that the degree of cellulose crystallinity affects the absorption capacity of the cellulose to attach enzymes onto its surface therefore leading to different rates of digestion (Hall *et al* 2010).

Table 5.2: The percentage change in degree of cellulose crystallinity (ΔCr.R) calculated from ATR-FTIR data for Brachypodium plant lines as well as Arabidopsis T-DNA lines. Increase or decrease is indicated by a – or +.

	A_{1430} (cm^{-1})	A_{893} (cm^{-1})	Cr.R	ΔCr.R (%)
AA	92.848	89.509	1.037	+ 0.3
BB	92.881	89.797	1.034	
Col.0	91.924	88.996	1.033	
GT43	84.672	82.610	1.025	- 0.8
CslA 2.1	91.984	88.836	1.035	+ 0.3
CslA 2.2	91.022	87.492	1.040	+ 0.7
CslA 9	90.192	86.880	1.038	+ 0.5
CslA 11.1	92.051	89.905	1.024	- 0.9
CslA 11.2	90.906	87.286	1.041	+ 0.9

ATR-FTIR is only a rough method for the analysis of the composition of the cell wall so any findings have to be backed up by further analysis (Hori *et al* 2003) therefore; further analysis focusing on cellulose was conducted.

5.3.3.3   Cellulose analysis

The first method used to analyse the cellulose content found within the plant cell walls was XRD analysis, which was conducted by Tengyao Jiang at the department of Green Chemistry, University of York. This is a rapid method that can be used to identify the phase of cellulose, for example if it is a cellulose Iβ, cellulose II or amorphous state. The data can also be used to determine the crystallinity index (CI), which is the relative amount of crystalline material within a sample using the peak height method (Harris *et al* 2008, Hall *et al* 2010, Park *et al* 2010);

$$CI (\%) = (I_{002} - I_{AM}/I_{002})*100$$

Where I_{002} = maximum intensity at $2\theta = 22.5º$

and I_{AM} = minimum intensity at $2\theta = 18°$

The Brachypodium lines were analysed according to allele type for the marker on chromosome 5. From the spectra it was observed that the plants with allele AA for marker BD1676_1 had a higher crystallinity (Figure 5.8).
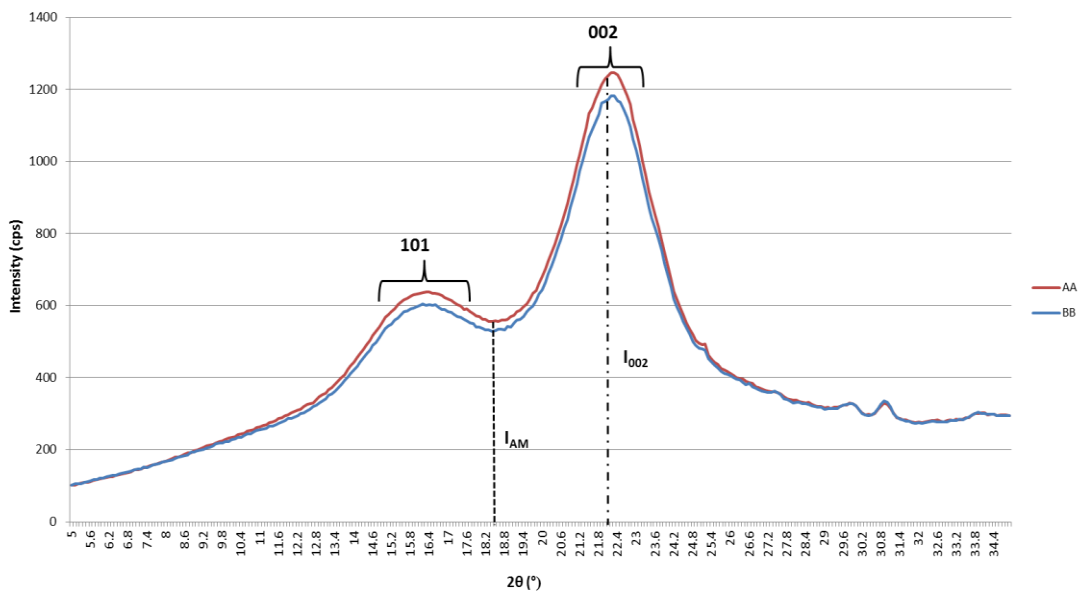
Figure 5.8: Comparison of the XRD analysis results of the Brachypodium RILs according to their genotype at marker BD1676_1. The samples were analysed by Tengyao Jiang at the department of Green Chemistry, University of York.

The CI values for the different Brachypodium alleles indicate that the AA plants have a slightly higher crystallinity than the BB alleles (Table 5.3). This difference in crystallinity (ΔCI) was observed as 0.05 which is not statistically significant. This result doesn't support those obtained during the ATR-FTIR analysis as those with allele AA appeared to have a decrease in the degree of crystallinity compared with allele BB plants whereas the XRD results appear to shown an increase in crystallinity for allele AA lines. The values of CI obtained for each allele is similar to those previously reported by Harris *et al* 2008 who analysed 22 different grasses and recorded values between 51.1 and 58.5%

Table 5.3: The change in the crystallinity index (ΔCI) calculated from XRD data for Brachypodium plant lines as well as Arabidopsis T-DNA lines.

	I002	IAM	CI (%)	ΔCI
AA	4983	2226	55.33	
BB	4711	2107	55.27	0.05
Col.0	2366	1254	47.00	
GT43	2536	1565	38.29	8.71
CslA 2.1	3148	1574	50.00	3.00
CslA 2.2	3132	1609	48.63	1.63
CslA 9	3309	1752	47.05	0.05
CslA 11.1	2928	1512	48.36	1.36
CslA 11.2	3711	1854	50.04	3.04

The XRD results from the analysis of the GT43 Arabidopsis T-DNA line was compared with those obtained from Col.0 (Figure 5.9) and it was noted that the GT43 lines had less crystalline cellulose than the wild type. This is an 8.71 difference in CI (Table 5.3).
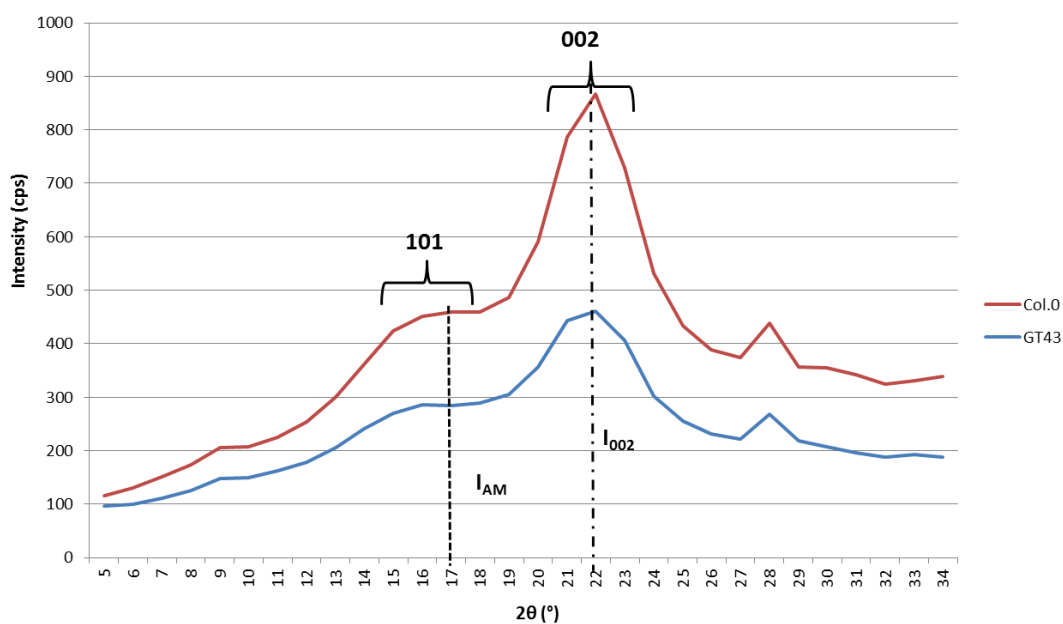
Figure 5.9: Comparison of the XRD analysis results of the GT43 Arabidopsis T-DNA line to wild type Col.0. The samples were analysed by Tengyao Jiang at the department of Green Chemistry, University of York.

The various CslA Arabidopsis T-DNA lines were compared to the wild type (Figure 5.10). From the spectra and the CI calculations it was determined that all the CslA

lines have an increase in cellulose crystallinity in comparison to the wild type (Table 5.3). These results don't support the ATR-FTIR findings except for CslA11.1 which had an increase in cellulose crystallinity (Table 5.2). The CI value determined during XRD is similar to previously reported observations for Arabidopsis of 54.8% for stem only measurements (Harris *et al* 2008).
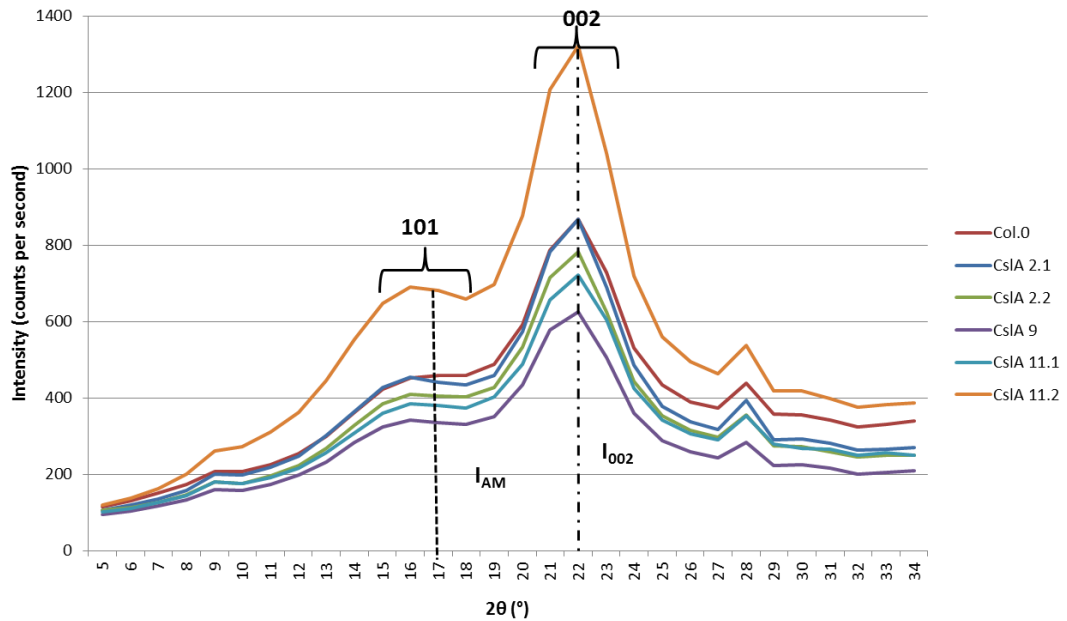
Figure 5.10: Comparison of the XRD analysis results of the Arabidopsis T-DNA lines for various CslA genes to wild type Col.0. The samples were analysed by Tengyao Jiang at the department of Green Chemistry, University of York.

The amount of crystalline cellulose within the cell wall was also analysed using an alcohol-insoluble residue (AIR) method that involves removing soluble sugars by conducting a number of ethanol washes. The resultant cellulose, which is from the crystalline fraction, is then measured using a colourimetric test involving Anthrone reagent (Foster *et al* 2010b).

The Brachypodium lines were screened and compared based on the allele at marker BD1676_1 (Figure 5.11A). Plants containing allele AA appeared to contain less crystalline cellulose, which agrees with the results that were observed during the ATR-FTIR analysis at the 1430 cm^{-1} peak, though again this is not a statistically significant difference. The difference in the percentage of crystalline cellulose between plants with allele AA and those with allele BB was calculated as 2.4%. The

Arabidopsis T-DNA lines also appear to contain less crystalline cellulose compared to the wild type, Col.0, however the only one that appears to be statistically significant is CslA2.1 (two-sample t-test: $t = 4.829; df = 4; p\text{-}value = 0.008$). A two-sample t-test was chosen as two independent groups of continuous data were being compared which followed a normal distribution and had equal variance. A one-way ANOVA could have been used for the analysis instead as either is an acceptable method used for analysing the difference between independent groups of data. The values obtained during this analysis are a little lower to those previously reported for soluble cellulose content for grasses as well as Arabidopsis. In the case of grasses it has been observed that Napier grass and reed canary grass has a cellulose content of 29.76 – 38.75%, whereas corn stover has been determined to contain 32% soluble cellulose (Thygesen *et al* 2005, Oleszek M *et al* 2014, Mohammed *et al* 2015). In the case of Arabidopsis Col.0 it has been measured as 30 – 35% and for *irx* mutants it has been observed as low as 5 – 15% (Turner *et al* 1997). The variation observed between the results could be due to differences within the methods used for soluble cellulose determination.
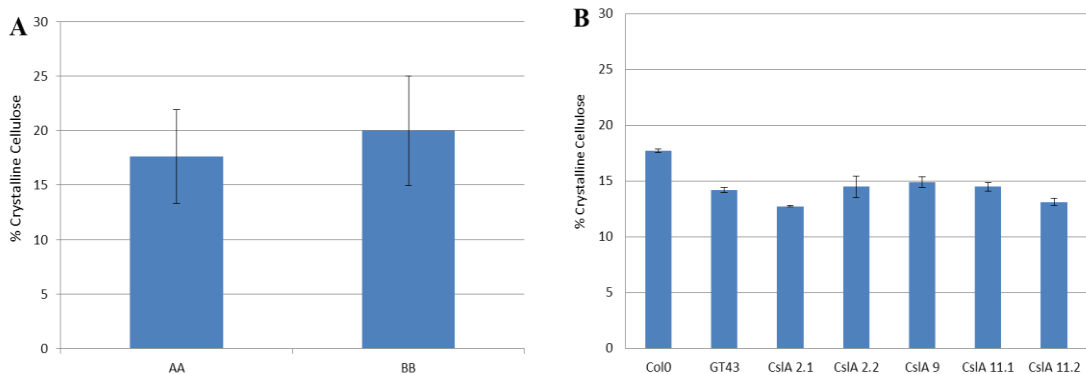
Figure 5.11: The crystalline cellulose content of the ground material of the (A) Brachypodium RILs compared according to allele at marker BD1676_1 and the (B) Arabidopsis T-DNA lines were determined using the Updegraff method (Foster *et al* 2010b). The insoluble cellulose content was quantified in triplicate using a glucose standard and absorption was measured at 620nm. The results are the mean and standard deviation.

In summary, in terms of cellulose crystallinity it appears that the Brachypodium AA allele lines contain a small decrease in crystalline cellulose when compared with plants containing allele BB. The Arabidopsis GT43 and CslA T-DNA lines also

show a small decrease in cellulose crystallinity compared to wild type, except for CslA11.1.

5.3.3.4   Hemicellulose analysis

The xylan chain length was analysed using a SEC method by Dr Leonardo Gomez and Dr Andrew Leech, University of York. The Brachypodium lines which represented the two alleles were analysed and the results in Figure 5.12 indicate that there is no difference between the alleles because the peak that represents xylan eluted at the same time point of 15 min for both alleles. This indicates that the xylan chains in all the samples have the same molecular weight and therefore the same length. In previous work, conducted by Brown *et al* 2009, they determined that *irx9* and *irx10* mutant's show an increase in elution time and therefore a decrease in molecular weight, which is related to a decrease in xylan chain length.
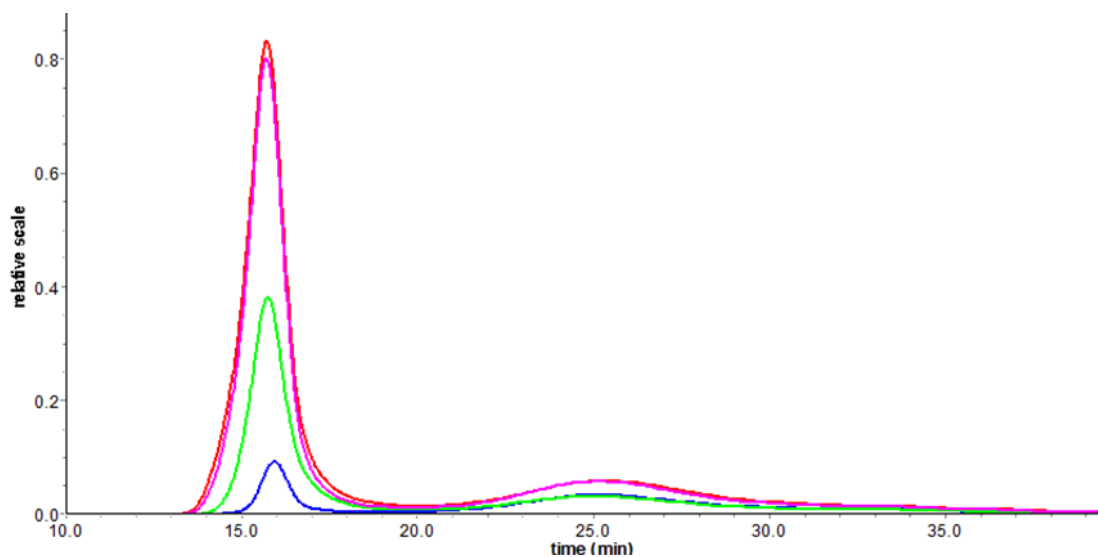
Figure 5.12: Xylan chain length determination using SEC analysis of the ground material from the Brachypodium RILs representing allele AA; Bd149 (red) and Bd166 (blue) as well as allele BB; Bd163 (green) and Bd176 (magenta). A 100ul sample injection was used and data was analysed using AstraV software using the Zimm Fit method to estimate molecular weight. The sample refractive index increment was 0.145.

5.3.3.5 Lignin analysis

Lignin analysis was undertaken using an acid soluble lignin method to determine if there was a difference in the amount present between the two allele groups of Brachypodium lines (Figure 5.13A) as well as between the Arabidopsis T-DNA lines and Col.0 (Figure 5.13B).

It was observed that the Brachypodium lines containing allele AA for marker BD1676_1 had a slightly higher percentage of lignin within their walls at 17.5% compared with those containing the allele BB at 16.5%. This difference in the amount of lignin was determined to be not statistically significant. In the case of the Arabidopsis T-DNA, they all had an increase in lignin compared to Col.0 except for CslA11.1. Only CslA2.1 is significantly different to the wild type (two-sample t-test: $t = -3.197; df = 4; p\text{-}value = 0.033$). It has previously been reported that some grasses such as Napier grass can have a higher lignin content (27%) compared with other grasses such as reed canary grass (8%). Brachypodium appears to have towards the higher range of lignin content (Oleszek *et al* 2014, Mohammed *et al* 2015). Whereas the agricultural waste biomass, corn stover has been reported to contain 6% lignin (Thygesen *et al* 2005).
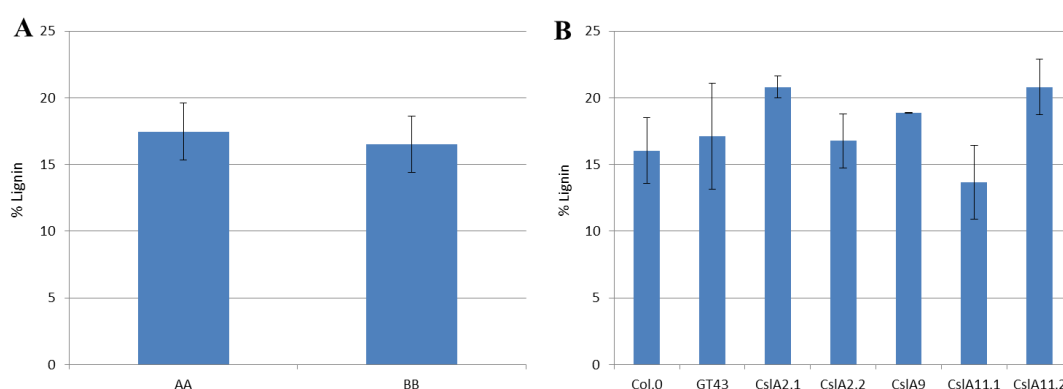
Figure 5.13: The lignin content of the (A) Brachypodium RILs compared according to the allele at marker BD1676_1 and the (B) Arabidopsis T-DNA lines were determined using an acid soluble method involving acetyl bromide (Foster *et al* 2010a). The results are the mean and indicate the standard deviation in triplicate.

5.4   Conclusion

The aim of this research was two-fold. In the first instance, it was to determine if there was a phenotypic difference in terms of cell wall composition between the alleles of selected Brachypodium lines. Secondly, the cell wall composition of these Brachypodium lines was compared to that of selected Arabidopsis T-DNA lines to determine if they have similar changes in cell wall structure.

When the Brachypodium alleles were compared it was determined that there was a change in saccharification potential with allele AA being significantly more digestible than allele BB. ATR-FTIR was used to provide a snapshot of the entire cell wall composition to determine if there were any changes to the cell wall that further research should focus on. From this analysis it was determined that the Brachypodium lines containing allele AA appeared to have a lower degree of cellulose crystallinity compared with allele BB. This decrease in crystallinity was further supported by work conducted analysing the cellulose crystallinity of AIR samples, however this was not seen in the XRD results therefore it is unclear if the difference in allele is responsible for a change in cellulose crystallinity though it has been reported that Arabidopsis *irx* mutants contain less cellulose (Brown *et al* 2007).

The composition of the cell walls of the Brachypodium lines were studied further and it was determined that there was no significant difference in the amount of lignin present between the two alleles. It was also discovered that there was also was no significant difference in the monosaccharide composition between the alleles in terms of neutral sugars found within the cell wall. This result is different to what has been previously published by Brown *et al* 2007 who reported that Arabidopsis *irx9, 10* and *14* mutants contained less xylose and in terms of grasses it has been reported by Chen *et al* 2013 that *irx10* mutants contained a loss in xylose content and an increase in arabinose.

Finally, the hemicellulose analysis was undertaken to determine if there is a difference in xylan chain length when comparing the Brachypodium alleles as it has been suggested that IRX9, 10 and 14 are all involved in the biosynthesis of the xylan

backbone (Brown *et al* 2007). This study it was noted that there was no difference in chain length between alleles AA and BB for the Bradi5g24290.1 gene. Chen *et al* 2013 also reported that they didn't see a change in the size of xylan when analysing *irx10* rice mutants. Further work in future could be conducted looking at the proportion of GlcUA and Me-GlcUA side chains as these have been shown to be affected in *irx* mutants of Arabidopsis (Brown *et al* 2007).

Phylogenetic trees of the two Brachypodium candidate genes, Bradi5g24290.1 and Bradi3g06740, were constructed using protein sequences from Arabidopsis, Brachypodium and rice. The trees confirmed that both genes grouped together with the expected orthologs, namely AtIRX14 and OsIRX14 for the BdGT43 gene and AtCslA3, 7 and 9 as well as OsCslA1 and 9 for the BdCslA3 gene. This information was used to select Arabidopsis T-DNA lines to undergo cell wall composition analysis so that the results could be compared with those that were obtained for the Brachypodium lines.

The saccharification analysis of the Arabidopsis T-DNA lines indicated that they all had a lower saccharification potential than the wild type, with GT43 being significantly less digestible. This result for GT43 was not expected as it has been reported that *irx* mutants have shown an increase in digestibility (Chen *et al* 2013). This change in digestibility for the Arabidopsis GT43 line is also different from what was observed during the analysis of the Brachypodium lines, which indicated an improvement in digestibility.

During the ATR-FTIR analysis of the Arabidopsis T-DNA lines it was observed that GT43, CslA2.2, CslA9 and CslA11.2 all appeared to have a reduction in cellulose crystallinity, which was confirmed during cellulose crystallinity analysis of AIR samples though again not by the XRD results. The cellulose crystallinity results are the same as those obtained for the Brachypodium allele AA lines.

Even though no difference in monosaccharide composition was detected during the analysis of the Brachypodium alleles there was a change observed in the Arabidopsis T-DNA lines when compared to the wild type. The GT43 line appeared to have a significant decrease in xylose and a small but significant increase in mannose. The

result for GT43 was expected as it has been reported in the literature that Arabidopsis *irx* mutants have a decrease in xylose (Brown *et al* 2007) and in rice this decrease in xylose is modest (Chen *et al* 2013). The CslA lines revealed a significant decrease in mannose for CslA9 and a significant increase in glucose for CslA2.1. The difference in the mannose content is expected for CslA genes as it has been reported previously and it has been shown that CslA9 is responsible for the majority of clucomannan synthesis in the cell wall of stems (Scheller *et al* 2010, Goubert *et al* 2009).

Finally, the amount of lignin was analysed and it was determined that there was a significant increase in the amount of lignin present in the Arabidopsis CslA2.1 T-DNA line when compared to the wild type but this change in lignin content was not seen between the Brachypodium alleles.

In conclusion this work has given some idea as to effect of the allele possibly has on cell wall biosynthesis but more work needs to be done because the differences seen with the Brachypodium lines only showed subtle changes in terms of cellulose crystallinity. In future, Brachypodium T-DNA lines could be produced for the relevant genes and undergo the same experiments but also include others to specifically looking at hemicellulose that was only touch upon in this instance. This could include looking at xylose and arabinose content of hemicellulose fractions as well as ferulic acid and *p*-coumeric acid composition.

Chapter 6: Discussion.

The burning of fossil fuels for energy is no longer sustainable or environmentally suitable as it has been stated that the transport sector in OECD countries are responsible for at least 23% of all CO_2 emissions globally, therefore other sources of energy are needed. In terms of transportation fuels, one of the most likely substitutes is biofuels either as biodiesel or bioethanol. Currently the transportation industry uses 3.8 trillion litres of fossil fuels globally and it is expected that biofuels will account for 27% of all transportation fuels by 2050 (Morales *et al* 2015, Parajuli *et al* 2015). However, because of the need for this technology to be sustainable and not impact on global food supplies; second generation biofuels generated from lignocellulosic biomass are a more viable source than first generation biofuels produced directly from food crops.

Fossil fuels are not sustainable as they rely on limited natural resources. New technology is always being developed to obtain harder to reach oil reserves such as shale oil, which leads to a drop in fuel prices, but this technology only offers a short term solution (Cheali *et al* 2015). Lignocellulosic biofuels relay instead on naturally according waste streams such as agricultural waste products and solid municipal waste. Therefore, they not only provide the needed transportation fuel but also solve the problem of waste removal. This can in practice have a big impact on human health. For example the practice of burning rice straw in the field to remove the waste leads to large amounts of pollution and CO_2 being generated at the same time, which has a knock on effect on health (Gadde *et al* 2009). If the straw was collected instead for biofuel production it would lead to improvements of air quality and reduce the incidence of poor lung health in these areas and it could also supply the farmer directly with another income stream.

Currently, the thinking is to develop lignocellulosic fuels using an integrated biorefinery model where biorefinery has been defined as the sustainable processing of biomass into a spectrum of marketable products and energy (Cheali *et al* 2015, de Jong *et al* 2015, Parajuli *et al* 2015). This refinery model is therefore based on what is used by the petrochemical industry in terms of producing fuels as well as various

high value petrochemicals that are used in a number of industries. A typical petrochemical refinery production consists of 70% transportation fuels and 30% chemicals, plastics and other products, however when comparing it to the revenue received, transportation fuels only account for 40 – 45% of the income created (Earhart *et al* 2015). It is hoped that the integrated biorefinery could produce many of the petrochemicals as well as other novel plant based chemicals together with other products such as animal feed, polymers, pharmaceuticals, materials and energy for power and heating (Falano *et al* 2014, Cheali *et al* 2015, de Jong *et al* 2015). In the past, focus has been in the production of the biofuels as the main product of biorefineries but this is no longer necessarily the main focus due to changes in economics due to a decrease in fossil fuel costs. The production of bioethanol from straw accounts for only 14 – 23% of the biomass input, therefore coproduction of other products can add value to the process as well as adding protection against fluctuations in market prices (Parajuli *et al* 2015). Bioethanol can also be used instead as an intermediate feedstock for the production of a large number of chemicals that have a higher value (Cheali *et al* 2015). In 2005, biobased chemicals accounted for 7% of all sales, which is equivalent to $77 billion. The EU contributed to approximately 30% of these sales. The global market for biobased products is growing at a rate of 10 – 20% annually and is expected to be worth $250 billion by 2020 (de Jong *et al* 2015, Parajuli *et al* 2015).

A number of studies have been conducted to determine the economic and sustainability of the biorefinery approach (Morales *et al* 2015). This is a complex area of research as a number of variables have to be considered such as the production of the feedstock, the type of process used to breakdown the lignocellulose to its sugars as well as the products being produced. All these factors can influence the effect the process will have on the environment in terms of GHG emissions as well as how profitable the business will be. A number of methods have been developed to study these processes which include cradle-to-grave assessments using life cycle analysis (LCA) as well as financial studies (Cheali *et al* 2015, Morales *et al* 2015, McKechnie *et al* 2015). These studies have shown that second generation biofuels have lower GHG emissions that first generation biofuels as well as fossil fuels however the amount of bioethanol present in the final product of blended fuels affects how large that reduction really is. For example, E10 fuels only have less than

a 10% reduction in GHG emissions compared with fossil fuels (Falano *et al* 2014, Jeswani *et al* 2015, Morales *et al* 2015). The feedstock used in the biorefinery also has an impact on GHG emissions; therefore agricultural waste has the lowest GHG emissions due to the amount of fertiliser used in the production of the biomass (Morales *et al* 2015, Muylle *et al* 2015, Parajuli *et al* 2015). In most cases lignocellulosic biofuel production is also energetically sustainable because lignin is burnt to provide heating for the production of bioethanol. However, this may not always be the case if lignin can be valorised instead of burning (Morales *et al* 2015, Parajuli *et al* 2015). For example, the value of lignin products can range in value from 200 $/tonne for lignin fertilizers to 8000 $/tonne for lignin-carbon fibres or vanillin (Eerhart *et al* 2015).

There are a number of barriers that need to be overcome before the commercialisation of an integrated biorefinery can be successful. These factors cover all areas of optimising the process conditions to release the maximum amount of sugars from the feedstock at the lowest cost (Morales *et al* 2015). It has been suggested that the purchasing of the feedstock can account for 40 – 60% of the total operating costs, which is equal to 30 – 32% of the total production cost of a gallon of ethanol produced. The pretreatment method selected can also have a large impact on the production cost as it can account for 19 – 20% of the final cost of a gallon of bioethanol. It is known that the cost of cellulase enzymes play a role in influencing the profitability of bioethanol in comparison to fossil fuels, therefore companies such as Novazymes Biotech are continually making improvements in technology and have recent reported a development that could lead to a decrease in the cost of cellulase from $5.40 per gallon of ethanol to 20 cents/gal (Parajuli *et al* 2015). Even with these financial considerations, US biorefineries are being set up to produce ethanol from agricultural residues. It is expected that they will produce bioethanol that costs approximately 0.50 – 1 $/gallon (Eerhart *et al* 2015).

Due to the cost of feedstocks having a large impact on final production costs it needs to be available at a moderate cost as well as be available in large quantities to meet the growing demands from a biobased economy (de Jong *et al* 2015, Muylle *et al* 2015, Parajuli *et al* 2015). Currently, approximately 5.1 billion dry tonnes of agricultural residues are produced globally every year. In the UK every year 27

million tonnes of biomass could be available, of which 80% consists of forestry and agricultural waste. A further 10% could come from energy crops (Falano *et al* 2014, Jeswani *et al* 2015). However, to keep up with demand and to improve biomass conversation to desired products, whether it is biofuels or other biobase products, developing new varieties will be necessary (de Jong *et al* 2015). For example, developing crops that have a high dry mater yield under lower input conditions such as varieties that are efficient nitrogen users (Muyelle *et al* 2015) or those that have changes in the cell wall structure that allow for improved digestibility (Gomez *et al* 2008).

Research needs to be conducted in order to understand cell wall biosynthesis and composition before new varieties of crops can be developed that have improved digestibility for example. To undertake this research a model plant is needed as agriculturally important crops tend to be difficult to use due to large and complicated genomes and lack of developed genetic tools (Bevan *et al* 2010, Christensen *et al* 2010). In this study Brachypodium was selected as the model plant for a number of reasons as mentioned throughout this thesis but it was selected predominantly because it is has a cell wall that is very similar in structure to other agricultural crops, unlike Arabidopsis (Garvin *et al* 2008, Bevan *et al* 2010). Brachypodium has been used successfully as a model plant in a number of studies related to different areas of research such as seed development, plant-pathogen interactions and root systems. In terms of cell wall research it has been used to study lignin composition as well as polysaccharide composition (Girin *et al* 2014). In this study, we have successfully used Brachypodium to determine the presence of variation in terms of cell wall composition and digestibility within natural varieties. The information gathered from the natural Brachypodium varieties can be used to develop new populations for further cell wall research as was demonstrated with the initial development of the Bd21 x BdTr1-1f RIL population, which shows a variation in digestibility ranging from 19.39 to 44.02 nmol reducing sugars release/mg material.h^{-1} when an alkaline pretreatment is used. This study was also able to successfully use an existing RIL population, Bd21 x Bd3-1, to identify possible genes involved in cell wall biosynthesis that could influence saccharification.

A forward genetic approach was used to identify genes linked to cell wall digestibility. The methodology followed was QTL analysis of the Brachypodium Bd21 x Bd3-1 RIL population. This population has been used successfully for QTL analysis in the identification of the barley stripe mosaic virus resistant gene *bsr1* (Cui *et al* 2012). QTL analysis is a good method to use for the studying of complex traits that are influenced by the environment. However, this study confirmed that the design of the experiment is very important as too much environmental variation can led to no QTLs being detected as was seen in the initial analysis of samples grown in Versailles, France. It was found that these non-genetic factors where probably being introduced during growth and sampling as well as in the laboratory due to systematic variation found on the semi-automatic liquid handling robot. Oakey *et al* (2013) reported that laboratory variation accounted for the largest cause of environmental variation found in an experiment analysing the saccharification potential of elite barley cultivars. Therefore, the environmental factors must be accounted for and removed through introducing replication and randomisation into experimental design (Borevitz *et al* 2004, Oakey *et al* 2013). Another method for decreasing environmental variation is to include corrections to the phenotypic data before QTL analysis takes place. Murray *et al* (2005) reported including corrections to their data led to improvements in trait normality, heritability, LOD scores as well as causing the marker LOD interval to narrow.

In this study a single QTL was detected on chromosome 5 linked to marker BD1676_1. It was determined that this digestibility trait had a strong genetic component as it had a broad-sense heritability of 0.45 and the QTL on chromosome accounted for 11.83% of the genetic variation of the phenotype. The calculated heritability ratio is similar to that observed by Ingram *et al* (2012), when working with the same population while analysing the heritability of various root architecture traits. Further analysis of the QTL resulted in the detection of another QTL on chromosome 3 linked to marker BD1415_1, which acted in epistasis. It is believed to be good practice to include epistasis analysis otherwise biased estimates may result. However, it is not always easy to conduct this type of analysis because genetic and environmental interactions may cause limitations (Shiringani *et al* 2010).

In total, only two QTLs could be identified in this project however in other studies involving maize and sorghum for example they detected many more. This may be due to the small amount of variation in the trait being investigate, evident in between the two parental lines used to produce the RIL population. For the sorghum research they were able to identify a large number of QTLs by analysing many different traits associated to digestibility, such as height, stalk mass and hemicellulose to name but a few. In total, Murray *et al* (2008) identified 110 QTL for 31 traits and Shiringani *et al* (2011) found 72 QTL for 9 traits. In the case of the maize studies, Krokowsky *et al* (2005) analysed their RIL population in two different locations over two years and Cardinal *et al* (2003) analysed different tissue types for various cell wall components. In both cases they were able to detect a number of QTL distributed throughout the genome. Another way to increase the detection rate of QTL is to increase the size of the population analysed. It is known that the larger the population the easier it is to detect QTLs that have a low effect on the phenotype (Parker *et al* 1998, Xu *et al* 2005). It may also be possible to increase the number of QTLs detected by using a population that has been developed from more divergent parents. In the case of the Bd21 x Bd3-1 population the parental lines had a small difference in their digestibility potential. Using the Bd21 x BdTr1-1f RIL population that is currently in development may increase the number of QTLs detected as the parental lines were selected according to their differing saccharification results. However, this study has shown that using a population that has not been specifically developed for the analysis of the trait of interest can still identify possible QTLs successfully.

Any QTLs identified during analysis need to be confirmed. Usually, this is rarely done within the same project as it normally involves additional costs due to growing the same population either in additional locations or over numerous years to determine if the same QTLs are detected. In this project a single-marker analysis approach was taken, which relies on trying to link a single marker to the trait of interest. This is the simplest method for detecting QTL and uses statistical techniques such as t-test, ANOVA or linear regression (Collard *et al* 2005). Using this method the presence of the QTLs linked to markers BD1676_1 and BD1415_1 were confirmed.

In this study the regions around the identified QTLs on chromosome 3 and 5 were analysed and the candidate genes Bradi3g06740 and Bradi5g24290.1 were selected respectively. Bradi3g06740 is predicted to be a Csl gene belonging to subfamily A, which has been classified as having a possible glucomannan 4-beta-mannosyltransferase function during cell wall biosynthesis according to the Pfam database and is involved in hemicellulose biosynthesis (Scheller *et al* 2010). Bradi5g24290.1 is predicted to belong to the glycosyltransferase family 43 and has a beta-1.3-glucuronyltransferase function, which is involved in the backbone synthesis of xylan found in hemicellulose (Carpita *et al* 2012, Dhugga 2012).

Ideally the candidate genes would have been studied further using Brachypodium T-DNA lines because using Arabidopsis T-DNA lines as a model to confirm gene function has it limitations. This is especially true when studying cell wall genes as the composition of Arabidopsis cell walls are different to those found in grasses (Vogel 2008, Bevan *et al* 2010). Unfortunately, at the time of this research the development of Brachypodium T-DNA lines was still in its infancy in comparison to Arabidopsis and rice T-DNA libraries that contained approximately 700 000 available lines. In 2008, the first 465 Brachypodium T-DNA lines were available for distribution but by 2010 this had increased to more than 1000 lines (Thole *et al* 2010). Today there are 23 000 lines available from the Join Genome Institute (JGI) as well as 13 000 lines from the BrachyTag programme and the USDA-ARS Western Regional Research Centre (An *et al* 2016). Therefore, at the time of conducting this research there were no Brachypodium T-DNA lines available for the candidate genes of interest however there are now two available for Bradi3g06740 and one for Bradi5g24290.

Due to this limitation it was decided to select Arabidopsis T-DNA lines for further analysis as well as compare the Brachypodium lines containing differing genotypes at the QTL linked markers. This method was able to give some insight in to the changes in cell wall composition for example it was noted the AtCslA9 T-DNA line contained a reduced amount of mannose with in the cell wall. This has been reported before by Goubert *et al* (2009) when they studied various AtCslA mutants. They determined that AtCslA9 was a significant contributor to mannan synthesis within stems as the mutant line showed a decrease of mannose by approximately 81% when

compared to wild type, whereas mutants of the other genes studied showed no detectable phenotype. The authors went on to show that AtCslA2 and AtCslA3 play a minor but an important role in glucomannan synthesis within the plant cell wall (Scheller *et al* 2010, Liepman *et al* 2012).

When the AtGT43 T-DNA line was analysed it showed the expected decrease in xylose. Brown *et al* (2007) reported that Arabidopsis T-DNA mutants involved in xylan biosynthesis contained a decrease in xylose content of the stem and more specifically that the *irx14* mutant appeared developmentally normal but also contained a less severe decrease in xylose. During composition analysis the Brachypodium allelic lines were also analysed to determine if there was a difference in the length of the xylan backbone as it has been reported that the *irx9, 10* and *14* mutants in Arabidopsis had a shorter chain length (Lovegrove *et al* 2013). However, it appears that this was not the case in the Brachypodium studied in this project. In Lovegrove *et al* (2013) paper they also mentioned that the side chains showed an increase in arabinose substitutions for the wheat transgenic lines that they studied but not in the Arabidopsis lines. The wheat lines also revealed a loss of arabinoxylan chains but again this was not present in the Arabidopsis lines. Therefore, future work could look at whether there is a difference in xylan backbone length in either Arabidopsis or Brachypodium T-DNA lines as well as including work on side chain structures.

It was also noted during cell wall composition analysis that the AtCslA2, AtCslA9 and the AtGT43 T-DNA lines have a decrease in cellulose crystallinity as well as an increase in lignin content. These differences were also observed in the Brachypodium allelic lines. Neither of these observations has been mentioned before in relation to changes in expression of the CslA genes. However, it is known that glucomannan, together with xylan, bind tightly to cellulose to form crosslinks between the cellulose microfibrils as well as links to lignin (Goubert *et al* 2009). It is therefore, possible that a change in glucomannan concentration or composition could affect both cellulose crystallinity and lignin content. This relationship would need to be studied further to determine how this affect is introduced. It has however, been reported that the AtGT43 T-DNA line has a decrease in cellulose content, which could be directly or indirectly due to alterations in xylan biosynthesis that have led to

changes in to the structure of the plant (Brown *et al* 2007). In the case of lignin, Petersen *et al* (2012) reported that Arabidopsis transformants of *irx7, 8* and *9* contained a decrease possibly due to a decrease in glucuronic acid, which could potentially affect the degree of lignification if it is dependent on xylan biosynthesis.

Modification of the mannan or glucomannan content of plant cell walls may be a good target for improved biofuel production because it can also result in changes to the hexose/pentose ratio, thereby increasing the hexose content and providing more readily available sugars for fermentation (Goubert *et al* 2009). The mannan and glucomannan content of grasses has been reported as relatively low as it contributes less than 5% towards the hemicellulose composition of the secondary cell wall (Scheller *et al* 2010, Dhugga 2012). Xylan modification would also make a good target for improved biomass as it has been suggested that a 20% decrease in xylose together with a 10% decrease in lignin could result in a 10 − 15% decrease in the selling price of bioethanol (Petersen *et al* 2012).

Therefore in conclusion this study has been able to determine that there is a variation in cell wall digestibility amongst natural Brachypodium accessions and this variation can be used in the development of specific populations for further cell wall studies. The study has also shown that QTLs linked to digestibility are present in a Brachypodium RIL and they can be used to identify candidate genes that are involved in cell wall biosynthesis. These candidates can then be studied further in Arabidopsis T-DNA lines as well as to some degree in Brachypodium allelic variants. The analysis of the cell wall of these plants has indicated that Bradi5g24290.1 and Bradi3g06740 possibly play a role in lignocellulose recalcitrance.

Appendices

Appendix A

Brachypodium RIL plant lines harvested from Block D and E at INRA, Versailles, France.

Plant Lines Collected	
Block D	**Block E**
L001	L002
L002	L003
L003	L005
L004	L006
L005	L007
L007	L008
L008	L009
L009	L010
L010	L012
L012	L013
L013	L015
L015	L018
L016	L019
L017	L020
L018	L021
L019	L022
L020	L023
L021	L024
L022	L025
L023	L027
L024	L029
L027	L031
L028	L032
L029	L033
L031	L034
L032	L035
L033	L036
L034	L037
L035	L038
L036	L042
L037	L044
L038	L045
L042	L047
L044	L049

L045	L050
L047	L051
L051	L052
L052	L053
L053	L054
L054	L055
L055	L056
L056	L057
L057	L058
L058	L059
L059	L060
L060	L061
L061	L062
L062	L063
L063	L064
L064	L065
L065	L066
L066	L068
L068	L069
L069	L070
L070	L071
L071	L073
L072	L074
L073	L076
L074	L077
L076	L078
L077	L079
L078	L080
L079	L081
L080	L082
L081	L083
L082	L085
L083	L086
L085	L087
L086	L088
L087	L089
L088	L091
L089	L096
L090	L097
L091	L098
L092	L099
L094	L101
L095	L102
L096	L103
L097	L104

L098	L106
L099	L107
L100	L109
L101	L111
L102	L112
L103	L113
L104	L114
L106	L115
L107	L117
L108	L120
L109	L121
L111	L122
L112	L124
L113	L125
L114	L127
L115	L128
L116	L129
L117	L130
L118	L131
L119	L136
L120	L137
L121	L138
L122	L139
L123	L140
L124	L141
L125	L142
L127	L143
L128	L144
L129	L145
L130	L146
L131	L149
L132	L150
L136	L153
L137	L154
L138	L155
L139	L157
L140	L158
L141	L159
L142	L161
L143	L162
L144	L163
L145	L164
L146	L165
L147	L170
L149	L174

L150	L175
L153	L176
L154	L183
L155	
L157	
L158	
L159	
L161	
L162	
L163	
L164	
L165	
L166	
L168	
L169	
L170	
L171	
L172	
L174	
L175	
L176	
L183	

Appendix B

Brachypodium RIL plant lines harvested from Block 1 – 3 at the University of York, UK.

Plant Lines Collected		
Block 1	**Block 2**	**Block 3**
L001	L001	L001
L002	L002	L003
L003	L003	L004
L004	L004	L005
L005	L005	L007
L007	L007	L008
L008	L008	L009
L009	L009	L010
L010	L010	L012
L012	L012	L013
L013	L013	L015
L015	L015	L016
L016	L016	L017
L017	L017	L018
L018	L018	L019
L019	L019	L020
L020	L020	L021
L021	L021	L022
L022	L022	L023
L023	L023	L024
L024	L024	L027
L027	L027	L028
L028	L028	L029
L029	L029	L031
L031	L031	L032
L033	L033	L033
L034	L034	L034
L035	L035	L035
L037	L037	L037
L038	L038	L038
L039	L039	L039
L042	L042	L042
L045	L045	L045
L047	L047	L047
L051	L051	L051
L052	L052	L052
L053	L053	L053

L054	L054	L054
L055	L055	L055
L056	L056	L057
L057	L057	L058
L058	L058	L059
L059	L059	L060
L060	L060	L061
L061	L061	L062
L062	L062	L063
L063	L063	L064
L064	L064	L065
L065	L065	L066
L066	L066	L068
L068	L068	L069
L069	L069	L070
L070	L070	L071
L071	L071	L072
L072	L072	L073
L073	L073	L076
L074	L074	L077
L076	L076	L078
L077	L077	L079
L078	L078	L080
L079	L079	L081
L080	L080	L082
L081	L081	L083
L082	L082	L085
L083	L083	L086
L085	L085	L087
L086	L086	L088
L087	L087	L089
L088	L088	L090
L089	L089	L091
L090	L090	L092
L091	L091	L094
L092	L092	L095
L094	L094	L096
L095	L095	L097
L096	L096	L098
L097	L097	L099
L098	L098	L100
L099	L099	L101
L100	L100	L102
L101	L101	L103
L102	L102	L104

172

L103	L103	L106
L104	L104	L107
L106	L106	L108
L107	L107	L109
L108	L108	L111
L109	L109	L112
L111	L111	L113
L112	L112	L114
L113	L113	L115
L114	L114	L116
L115	L115	L117
L116	L116	L118
L117	L117	L119
L118	L118	L120
L119	L119	L121
L120	L120	L122
L121	L121	L123
L122	L122	L124
L123	L123	L125
L124	L124	L127
L125	L125	L128
L127	L127	L129
L128	L128	L130
L129	L129	L131
L130	L130	L132
L131	L131	L136
L132	L132	L137
L136	L136	L138
L137	L137	L139
L138	L138	L140
L139	L139	L141
L140	L140	L142
L141	L141	L143
L142	L142	L144
L143	L143	L145
L144	L144	L146
L145	L145	L147
L146	L146	L149
L147	L147	L150
L149	L149	L153
L150	L150	L154
L153	L153	L155
L154	L154	L157
L155	L155	L158
L157	L157	L159

L158	L158	L161
L159	L159	L162
L161	L161	L163
L162	L162	L164
L163	L163	L165
L164	L164	L166
L165	L165	L168
L166	L166	L169
L168	L168	L170
L169	L169	L171
L170	L170	L172
L171	L171	L174
L172	L172	L175
L174	L174	L176
L175	L175	L183
L176	L176	
L183	L183	

Appendix C

Rank order of Brachypodium RIL plant lines from Block 1 – 3 according to digestibility potential based on saccharification data.

Sample ID	Rank Order		
	Block 1	Block 2	Block 2
L001	75	120	107
L003	68	131	9
L004	123	86	87
L005	26	13	5
L007	14	106	104
L008	12	102	127
L009	9	119	98
L010	5	99	56
L012	111	105	7
L013	110	2	11
L015	7	42	2
L016	22	95	112
L017	112	51	1
L018	44	16	3
L019	85	71	122
L020	63	41	96
L021	126	126	108
L022	43	111	63
L023	88	10	101
L024	106	134	138
L027	57	68	39
L028	4	26	25
L029	62	52	38
L031	139	137	81
L033	20	70	89
L034	55	136	125
L035	60	78	102
L037	34	15	27
L038	95	24	99
L039	13	8	137
L042	67	73	93
L047	50	33	37
L051	97	32	71
L052	69	104	86
L053	135	62	109
L054	18	9	30
L055	15	45	76

L057	120	54	124
L058	33	121	129
L059	104	61	22
L060	66	103	132
L061	37	75	79
L062	138	19	46
L063	61	11	32
L064	94	50	65
L065	79	34	26
L066	10	65	73
L068	125	36	55
L069	74	116	68
L070	124	55	140
L071	87	98	19
L072	103	77	114
L073	78	58	41
L076	119	100	83
L077	89	94	20
L078	52	74	31
L079	134	80	4
L080	56	108	118
L081	17	83	97
L082	82	130	136
L083	93	133	53
L085	64	44	59
L086	65	79	139
L087	28	66	82
L088	127	125	128
L089	100	38	14
L090	48	81	121
L091	140	84	45
L092	128	72	50
L094	71	53	110
L095	91	57	113
L096	51	22	85
L097	132	69	57
L098	29	128	72
L099	105	115	134
L100	32	76	58
L101	86	135	100
L102	27	59	95
L103	131	20	43
L104	8	113	70
L106	83	1	28
L107	81	101	47

L108	136	7	61
L109	24	5	16
L111	137	37	80
L112	58	23	8
L113	109	82	64
L114	16	64	35
L115	35	56	12
L116	72	88	49
L117	90	21	135
L118	101	25	126
L119	80	132	115
L120	1	117	54
L121	130	140	106
L122	39	107	69
L123	59	97	42
L124	73	138	131
L125	42	67	33
L127	30	139	111
L128	117	93	48
L129	92	18	90
L130	31	43	60
L131	23	14	67
L132	49	63	18
L136	113	90	75
L137	3	40	94
L138	96	89	62
L139	70	123	116
L140	40	129	120
L141	53	49	36
L142	99	92	130
L143	133	27	34
L144	98	127	117
L145	121	85	77
L146	2	112	91
L147	116	109	103
L149	41	124	84
L150	122	122	133
L153	25	12	13
L154	38	48	105
L155	108	91	66
L157	118	39	6
L158	36	17	17
L159	21	47	29
L161	54	29	119
L162	77	87	78

L163	115	31	40
L164	47	35	74
L165	114	96	51
L166	84	28	92
L168	129	6	21
L169	46	3	44
L170	107	46	15
L171	76	114	88
L172	102	118	123
L174	6	30	52
L175	11	110	23
L176	19	60	10
L183	45	4	24

Appendix D

Genes identified on chromosome 3 around the epistatic QTL linked to marker BD1415_1 (position shown highlighted in red).

Gene Name		Function	Pfam	Panther
Bradi3g0	6540.1	Putative Gene	Mediator complex protein	
	6550.1	Putative Gene	GDSL-like Lipase/Acylhydrolase	Zinc finger FYVE domain containing protein
	6556.1	Putative Gene	Cleavage site for pathogenic type III effector avirulence factor Avr	
	6562.1	Putative Gene	AP2 domain	
	6570.1	Acetyltransferase activity		
	6577.1	Putative Gene	ABC transporter	
	6590.1	Leucine-rich repeat protein kinase, putative, expressed, subfamily LRR-V		
	6597.1	Putative Gene	Lycopene cyclase protein	Flavoprotein-ubiquinone oxidoreductase
	6610.1	Putative Gene		
	6620.1	Putative Gene	Prolyl oligopeptidase family	
	6630.1	Gamma-tubulin binding	Biogenesis of lysosome-related organelles complex-1 subunit 2	Phytoene dehydrogenase
	6640.1	Putative Gene	Transmembrane amino acid transporter protein	Amino acid transporter
	6640.3	Hyrogen:amino acid symporter activity	Transmembrane amino acid transporter protein	
	6650.1	Putative Gene		
	6660.1	RING, subfamily zinc finger (C3HC4-type RING finger) family protein	Zinc finger, C3HC4 type (RING finger)	Ring Zinc finger protein

	6670.1	bZIP transcription factor	bZIP transcription factor	CAMP-response element binding protein-related
	6680.2	Putative Gene	Spindle pole body ineracting protein	Phosphatase 2A regulatory subunit-related
	6680.3	Putative Gene	Spindle pole body ineracting protein	Phosphatase 2A regulatory subunit-related
	6690.1	Putative Gene		
	6695.1	Putative Gene		
	6700.1	Ligand-dependent nuclear receptor transcription coactivator activity		
	6700.2	Ligand-dependent nuclear receptor transcription coactivator activity		
	6710.1	Zinc ion binding		
	6720.1	Acetyltransferase activity	PHD-finger	
	6727.1	Putative Gene	PHD-finger	Nuclear protein 95
	6740.1	Cellulose synthase-like (CSL), subfamily A	Glycosyl transferase family 2	N-acetylglucosaminyltransferse-related
	6740.2	Cellulose synthase-like (CSL), subfamily A	Glycosyl transferase family 2	N-acetylglucosaminyltransferse-related
	6750.1	Thioredoxin peroxidase activity	Redoxin	Peroxiredoxin-5
	6757.1	Putative Gene	Protein kinase domain	
	6770.1	Protein kinase family protein, putative, expressed, subfamily LysM-II		
	6780.1	Putative Gene	Cytochrome P450	
	6790.1	Putative Gene		
	6800.1	Putative Gene		
	6810.1	Putative Gene	PLATZ transcription factor	
	6820.1	Putative Gene	PLATZ transcription factor	

180

	6830.1	Putative Gene		
	6840.1	Protein kinase activity	Mad3/BUB1 homology region 1	Mitotic checkpoint Serine/Threonine-protein kinase BUB1 and BUBR1
	6840.2	Protein kinase activity	Mad3/BUB1 homology region 1	Mitotic checkpoint Serine/Threonine-protein kinase BUB1 and BUBR1
	6847.1	Putative Gene	DUF1981	Guanyl-nucleotide exchange factor
	6860.1	Putative Gene		
	6870.1	MYB-related transcription factor	Myb-like DNA-binding domain	SWI/SNF complex-related
	6870.2	MYB-related transcription factor	Myb-like DNA-binding domain	SWI/SNF complex-related
	6880.1	Putative Gene	YIF1	YIP1 interacting factor homolog
	6890.1	Putative Gene		
	6900.1	Putative Gene	Elongation factor TS	Elongation factor TS
	6910.1	CYCLIN, subfamily CYCD7	Cyclin	Cyclins
	6916.1	Putative Gene	F-box domain	
	6922.1	Putative Gene	F-box domain	
	6930.1	Citrate (Si)-synthase activity, 2-methylcitrate synthase activity	Citrate synthase	Citrate synthase
	6940.1	KH domain-containing protein similar to KH domain-containing protein/Zinc finger (CCCH type) family protein	KH domain	KH RNA binding domain protein
	6950.1	KH domain-containing protein similar to zinc finger C-x8-C-x3-H type protein, [Orysa sativa] LOC_Os0210080	KH domain	RNA-binding protein related
	6957.1	Putative Gene	KH domain	RNA-binding protein related
	6970.1	Putative Gene	Nse4	
	6980.1	Leucine-rich repeat protein kinase, putative, subfamily LRR-Xb	Leucine rich repeat	

6990.1	Leucine-rich repeat protein kinase, putative, subfamily LRR-V		
7000.2	Iron ion binding, lipoxygenase activity	Lipoxygenase	Lipoxygenase
7010.1	Iron ion binding, lipoxygenase activity	Lipoxygenase	Lipoxygenase
7020.1	Putative Gene	Tubulin binding cofactor C	
7020.2	Putative Gene	Tubulin binding cofactor C	
7030.1	bZIP transcription factor	bZIP transcription factor	CAMP-response element binding protein-related
7040.1	Putative Gene	DUF2048	
7040.2	Putative Gene	DUF2048	
7047.1	Putative Gene	Putative glycosyl hydrolase of unknown function (DUF1680)	
7060.1	Zinc ion binding	Zinc finger	PRK1-associated Zinc finger protein
7060.2	Zinc ion binding	Zinc finger	PRK1-associated Zinc finger protein
7070.1	Putative Gene	DnaJ domain	
7070.3	Putative Gene	DnaJ domain	
7070.4	Putative Gene	DnaJ domain	
7080.2	Zinc ion transmembrane transporter activity	ZIP Zinc transporter	GUFA protein-related
7090.2	Putative Gene	Surface antigen	Sorting and assembly machinery (SAM50) protein
7090.3	Putative Gene	Surface antigen	Sorting and assembly machinery (SAM50) protein
7100.1	Putative Gene		
7110.1	Heavy metal P-type ATPase subfamily P1 cluster 3 from PMID:12805592. Similar to AtHMA 5 Cu transporter	Heavy-metal-associated domain	
7120.1	Glycosyl hydrolase (GH), subfamily GH28	Glycosyl hydrolases family 28	

	7130.1	Fumarylacetoacetase activity	Fumarylacetoacetate (FAA) hydrolase family	Fumarylacetoacetate hydrolase
	7140.1	3-oxoacyl-[acyl-carrier-protein] synthase activity	Beta-ketoacyl synthase	Polyketide synthase-related
	7150.1	Putative Gene	Mlo family	
	7160.1	GRAS transcription factor	GRAS family transcription factor	
	7170.1	Putative Gene		
	7180.1	Inositol or phosphatidylinositol phosphatase activity	Regulator of chromosome condensation (RCC1) repeat	Regulator of chromosome condensation
	7190.1	Putative Gene	Chlorophyll A-B binding protein	Chlorophyll A/B binding protein
	7200.1	Putative Gene		
	7210.1	Putative Gene		Signal recognition particle 68 KDA protein
	7220.1	Putative Gene	X-Pro dipeptidyl-peptidase (S15 family)	
	7220.2	Putative Gene	X-Pro dipeptidyl-peptidase (S15 family)	
	7230.1	Putative Gene		Calmodulin
	7230.2	Putative Gene		Calmodulin
	7230.3	Putative Gene		Calmodulin
	7237.1	Putative Gene		Centaurin/ARF-related
	7237.2	Putative Gene		Centaurin/ARF-related
	7237.3	Putative Gene		Centaurin/ARF-related
	7250.1	Pectin methylesterase inhibitor (PMEI)		
	7260.1	Putative Gene		
	7260.2	Putative Gene		
	7270.1	Nuclear hormone receptor binding, transcription coactivator activity	Transcription initiation factor TFIID subunit A	
	7275.1	Putative Gene		

	7280.1	Serine-type endopeptidase activity		Subtilisin/Kexin-related Serine protease
	7286.1	Putative Gene		Aspartyl protease DDI-related
	7292.1	Putative Gene	DUF868	
	7300.2	Putative Gene	Ribosomal protein L7/L12 C-terminal domain	Ribosomal protein L7/L12
	7310.1	Putative Gene		
	7315.1	Putative Gene	NB-ARC domain	
	7320.1	Putative Gene	FKBP-type peptidyl-prolyl cis-trans isomerase	FK506 binding protein
	7320.2	Putative Gene	FKBP-type peptidyl-prolyl cis-trans isomerase	FK506 binding protein
	7320.3	Putative Gene	FKBP-type peptidyl-prolyl cis-trans isomerase	FK506 binding protein
	7330.1	F-Box		
	7340.1	F-Box		
	7350.1	Putative Gene	PPR repeat	

Appendix E

Genes identified on chromosome 5 around the QTL linked to marker BD1676_1 (position shown highlighted in red).

Gene Name		Function	Pfam	Panther
Bradi5g	24160.1	Motor activity, ATPase activity, Rab GTPase binding		GOLGIN-84
	24160.2	Motor activity, ATPase activity, Rab GTPase binding		GOLGIN-84
	24170.1	Secondary active sulfate transmembrane transporter activity, chloride channel activity, anion exchanger activity	Sulfate transporter family	Sulfate transporter
	24170.2	Secondary active sulfate transmembrane transporter activity, chloride channel activity, anion exchanger activity	Sulfate transporter family	Sulfate transporter
	24180.1	WAK receptor-like protein kinase, expressed, subfamily WAKb	Protein kinase domain	
	24190.1	WAK receptor-like protein kinase, expressed, subfamily WAKb	Protein kinase domain	
	24200.1	Putative Gene	Peroxidase	
	24207.1	Putative Gene	Alpha-N-acetylglucosaminidase (NAGLU)	Alpha-N-acetylglucosaminidase
	24220.1	Transcription corepressor activity, phosphoglycerate dehydrogenase activity		2-Hydroxyacid dehydrogenase
	24227.1	Putative Gene	Asparaginase	Protease T2 asparaginase
	24240.1	Putative Gene	Sec1 family	Vesicle protein sorting-associated
	24250.1	Single-standed DNA specific 3'-5' exodeoxyribonuclease activity	Exonuclease	Exonuclease-related
	24257.1	Putative Gene	DUF2930	
	24267.1	Putative Gene	Dedicator of cytokinesis	dedicator of cytokinesis (DOCK)
	24280.1	UDP-galactosyltransferase activity	UDP-glucoronosyl and UDP-glucosyl transferase	UDP-glucuronosyltransferase related

24290.1	Putative xylosyltransferase, CAZy family GT43	Glycosyltransferase family 43	Beta-1,3-glucuronyltransferase-related
24300.1	Serine-type carboxypeptidase activity	Serine carboxypeptidase	Serine carbosypeptidase II (carboxypeptidase D)
24310.1	WAK receptor-like protein kinase, expressed, subfamily WAKb	Protein kinase domain	
24320.1	Putative Gene		
24330.1	Putative Gene		
34340.1	Putative Gene		
34350.1	Putative Gene		
24360.1	Putative Gene	AP2 domain	Protein kinase
24370.1	NADPH-hemoprotein reductase activity, iron ion binding, FAD binding, nitric-oxide synthase activity		NADPH-cytochrome P450 reductase
24380.1	Putative Gene	AUX/IAA family	
24387.1	Putative Gene	Sodium/calcium exchanger protein	
24397.1	Putative Gene	Transmembrane proteins 14C	
24410.1	Putative Gene	Tify domain	
24410.2	Putative Gene	Tify domain	
24420.1	Aminoacyl-tRNA hydrolase activity	Peptidyl-tRNA hydrolase PTH2	Peptidyl-tRNA hydrolase 2
24430.1	Endonuclease activity	S1/P1 nuclease	
24430.2	Endonuclease activity	S1/P1 nuclease	
24440.1	Putative Gene		
24450.1	Protein kinase family protein, putative, expressed, subfamily RLCK-VI	Protein tyrosine kinase	
24460.1	Putative Gene	Calmodulin binding protein-like	
24460.2	Putative Gene	Calmodulin binding protein-like	

186

	24460.3	Putative Gene	Calmodulin binding protein-like	
	24470.1	Metalloendopeptidase activity	Mov34/MPN/PAD-1 family	JUN activation domain binding protein
	24480.1	Protein kinase family protein, putative, subfamily, SD-2a	Protein kinase domain	
	24490.1	F-Box	F-Box	
	24490.2	F-Box	F-Box	
	24490.3	F-Box	F-Box	
	24500.1	Serine-type endopeptidase activity		Subtilinsin/Kexin-related serine protease
	24510.1	Putative Gene		
	24520.1	Serine-type endopeptidase activity	Peptidase inhibitor I9	Subtilinsin/Kexin-related serine protease
	24530.1	Protein serine/threonine phosphatase activity	Protein phosphatase 2C	Protein phosphatase 2c
	24536.1	Putative Gene		
	24542.1	Putative Gene	Cupin superfamily protein	MINA53 (MYC induced nuclear antigen
	24550.1	Glutamate-ammonia ligase activity, ATP binding	Glutamine synthetase	Glutamine synthetase
	24550.2	Glutamate-ammonia ligase activity, ATP binding	Glutamine synthetase	Glutamine synthetase
	24560.1	Putative Gene		
	24570.1	Protein kinase family protein, putative, subfamily, RLCK-OS1	Protein kinase domain	
	24580.1	Putative Gene	Nucleolar protein, Nop52	NNP-1 protein (novel nuclear protein 1, NOP52
	24590.1	ATP binding	ATPase family associated with various cellular activities (AAA)	
	24600.1	Putative Gene		
	24610.2	Glucose-6-phosphate 1-epimerase activity	Aldose 1-epimerase	Apospory-associated protein c-related

24610.3	Glucose-6-phosphate 1-epimerase activity	Aldose 1-epimerase	Apospory-associated protein c-related
24610.4	Glucose-6-phosphate 1-epimerase activity	Aldose 1-epimerase	Apospory-associated protein c-related
24620.1	Putative Gene		
24630.1	ATP binding	Lipase (class 3)	Alpha/beta hydrolase related
24640.1	Protein farnesyltransferase activity	Polyprenyl synthetase	Farnesyl-prophosphate synthetase
24640.2	Protein farnesyltransferase activity	Polyprenyl synthetase	Farnesyl-prophosphate synthetase
24650.1	Putative Gene	Peroxidase	
24660.1	Putative Gene	Universal stress protein family	
24670.1	Putative Gene	SBP domain	
24680.1	Putative Gene	Ribosomal protein L7Ae/L30e/S12e/Gadd45 family	60S Ribosomal protein 10A-related
24690.1	Autoinhibited H+ P-type ATPase subfamily P3 cluster 2 from PMID:12805592. Similar to AtAHA1 plasma membrane H+ transporter		
24700.1	Putative Gene	AP2 domain	
24710.1	Putative Gene	AP2 domain	
24720.1	Putative Gene	AP2 domain	
24730.1	Protein kinase activity		
24730.2	Protein kinase activity		
24737.1	Putative Gene	Protein tyrosine kinase	
24750.1	Protein kinase family protein, putative, expressed, subfamily RLCK-Os4	Protein tyrosine kinase	
24760.1	Protein kinase family protein, putative, expressed, subfamily RLCK-Os4	Protein tyrosine kinase	
24770.1	Putative Gene		

24780.1	Serine-type endopeptidase activity		Subtilisin/Kexin-related Serine protease
24790.1	BTB	BTB/POZ domain	
24800.1	Neutral amino acid transmembrane transporter activity, L-amino acid transmembrane transporter activity	Transmembrane amino acid transporter protein	Amino acid transporter
24810.1	Neutral amino acid transmembrane transporter activity, L-amino acid transmembrane transporter activity	Transmembrane amino acid transporter protein	Amino acid transporter
24820.1	Endoribonuclease activity	eRF1 domain	PELOTA
24830.1	Putative Gene		
24836.1	Putative Gene	ABC1 family	ABC transporter-related
24842.1	Putative Gene		
24850.1	Similar to UDP-arabinopyranose mutase. Reversibly glycosylated polypeptide. CAZy family GT75	Reversibly glycosylated polypeptiede	
24860.1	Putative Gene		
24870.1	STE_MEKK_ste11_MAP3K.24 - STE kinases include homologs to sterile 7, sterile 11 and sterile 20 from yeast, expressed, subfamily	Protein kinase domain	
24870.2	STE_MEKK_ste11_MAP3K.24 - STE kinases include homologs to sterile 7, sterile 11 and sterile 20 from yeast, expressed, subfamily	Protein kinase domain	
24870.3	STE_MEKK_ste11_MAP3K.24 - STE kinases include homologs to sterile 7, sterile 11 and sterile 20 from yeast, expressed, subfamily	Protein kinase domain	
24880.1	Putative Gene	Heavy-metal-associated domain	Copper transport protein ATOX1-related
24890.1	Inositol pentakisphosphate 2-kinase activity, ATP binding	Inositol-pentakisphosphate 2-kinase	Inositol polyphosphate kinase 1
24900.1	Putative Gene		
24900.3	Putative Gene		

	24910.1	Putative Gene		
	24917.1	Putative Gene	Eukaryotic DNA topoisomerase I	DNA topoisomerase Type 1
	24930.1	Putative Gene		Molybdopterin biosynthesis protein
	24930.3	Putative Gene		Molybdopterin biosynthesis protein
	24930.4	Putative Gene		Molybdopterin biosynthesis protein
	24937.1	Putative Gene		
	24950.1	Transcription coactivator activity	MED7 protein	
	24950.3	Transcription coactivator activity	MED7 protein	
	24960.1	26S, subfamily 19S	Mov34/MPN/PAD-1 family	EIF3F-related
	24967.1	Putative Gene	BSD domain	
	24980.1	Putative Gene		40S Ribosomal protein S9

Appendix F

Summary of the genes that were identified in all four pairwise comparisons to have a significant change in expression levels. The Pvalue and logFC values are from the comparison of L166 and L163. The candidate gene is highlighted in red.

Brachypodium Gene	Pvalue	logFC	Arabidopsis Gene	Gene Name	Discription
Bradi4g21300.1	2.04E-35	-8.148			Unknown
Bradi1g16150.1	1.68E-33	4.616	AT5G19130.1		GPI transamidase component family protein / Gaa1-like family protein
Bradi4g03560.2	6.88E-27	-3.518	AT2G17570.1		Undecaprenyl pyrophosphate synthetase family protein
Bradi4g03620.1	1.58E-23	-6.167			Unknown
Bradi4g29960.1	2.28E-23	-2.531	AT1G73950.1		Transmembrane Fragile-X-F-associated protein
Bradi4g03560.1	1.07E-22	-3.229	AT2G17570.1		Undecaprenyl pyrophosphate synthetase family protein
Bradi4g21710.2	6.30E-22	-6.390			Unknown
Bradi3g26960.1	1.85E-21	7.258	AT3G11290.1		
Bradi3g26420.1	5.04E-21	4.710			Unknown
Bradi3g26420.2	2.66E-20	4.508	AT5G08020.1	ATRPA70B,RPA70B	RPA70-kDa subunit B
Bradi4g21310.1	1.94E-19	-7.596			Unknown
Bradi3g55320.1	2.81E-19	1.590			Unknown
Bradi2g34120.1	1.13E-15	-2.906	AT3G15630.1		
Bradi4g21710.3	2.06E-14	-5.669			Unknown
Bradi2g39820.1	2.41E-14	6.786	AT1G53500.1	ATMUM4,ATRHM2,MUM4, RHM2	NAD-dependent epimerase/dehydratase family protein
Bradi4g21710.1	2.51E-13	-5.366			Unknown
Bradi4g21280.1	9.51E-13	2.095			Unknown
Bradi4g21720.1	6.45E-12	-6.817			Unknown
Bradi5g24290.1	1.10E-11	1.877	AT5G67230.1	I14H,IRX14-L	Nucleotide-diphospho-sugar transferases superfamily protein
Bradi2g34120.2	1.18E-11	-2.499	AT3G15630.1		

Bradi2g08630.1	7.17E-11	6.641			Unknown
Bradi2g39670.1	1.82E-10	-2.388			Unknown
Bradi1g22440.1	2.79E-10	1.708	AT5G09530.1		hydroxyproline-rich glycoprotein family protein
Bradi2g39120.1	1.10E-09	3.842			Unknown
Bradi1g06450.1	1.37E-09	1.709	AT1G78380.1	ATGSTU19,GST8,GSTU19	glutathione S-transferase TAU 19
Bradi1g15630.1	2.50E-09	-2.091			Unknown
Bradi3g22950.1	6.01E-09	-6.161	AT3G29670.1		HXXXD-type acyl-transferase family protein
Bradi4g21730.1	1.08E-08	-2.708			Unknown
Bradi1g22270.1	1.40E-08	-1.679	AT4G27250.1		NAD(P)-binding Rossmann-fold superfamily protein
Bradi4g03650.1	4.59E-08	-1.199	AT3G47570.1		Leucine-rich repeat protein kinase family protein
Bradi1g55620.1	4.68E-08	1.954	AT4G14690.1	ELIP2	Chlorophyll A-B binding family protein
Bradi1g55560.1	6.67E-08	2.275	AT4G14690.1	ELIP2	Chlorophyll A-B binding family protein
Bradi2g54960.1	8.47E-08	3.024			Unknown
Bradi3g08310.1	2.14E-07	0.888			Unknown
Bradi1g58730.1	2.15E-07	3.433			Unknown
Bradi2g39050.1	2.56E-07	5.832			Unknown
Bradi3g28460.1	2.64E-07	-2.275	AT1G02205.2	CER1	Fatty acid hydroxylase superfamily
Bradi1g73120.1	4.52E-07	-1.352	AT4G39660.1	AGT2	alanine:glyoxylate aminotransferase 2
Bradi3g09500.1	5.07E-07	-5.230	AT3G45070.1		P-loop containing nucleoside triphosphate hydrolases superfamily protein
Bradi1g35600.1	5.39E-07	-1.371	AT5G38530.1	TSBtype2	tryptophan synthase beta type 2
Bradi2g39630.1	8.83E-07	-4.674	AT3G07040.1	RPM1,RPS3	NB-ARC domain-containing disease resistance protein
Bradi3g32320.1	9.41E-07	5.024	AT1G11475.1	NRPB10,NRPD10,NRPE10	RNA polymerases N / 8 kDa subunit
Bradi1g58430.1	2.32E-06	-1.243	AT1G30220.1	ATINT2,INT2	inositol transporter 2
Bradi1g15840.1	2.70E-06	-0.964	AT3G25780.1	AOC3	allene oxide cyclase 3
Bradi2g39540.1	3.43E-06	5.788			Unknown
Bradi2g33030.1	3.47E-06	-1.614	AT3G24315.1	AtSec20	Sec20 family protein
Bradi3g03100.1	4.40E-06	-2.821			Unknown
Bradi2g55300.1	8.71E-06	-0.937			Unknown

Bradi4g11560.1	1.87E-05	-1.058	AT1G44790.1		ChaC-like family protein
Bradi2g34750.1	2.08E-05	-0.965	AT1G23410.1		Ribosomal protein S27a / Ubiquitin family protein
Bradi1g58680.1	4.31E-05	5.460			Unknown
Bradi2g12160.1	5.15E-05	0.707	AT4G01850.1	AtSAM2,MAT2,SAM-2,SAM2	S-adenosylmethionine synthetase 2
Bradi1g22780.1	8.46E-05	-5.031	AT4G29090.1		Ribonuclease H-like superfamily protein
Bradi2g33700.1	8.61E-05	-1.003			Unknown
Bradi2g55200.1	0.000241	-0.556	AT3G09050.1		
Bradi2g49220.1	0.000303	-1.054	AT2G20585.3	NFD6	nuclear fusion defective 6
Bradi4g03620.4	0.000307	-5.144			Unknown
Bradi4g22740.1	0.000604	-5.045	AT3G14470.1		NB-ARC domain-containing disease resistance protein
Bradi1g18550.1	0.000772	-0.555			Unknown
Bradi3g13670.2	0.00112	0.845	AT1G64680.1		
Bradi4g03620.3	0.00116	-4.921			Unknown
Bradi1g78260.1	0.00123	-2.314	AT4G12490.1		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
Bradi3g08290.1	0.00148	-0.570	AT4G38510.5		ATPase, V1 complex, subunit B protein
Bradi3g27370.1	0.00223	3.336	AT5G55000.2	FIP2	potassium channel tetramerisation domain-containing protein / pentapeptide repeat-containing protein
Bradi2g39720.1	0.00224	1.290	AT1G66910.1		Protein kinase superfamily protein

List of Abbreviations

4CL – Hydroxycinnamate-CoA ligase

ΔCI – Change in crystallinity index

ΔCr.R – Change in degree of cellulose crystallinity

2n – Diploid

4n – Tetraploid

A – Absorbance

A – Adenine

ADF – Acid detergent fibre

ADL – Acid detergent lignin

AFEX – Ammonia fibre explosion

AIR – Alcohol insoluble residue

Alfalfa – *Medicago sativa*

ANOVA – Analysis of variance

Arabidopsis – *Arabidopsis thaliana*

ATR-FTIR – Attenuated total reflectance Fourier transform infrared spectroscopy

bm – Brown mid-rib mutant

Brachypodium – *Brachypodium distachyon*

brs – Barley stripe resistant mutant

C – Cytosine

C3H – 4-Courmarate-3-hydroxylase

C4H – Cinnamate-4-hydroxylase

CAD – Cinnamyl-alcohol dehydrogenase

CCoA-3H – 4-Hydroxycinnamoyl-CoA

CCoA-OMT – 5-Hydroxyferuloyl-CoA-*O*-methltransferase

CCR – Cinnamoyl-CoA-reductase

CI – Crystallinity index

CO_2 – Carbon dioxide

CRM – Certified reference materials

Cr.R – Ratio of crystallinity

Csl – Cellulose synthase-like

EU – European Union

E5 – Blend of fuel containing 5% ethanol

E10 – Blend of fuel containing 10% ethanol

F5H – Ferulate-5-hydroxylase

G – Guanine

G – Guaiacyl lignin monomer

GlcUA – Glucuronic acid

GT – Glycosyltransferases

H – *p*-Hydroxyphenyl lignin monomer

H^2 – Broad sense heritability

H_2SO_4 – Sulphuric acid

H_4SiO_4 – Monosilic acid

HCl – Hydrochloric acid

HIFs – Heterogeneous inbred lines

HPAEC – High performance anion exchange chromatography

IBL – International Brachypodium initiative

irx – Irregular xylem mutant

LOD – Logarithm of the odds

LOI – Lateral order index

LWA – Liquid hot water

MBTH – 3-methy-2benzothiazolinonehydrazone

Me-GlcUA – Methelated glucuronic acid

Miscanthus – *Miscanthus*

NaOH – Sodium hydroxide

NDF – Neural detergent fibre

NILs – Near isogenic lines

OMT – *O*-methyltranserase

PAL – Phenylalanine-ammonia lyase

PCA – Principle component analysis

Poplar – *Populus*

QTL – Quantitative trait loci

RIL – Recombinant inbred line

RIN – RNA integrity number

RNAseq – RNA sequencing

S – Syringyl lignin monomer

SEC – Size exclusion chromatography

Si – Silica

SIFT – Sorting intoleranct from tolerant

SiO_2 – Silicon dioxide

SNP – Single nucleotide polymorphism

T-DNA – Transfer DNA

USDA – United States Department of Agriculture

V_G – Genotype variance

V_E – Environmental variance

V_T – Total variance

XRD – X-ray diffraction

References

Abidi N., Cabrales L., Haigler C.H. Changes in the cell wall and cellulose content of developing cotton fibers investigated by FTIR spectroscopy. Carbohydrate Polymers 2014; 100: 9 – 16.

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

An T., Cai Y., Zhao S/. Zhou J., Song B., Bux H., Qi X. *Brachypodium distachyon* T-DNA insertion lines: a model pathosystem to study nonhost resistance to wheat stripe rust. Scientific Reports 6: 1 – 9.

Andersson-Gunnerås S., Mellerowics E.J., Love J., Segerman B., Ohmiya Y., Coutinho P.M., Nilsson P., Henrissat B., Moritz T., Sundberg B. Biosynthesis of cellulose-entriched tension wood in *Populus*: Global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis. The Plant Journal 2006; 45(2): 144 – 165.

Anthon G.E., Barrett D.M. Determination of reducing sugars with 3-methyl-2-benzothiazolinonehydrazone. Analytical Biochemistry 2002; 305: 287 – 289.

Barrière Y., Thomas J., Denoue D. QTL mapping for lignin content, lignin monomeric composition, p-hydroxycinnamate content, and cell wall digestibility in the maize recombinant inbred line progeny F838 x F286. Plant Science 2008; 175: 585 – 595.

Bevan M.W., Garvin D.F., Vogel J.P. *Brachypoidium distachyon* genomics for sustainable food and fuel production. Current Opinion in Biotechnology 2010; 21: 211 – 217.

Borevitz J.O., Chory J. Genomics tools for QTL analysis and gene discovery. Current Opinion in Plant Biology 2004; 7: 132 – 136.

BP Statistical Review of World Energy, June 2015.

Brenner E.A., Zein I., Chen Y., Andersen J.R., Wenzel G., Ouzunova M., Eder J., Darnhofer B., Frei U., Barrière Y., Lübberstedt T. Polymorphisms in O-methyltransferase genes are associated with stover cell wall digestibility in European maize (*Zea mays* L.). BMC Plant Biology 2010; 10: 27.

Broman K.W, Sen S. A guide to QTL mapping with R/qtl. Springer 2009.

Brown D.M., Goubet F., Wong V.W., Goodacre R., Stephens E., Dupree P., Turner S.R. Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. The Plant Journal 2007; 52: 1154 – 1168.

Brown D.M., Zhang Z., Stephens E., Dupree P., Turner S.R. Characterization of IRX10 and IRX10-like reveals an essential role in glucuronoxylan biosynthesis in Arabidopsis. The Plant Journal 2009; 57: 732 – 746.

Cardinal A.J., Lee M., Moore K.J. Genetic mapping and analysis of quantitative trait loci affecting fiber and lignin content in maize. Theoretical and Applied Genetics 2003; 106: 866 – 874.

Carpita N.C. Progress in the biological synthesis of the plant cell wall: new ideas for improving biomass for bioenergy. Current Opinion in Biotechnology 2012; 23: 330 – 337.

Carrillo F., Colom X., Suñol J.J., Saurina J. Structural FTIR analysis and thermal characterisation of lyocell and viscose-type fibres. European Polymer Journal 2004; 40: 2229 – 2234.

Carroll A., Somerville C. Cellulosic biofuels. Annual Review of Plant Biology 2009; 60: 165 – 182.

CCAFS-CIARG (Research programme in climate change, agriculture and food security) www.ccafs.ciarg.org.

Cheali P., Posada J.A., Gernaey K.V., Sin G. Upgrading of lignocellulosic biorefinery to value-added chemicals: Sustainability and economics of bioethanol-derivatives. Biomass and Bioenergy 2015; 75: 282 – 300.

Chaturvedi V., Verma P. An overview of key pretreatment processes employed for bioconversion of lignocellulosic biomass into biofuels and value added products. Biotech 2013; 3: 415 – 431.

Chen F., Dixon R.A. Lignin modification improves fermentable sugar yields for biofuel production. Nature Biotechnology 2007; 25(7): 759 – 761.

Chen X., Vega-Sánchez M.E., Verhertbruggen Y., Chiniquy D., Canlas P.E., Fagerström A., Prak L., Christensen U., Oikawa Ai., Chern M., Zuo S., Lin F., Auer M., Willats W.G.T., Bartley L., Harholt J., Scheller H.V., Ronald P.C. Inactivation of *Os*IRX10 leads to decreased xylan content in rice culm cell walls and improved biomass saccharification. Molecular Plant 2013; 6(2): 570 – 573.

Chiniquy D., Varanasi P., Oh T., Harholt J., Katnelson J., Singh S., Auer M., Simmons B., Adams P.D., Scheller H.V., Ronald P.C. Three novel rice genes closely related to the *Arabidopsis IRX9*, *IRX9L*, and *IRX14* genes and their roles in xylan biosynthesis. Frontiers in Plant Science 2013; 4(83): 1 – 13.

Christensen U., Alonso-Simon A., Scheller H.V., Willats W.G.T., Harholt J. Characterization of the primary cell walls of seedlings of Brachypodium distachyon – A potential model plant for temperate grasses. Phytochemistry 2010; 71(1): 62 – 69.

Chundawat S.P.S, Balan V., Dale B.E. High-throughput microplate technique for enzymatic hydrolysis of lignocellulosic biomass. Biotechnology and Bioengineering 2008; 99:1281 – 1294.

Ciolacu D., Ciolacu F., Popa V.I. Amorphous cellulose – structure and characterization. Cellulose Chemistry and Technology 2011; 45: 13 – 21.

Coimbra M.A., Barros A., Rutledge D.N. Delgadillo I. FTIR spectroscopy as a tool for the analysis of olive pulp cell-wall polysaccharide extracts. Carbohydrate Research 1999; 317: 145 – 154.

Collard B.C.Y., Jahufer M.Z.Z., Brouwer J.B., Pang E.C.K. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 2005; 142: 169 – 196.

Costa V., Angelini C., De Feis I., Ciccodicola A. Uncovering the complexity of transcriptomes with RNA-Seq. Journal of Biomedicine and Biotechnology 2010; 2010: 1 – 19.

Cui Y, Lee M.Y., Huo N., Bragg J., Yan L., Yuan C., Li C., Holditch S.J., Xie J., Luo M-C., Li D., Yu J., Martin J., Schackwitz W., Gu Y.Q., Vogel J.P., Jackson A.O., Liu Z., Garvin D.F. Fine mapping of the *Bsr1* Barley Stripe Mosic Virus resistance gene in the model grass Brachypodium *distachyon*. PLoS ONE 2012; 7(6): 1 – 10.

Currie H.A., Perry C.C. Silica in plants: Biological, biochemical and chemical studies. Annals of Botany 2007; 100: 1383 – 1389.

Dalmais M., Antelme S., Ho-Yue-Kuang S., Wang Y., Darracq O., Bovier d'Yvoire M., Cézard L., Legée F., Blondet E., Oria N., Troadec C., Brunaud V., Jouanin L., Höfte H., Bendahmane A., Lapierre C., Sibout R. A TILLING platform for functional genomics in *Brachypodium distachyon*. PLoS One 2013: 8(6); 1 – 10.

Decker S.R., Brunecky R., Tucker M.P., Himmel M.E., Selig M.J. High-throughput screening techniques for biomass conversion. Bioenergy Research 2009; 2: 179 – 192.

De Jong E., Jungmeier G. Biorefinery concepts in comparison to petrochemical refineries. In: Industrial Biorefineries and White Biotechnolog. Elsevier 2015: 3 – 33.

Demirbas A. Biofuels: Securing the planet's future energy needs. London: Springer 2009.

Dence C.W. The determination of lignin. In: Methods in Lignin Chemistry. Lin S, Y, Dence CW, editors. Springer-Verlag 1992: 33–61.

De Souza A.P., Alvim Kamei C.L., Torres A.F., Pattathil S., Hahn M.G., Trindade L.M., Buckeridge M.S. How cell wall complexity influences saccharification efficiency in *Miscanthus sinenis*. Journal of Experimental Botany 2015; 66(14): 4351 – 4365.

De Storme N., Mason A. Plant speciation through chromosome instability and ploidy change: Cellular mechanisms, molecular factors and evolutionary relevance. Current Plant Biology 2014; 1: 10 – 33.

Directive 2009/28/EC of the European Parliament and of the Council of 23 April 2009 on the promotion of the use of energy from renewable sources and amending and subsequently repealing directives 2001/77/EC and 2003/30/EC. OJ L 140, 5.6.2009 p16 – 62.

Dhugga K.S. Biosynthesis on non-cellulosic polysaccharides of plant cell walls. Phytochemistry 2012; 74: 8 – 19.

Di Maio D., Turley D. Lignocellulosic feedstock in the UK: A report for the lignocellulosic biorefinery network. NNFCC: The Bioeconomy consultants 2014; 1 – 45.

Ding S-Y., Himmel M.E. Anatomy and ultrastructure of maize cell walls: An example of energy plants. In: Himmel M.E., Biomass recalcitrance: Deconstructing the plant cell wall for bioenergy. Oxford: Wiley-Blackwell 2008. Chapter 3.

Doerge R.W. Mapping and analysis of quantitative trait loci in experimental populations. Nature Reviews Genetics 2002; 3: 43 – 52.

Draper J., Mur L.A.J., Jenkins G., Ghosh-Biswas G.C., Bablak P., Hasterok R., Routledge A.P.M. *Brachypodium distachyon*. A new model system for functional genomics in grasses. Plant Physiology 2001; 127: 1539 – 1555.

Eerhart A.J.J.E., Patel M.K., Faaij A.P.C. Fuels and plastics from lignocellulosic biomass via the furan pathway: an economic analysis. Biofuels, Bioproducts and Biorefining 2015; 9: 307 − 325.

Elgersma A., Nieboer I.G., Keizer L.C.P. The effect of temperature on seed set and seed development in detached spikelets of perennial Ryegrass (*Lolium perenne* L.). Annals of Botany 1993; 72: 337 − 340.

Epstein E. The anomaly of silicon in plant biology. PNAS 1994; 91: 11 − 17.

Fagard M., Höfte H., Vernhettes S. Cell wall mutants. Plant Physiology and Biochemistry 2000; 38: 15 − 25.

Falano T., Jeswani H.K., Azapagic A. Assessing the environmental sustainability of ethanol from integreated biorefineries. Biotechnology Journal 2014; 9: 753 − 765.

Fazio G., Staub J.E., Stevens M.R. Genetic mapping and QTL analysis of horticultural traits in cucumber (*Cucumis sativus* L.) using recombinant inbred lines. Theoretical and Applied Genetics 2003; 107: 864 − 874.

Feuillet C., Keller B. Comparative genomics in the grass family: Molecular characterization of grass genome structure and evolution. Annuals of Botany 2002; 89: 3 − 10.

Flint-Garcia S.A., Tornsberry J.M., Buckler IV E.S. Annual Review of Plant Biology 2003; 54: 357 − 374.

Foster C.E., Martin T.M., Pauly M. Comprehensive compositional analysis of plant cell walls (lignocellulosic biomass). Part I: Lignin. The Journal of Visualized Experiment 2010a; 37.

Foster C.E., Martin T.M., Pauly M. Comprehensive compositional analysis of plant cell walls (lignocellulosic biomass). Part II: Carbohydrates. The Journal of Visualized Experiment 2010b; 37.

Foust T.D., Ibsen K.N., Dayton D.C., Hess J.R., Kenney K.E. The Biorefinery. In: Himmel M.E. Biomass recalcitrance: Deconstructing the plant cell wall for bioenergy. Oxford: Wiley-Blackwell 2008. Chapter 2.

Gadde B., Bonnet S., Menke C., Garivait S. Environmental Pollution 2009; 157 (5): 1554 – 1558.

Garvin D.F. *Brachypodium distachyon*: A new model system for structural and functional analysis of grass genomes. In: Varshney R.K., Koebner R.M.D. editiors. Model Plants and Crop Improvement: Taylor & Francis, Boca Raton, Fl. 2007: 109 – 123.

Garvin D.F., Gu Y-G., Hasterok R., Hazen S.P., Jenkins G., Mockler T.C., Mur L.A.J., Vogel J.P. The Plant Genome 2008; 1: S69 – S84.

Gery C., Zuther E., Schultz E., Legoupi J., Chauveau A., McKhann H., Hincha D.K., Téoulé E. Natural variation in the freezing tolerance of *Arabidopsis thaliana*: Effects of RNAi-induced CBF depletion and QTL localisation vary among accessions. Plant Science 2011; 180: 12 – 23.

Giovane A., Servillo L., Balestrieri C., Raiola A., D'Avino R., Tamburrini M., Ciardiello M.A., Camardella L. Pectin methylesterase inhibitor. Biochimica et Biophysica Acta (BBA) – Proteins and Proteomics 2004; 1696(2): 245 – 252.

Girin T., David L.C., Chardin C., Sibout R., Krapp A., Ferrario-Méry S., Daniel-Vedele F. Brachypodium: a promising hub between model species and cereals. Journal of Experimental Botany 2014; 65 (19): 5683 – 5696.

Gomez L.D., Steele-King C.G., McQueen-Mason S.J. Sustainable liquid biofuels from biomass: The writing's on the walls. New Phytologist 2008; 178(3): 473 – 485.
Gomez L.D., Bristow J.K., Statham E.R., McQueen-Mason S.J. Analysis of saccharification in *Brachypodium distachyon* stems under mild conditions of hydrolysis. Biotechnology for Biofuels 2008; 1(15): 1 – 12.

Gomez L.D., Whitehead C., Barakate A., Halpin C., McQueen-Mason S.J. Automated saccharification assay for determination of digestibility in plant materials. Biotechnology for Biofuels 2010; 3(23): 1 – 12.

Gomez L.D., Whitehead C., Roberts P., McQueen-Mason S.J. High-throughput saccharification assay for lignocellulosic materials. Journal of Visual Experiments 2011; 53, e3240.

Goubet F., Barton C.J., Mortimer J.C., Yu X., Zhang Z., Miles G.P., Richens J., Lipeman A.H., Seffen K., Dupree P. Cell wall glucomannan in Arabidopsis is synthesised by CSLA glylcosyltranserases, and influences the progression of embryogenesis. The Plant Journal 2009; 60: 527 – 538.

Grass L., Burris J.S. Effect of heat stress during seed development and maturation on wheat (*Triticum durum*) seed quality. I. Seed germination and seedling vigor. Canadian Journal of Plant Science 1995; 75(4): 821 – 829.

Hall M., Bansal P., Lee J.H., Realff M.J., Bommarius A.S. Cellulose crystallinity – a key predictor of the enzymatic hydrolysis rate. The FEBS Journal 2010; 277: 1571 – 1582.

Halpin C., Holt K., Chojecki J., Oliver D., Chabbert B., Monties B., Edwards K., Barakate A., Foxon G.A. *Brown-midrib* maize (*bm1*) – a mutation affecting the cinnamyl alcohol dehydrogenase gene. The Plant Journal 1998; 14(5): 545 – 553.

Han L., Feng J., Zhang S., Ma Z., Wang Y., Zhang X. Alkali pretreated of wheat straw and its enzymatic hydrolysis. Brazilian Journal of Microbiology 2012; 43: 53 – 61.

Harmsen P.F.H., Huijgen W.J.J., Bermúdez López L.M., Bakker R.R.C. Literature review of physical and chemical pretreatment processes for lignocellulosic biomass. Biosynergy Project Report 2010; 1 – 49.

Harris D., DeBolt S. Relative crystallinity of plat biomass: Studies on assembly, adaptation and acclimation. Plos One 2008; 3(8): 1 – 10.

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

Himmel M.E., Picataggio S.K. Our challenge is to acquire deeper understanding of biomass recalcitrance and conversion. In: Himmel M.E., 2008, Biomass recalcitrance: Deconstructing the plant cell wall for bioenergy. Oxford:Wiley-Blackwell. Ch1.

Hoekman S.K. Biofuels in the US – challenges and opportunities. Renewable Energy 2009; 34: 14 – 22.

Hori R., Sugiyama J. A combined FR-IR microscopy and principal component analysis on softwood cell wall. Carbohydrate Polymers 2003; 52: 449 – 453.

Ho-Yue-Kuang S., Alvarado C., Antelme S., Bouchet B., Cézard L., Le Bris P., Legée F., Maia-Grondard A., Yoshinaga A., Saulnier L., Guillon F., Sibout R., Lapierre C., Chateigner-Boutin A-L. Mutation in *Brachypodium* caffeic acid *O*-methyltransferase 6 alters stem and grain lignins and improves straw saccharification without deterioration grain quality. Journal of Experimental Botany. 2015; 1 – 11.

Huo N., Garvin D.F., You F.M. McMahon S., Luo M-C., Gu Y.Q., Lazo G.R., Vogel J.P. Comparison of a high-density genetic linkage map to genome features in the model grass *Brachypodium distachyon*. Theoretical and Applied Genetics 2011; 123: 455 – 469.

Huson D.H., Richter D.C., Rausch C., Dezulian T., Franz M., Rupp R. Dendroscope: an interactive viewer for large phylogenetic trees. BMC Bioinformatics 2007; 8(1): 460.

Ingram P.A., Zhu J., Shariff A., Davis I.W., Benfey P.N., Elich T. High-throughput imaging and analysis of root system architecture in *Brachypodium distachyon* under differential nutrient availability. Philosophical Transactions of the Royal Society B 2012; 367: 1559 – 1569.

Jeswani H.K., Falano T., Azapagic A. Biofuels, Bioproducts and Biorefining 2015; 9: 661 – 676.

Jin S., Chen H. Structurla properties and enzymatic hydrolysis of rice straw. Process Biochemistry 2006; 41: 1261 – 1264.

Jones L.H.P, Handreck K.A. Silica in soils, plants and animals. Advances in Agronomy 1967; 19: 107 – 149.

Jones L., Milne J.L, Ashford D., McQueen-Mason S.J. Cell wall arabinan is essential for guard cell function. PNAS 2003; 100: 11783 – 11788.

Jung H.G. Forage lignins and their effects on fiber digestibility. Agronomy Journal 1989; 81 (1): 33 – 38.

Kačuráková M., Capek P., Sasinková V., Wellner N., Ebringerová A. FT-IR study of plant cell wall model compounds: pectic polysaccharides and hemicelluloses. Carbohydrate Polymers 2000; 43: 195 – 203.

Kačuráková M., Wilson R.H. Developments in mid-infrared FT-IR spectroscopy of selected carbohydrates. Carbohydrate Polymers 2001; 44: 291 – 303.

Koh L.P., Ghazoul J. Biofuels, biodiversity, and people: Understanding the conflicts and finding opportunities. Biological Conservation 2008; 141: 2450 – 2460.

Kohorn B.D. WAKs; cell wall associated kinases. Current Opinion in Cell Biology 2001; 13(5): 529 – 533.

Krakowsky M.D., Lee M., Coors J.G. Quantitative trait loci for cell-wall components in recombinant inbred lines of maize (*Zea mays* L.)I: Stalk tissue. Theoretical and Applied Genetics 2005; 111: 337 – 346.

Kumar R., Mago G., Balan V., Wymand C.E. Physical and chemical characterizations of corn stover and poplar solids resulting from leading pretreatment technologies. Bioresource Technology 2009; 100(17): 3948 – 3962.

Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A, McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gobson T.J., Higgins D.G. Clustal W and Clustal X version 2.0. Bioinformatics 2007; 23: 2947 – 2948.

Lee J. Kwon S.W. Analysis of quantitative trait loci associated with seed germination and coleoptile length under low temperature condition. Journal of Crop Science and Biotechnology 2015; 18(4): 273 – 278.

Le Gall H., Philippe F., Domon J-M., Gillet F., Pelloux J., Rayon C. Cell wall metabolism in response to abiotic stress. Plants 2015; 4: 112 – 166.

Leroxel O., Cavalier D.M., Liepman A.H., Keegstra K. Biosynthesis of plant cell wall polysaccharides – a complex process. Current Opinion in Plant Biology 2006; 9: 621 – 630.

Li C., Knierim B., Manisseri C., Arora R., Scheller H.V., Auer M., Vogel K.P., Simmons B.A., Singh S. Comparison of dilute acid and ionic liquid pretreatment of switchrass: Biomass recalcitrance, delignification and enzymatic saccharification. Bioresource Technology; 101(13): 4900 – 4906.

Li F., Ren S., Zhang W., Xu Z., Xie G., Chen Y., Tu Y., Li Q., Zhou S., Li Y., Tu F., Liu L., Wang Y., Jiang J., Qin J., Li S., Li Q., J H-C., Zhou F., Gutterson N. Arabinose substitution degree in xylan positively affects lignocellulose enzymatic digestibility after various NaOH/H_2SO_4 pretreatment in *Miscanthus*. Bioresource Technology 2013; 130; 629 – 637.

Li X, Ximenes E., Kim Y., Slininger M., Meilan R., Ladisch M., Chapple C. Lignin monomer composition affects Arabidopsis cell-wall degradability after liquid hot water pretreatment. Biotechnology for Biofuels 2010; 3(27): 1 – 7.

Li Z., Zhao C., Zha Y., Wan C., Si S., Liu F., Zhang R., Li F., Yu B., Yi Z., Xu N., Peng L., Li Q. The minor wall-networks between monolignols and interlinked-phenolics predominantly affect biomass enzymatic digestibility in *Miscanthus*. PLoS ONE 2014; 9(8): 1 – 9.

Liepman A.H., Wilkerson C.G., Keegstra K. Expression of cellulose synthase-like (*Csl*) genes in insect cells reveals that *CslA* family members encode mannan synthases. PNAS 2005; 102(6): 2221 – 2226.

Liepman A.H., Nairn C.J., Willats W.G.T., Sørensen I., Roberts A.W., Keegstra K. Functional genomic analysis supports conservation of function among cellulose synthase-like A gene family members and suggests diverse roles of mannans in plants. Plant Physiology 2007; 143: 1881 – 1893.

Liepman A.H., Cavalier D.M. The cellulose synthase-like A and cellulose synthase-like C families: Recent advances and future perspectives. Frontiers in Plant Science 2012; 3(109): 1 – 7.

Lionikas A., Meharg C., Derry J.M.J., Ratkevicius A., Carroll A.M., Vandenbergh D.J., Blizard D.A. Resolving candidate genes of mouse skeletal muscle QTL via RNA-Seq and expression network analyses. BMC Genomics 2012; 13(592): 1 – 14.

Liu Y. Recent progress in Fourier transform infrared (FTIR) spectroscopy study of compositional, structural and physical attributes of developmental cotton fibers. Materials 2013; 6: 299 – 313.

Li X., Ximenes E., Kim Y., Slininger M., Meilan R., Ladisch M., Chapple C. Lignin monomer composition affects Arabidopsis cell-wall degradability after liquid hot water pretreatment. Biotechnology for Biofuels 2010; 3(27): 1 – 7.

Loqué D., Scheller H.V., Pauly M. Engineering of plant cell walls for enhanced biofuel production. Current Opinion in Plant Biology 2015; 25: 151 – 161.

Luo N., Liu J., Yu X., Jiang Y. Natural variation of drought response in *Brachyppdium distachyon*. Physiologia Plantarum 2011; 141: 19 – 29.

Lovegrove A., Wilkinson M.D., Freeman J., Pellny T.K., Tosi P., Saulnier L., Shewry P.R., Mitchell R.A.C. RNA interference suppression of genes in glycosyl transferase families 43

and 47 in wheat starchy endosperm causes large decreases in arabinoxylan content. Plant Physiology 2013; 163: 95 – 107.

Marriott P.E., Gomez L.D., McQueen-Mason S.J. Unlocking the potential of lignocellulosic biomass through plant science. New Phytologist 2016; 209: 1366 – 1381.

McKechnie J., Pourbafrani M., Saville B.A., MacLean H.L. Environmental and financial implications of ethanol as a bioethylene feedstock versus as a transportation fuel. Environmental Research Letters 2015; 10: 124018.

Mellidou I., Chagné D., Laing W.A., Keulemans J., Davey M.W. Allelic variation in paralogs of GDP-L-galactose phosphorylase is a major determinant of vitamin C concentration in apple fruit. Plant Physiology 2012; 160: 1613 – 1629.

Meides E., Vanholme R., Boerjan W., Molina A. The role of the secondary cell wall in plant resistance to pathogens. Frontiers in Plant Science 2014; 5: 358.

Mohammed I.Y., Abakr Y.A., Kazi F.K., Yusup S., Alshareef I., Chin S.A. Comprehensive characterization of napier grass as a feedstock for thermochemical conversion. Energies 2015; 8(5): 3403 – 3417.

Mohanram S., Amat D., Choudhary J., Arora A., Nain L. Novel perspectives for evolving enzyme cocktails for lignocellulose hydrolysis in biorefineries. Sustainable Chemical Processes 2013; 1(15): 1 – 12.

Morales M., Quintero J., Conejeros R., Aroca G. Life cycle assessment of lignocellulosic bioethanol: Environmental impacts and energy balance. Renewable and Sustainable Energy Reviews 2015; 42: 1349 – 1361.

Moran J. Tropical dairy farming: feed management for small holder dairy farmers in the humid tropics. Chapter 5: How the rumen works. Landlinks Press 2005; 312.

Morozova O., Hirst M., Marra M.A. Applications of new sequencing technologies for transcriptome analysis. Annual Review of Genomics and Human Genetics 2009; 10: 135 – 151.

Murray S.C., Rooney W.L., Mitchell S.E., Sharma A., Klein P.E., Mullet J.E., Kresovich S. Genetic improvement of sorghum as a biofuel feedstock: II. QTL for stem and leaf structural carbohydrates. Crop Science 2008; 48: 2181 – 2193.

Muylle H., Van Hulle S., De Vliegher A., Baert J., Van Bockstaele E., Roldán-Ruiz I. Yield and energy balance of annual and perennial lignocellulosic crops for bio-refinery use: A 4-year field experiment in Belgium. European Journal of Agronomy 2015; 63: 62 – 70.

Naik S.N., Goud V.V., Rout P.K., Dalai A.K. Production of first and second generation biofuels: A comprehensive review. Renewable and Sustainable Energy Reviews 2010: 14; 578 – 597.

Ng P.C, Henikoff S. Predicting deleterious amino acid substitutions. Genome Research 2001; 11(5): 863 – 874.

Norambuena L., Marchant L., Berninsone P., Hirschberg C.B., Silva H., Orellana A. Transport of UDP-galactose in plants: identification and functional characterization of AtUTr1, an *Arabidopsis thaliana* UDP-galactose/UDP-glucose transporter. The Journal of Biological Chemistry 2002; 277: 32923 – 32929.

Oakey H., Shafiei R., Comadran J., Uzrek N., Cullis B., Gomez L.D., Whitehead C., McQueen-Mason S.J., Waugh R., Halpin C. Identificaiton of crop cultivars with consistently high lignocellulosic sugar release requires the use of appropriate statistical design and modelling. Biotechnology for Biofuels 2013; 6(185): 1 – 12.

Oikawa A., Lund C.H., Sakuragi Y., Scheller H.V. Golgi-localized enzyme complexes for plant cell wall biosynthesis. Trends in Plant Science 2013; 18(1): 49 – 58.

Oleszek M., Król A., Tys J., Matyka M., Kulik M. Comparison of biogas production from wild and cultivated varieties of reed canary grass. Bioresource Technology 2014; 156: 303 – 306.

Opanowicz M., Vain P., Draper J., Parker D., Doonan J.H. *Brachypodium distachyon*: Making hay with a wild grass. Trends in Plant Science 2008; 13(4): 172 – 177.

Pacheco-Villalobos D., Hardtke C.S. Natural genetic variation of root system architecture from *Arabidopsis* to *Brachypodium*: Towards adaptive values. Philosophical Transactions of the Royal Society B 2012; 367: 1552 – 1558.

Parajuli R., Dalgaard T., Jørgensen U., Adamsen A.P.S., Knudsen M.T., Birkved M., Gylling M., Schjørring J.K. Biorefining in the prevailing energy and materials crisis: A review of sustainable pathways for biorefinery value chains and sustainability assessment methodologies. Renewable and Sustainable Energy Reviews 2015; 43: 244 – 263.

Park S., Baker J.O., Himmel M.E., Parilla P.A., Johnson D.K. Cellulose crystallinity index: Measurement techniques and their impact on interpreting cellulose performance. Biotechnology for Biofuels 2010; 3(10): 1 – 10.

Parker G.D., Chalmers K.J., Rathjen A.J., Langridge R. Mapping loci associated with flour colour in wheat (*Tricticum aestivum* L.). Theoretical and Applied Genetics 1998; 97: 238 – 254.

Persson S., Wei H., Milne J., Page G.P., Somerville C.R. Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. PNAS 2005; 102(24): 8633 - 8638.

Petersen P.D., Lau J., Ebert B., Yang F., Verhertbruggen Y., Kim J.S., Varanasi P., Suttangkakul A., Auer M., Loqué D., Scheller H.V. Engineering of plants with improved properties as biofuels feedstocks by vessel-specific complementation of xylan biosynthesis mutants. Biotechnology for Biofuels 2012; 5(84): 1 – 19.

Pinzón-Latorre D., Keyholos M.K. Characterization and transcript profiling of the pectin methylesterase (PME) and pectin methylesterase inhibitor (PMEI) gene families in flax (*Linum usitatissimum*). BMC Genomics 2013: 14: 742.

Prioul J-L., Pelleschi S., Séne M., Thévenot C., Causse M., de Vienne D., Leonardi A. From QTLs for enzyme activity to candidate genes in maize. Journal of Experimental Botany 1999: 50(337): 1281 – 1288.

Reidinger S., Ramsey M.H., Hartley S.E. Rapid and accurate analyses of silicon and phosphorus in plants using a portable X-ray fluorescence spectrometer. New Phytologist 2012; 195: 699 – 706.

Ren Y., Hansen S.F., Ebert B., Lau J., Scheller H.V. Site-directed mutagenesis of IRX9, IRX9L and IRX14 proteins involved in xylan biosynthesis: Glycosyltransferase activity is not required for IRX9 function in Arabidopsis. PloS ONE 2014; 9(8): 1 – 9.

Richmond T.A., Somerville C.R. The cellulose synthase superfamily. Plant Physiology 2000; 124: 495 – 498.

Rollin J.A., Zhu Z., Sathitsuksanoh N., Zhang Y-H.P. Increasing cellulose accessibility is more important than removing lignin: A comparison of cellulose solvent-based lignocellulose fractionation and soaking in aqueous ammonia. Biotechnology and Bioengineering 2001; 108: 22 – 30.

Schartz C.J. Doyle M.R. Manzaneda A.J., Rey P.J., Mitchell-Olds T., Amazino R.M. Natural variation of flowering time and vernalization responsiveness in *Brachypodium distachyon*. Bioenergy Research 2010; 3: 38 – 46.

Scheller H.V., Ulskov P. Hemicelluloses. Annual Review of Plant Biology. 2010; 61: 263 – 289.

Schwartz C.J., Doyle M.R., Manzaneda A.J., Rey P.J., Mitchell-Olds T., Amasino R.M. Natural variation of flowering time and vernalization responsiveness in *Brachypodium distachyon*. Bioenergy Research 2010; 3: 38 – 46.

Semagn K., Bjørnstad Å., Ndjiondjop M.N. An overview of molecular marker methods for plants. African Journal of Biotechnology 2006; 5(25): 2540 – 2568.

Serba D.D., Daverdin G., Bouton J.H., Devos K.M., Brummer C.E., Saha M.C. Quantitative trait loci (QTL) underlying biomass yield and plant height in switchgrass. BioEnergy Research 2015; 8(1): 307 – 324.

Shiringani A.L., Frisch M., Friedt W. Genetic mapping of QTLs for sugar-related traits in a RIL population of *Sorghum biocolour* L. Moench. Theoretical and Applied Genetics 2010; 121: 323 – 336.

Shiringani A.L., Friedt W. QTL for fibre-related traits in grain sweet sorghum as a tool for enhancement of sorghum as a biomass crop. Theoretical and Applied Genetics 2011; 123: 999 – 1011.

Sills D.L., Gossett J.M. Using FTIR to predict saccharification from enzymatic hydrolysis of alkali-pretreated biomasses. Biotechnology and Bioengineering 2012; 109: 353 – 362.

Sim N-L., Kumar P., Hu J., Henikoff S., Schneider G., Ng P.C. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Research 2012; 40; W452 – W457.

Slavov G., Allison G., Bosch M. Advances in the genetic dissection of plant cell walls: Tools and resources available in *Miscanthus*. Frontiers in Plant Science 2013; 4(217): 1 – 21.

Smith A.B., Cullis B.R., Thompson R. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. The Journal of Agricultural Science 2005; 146 (6): 449 – 462.

Studer M.H., DeMartini J.D., Brethauer S., McKenzie H.L., Wyman C.E. Engineering of a high-throughput screening system to identify cellulosic biomass, pretreatments, and enzyme formations that enhance sugar release. Biotechnology and Bioengineering 2010; 105: 231 – 238.

Sumiyoshi M., Inamura T., Nakamura A., Aohara T., Ishii T., Satoh S., Iwai H. UDP-arabinopyranose mutase 3 is required for pollen wall morphogenesis in rice (*Oryza sativa*). Plant Cell Physiology 2015; 56(2): 232 – 241.

Takahashi E., Miyake Y. Silica and plant growth. Proceedings of the International Seminar on Soil Environment and Fertility Management in Intensive Agriculture 1977: 603 – 611.

Tan Y.F., Sun M., Xing Y.Z., Hua J.P., Sun X.L., Zhang Q.F., Corke H. Mapping quantitative trait loci for milling quality, protein content and color characteristics of rice using a recombinant inbred line population derived from an elite rice hybrid. Theoretical and Applied Genetics 2001; 103: 1037 – 1045.

The International Brachypodium Initiative. Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature 2010; 463: 763 – 768.

Thole V., Worland B., Wright J., Bevan M.W., Vain P. Distribution and characterization of more than 1000 T-DNA tags in the genome of *Brachypodium distachyon* community standard line Bd21. Plant Biotechnology Journal 2010; 8 (6): 734 – 747.

Thomas J., Guillaumie S., Verdu C., Denoue D., Pichon M., Barrière Y. Cell wall phenylpropanoid-related gene expression in early maize recombinant inbred lines differing in parental alleles at a major lignin QTL position. Molecular Breeding 2010; 25: 105 – 124.

Thygesen A., Oddershede J., Lilholt H., Thomsen A.B., Ståhl K. On the determination of crystalline cellulose content in plant fibres. Cellulose 2005; 12: 563 – 576.

Till B.J., Reynolds S.H., Weil C., Springer N., Burtner C., Young K., Bowers E., Codomo C.A., Enns L.C., Odden A.R., Greene E.A., Comai L., Henikoff S. Discovery of induced point mutations in maize genes by TILLING. BMC Plant Biology 2004; 4: 12.

Timpano H., Sibout R., Devaux M-F., Alvarado C., Looten R., Falourd X., Pontoire B., Martin M., Legée F., Cézard L., Lapierre C., Badel E., Citerne S., Vernhettes S., Höfte H.,

Guillon F., Gonneau M. *Brachypodium* cell wall mutant with enhanced saccharification potential despite increased lignin content. Bioenergy Research 2014; 8: 53 – 67.

Torres A.F., Visser R.G.F., Trindade L.M. Bioethanol from maize cell walls: Genes, molecular tools, and breeding prospects. Global Change Biology Bioenergy 2015; 7: 591 – 607.

Turner S.R., Somerville C.R. Collapsed xylem phenotype of Arabidopsis identifies mutants deficient in cellulose deposition in the secondary cell wall. The Plant Cell 1997; 9: 689 – 701.

Tyler L., Bragg J.N., Wu J., Yang X., Tuskan G.A., Vogel J.P. Annotation and comparative analysis of the glycoside hydrolase genes in *Brachypodium distachyon*. BMC Genomics 2010; 11: 600.

Tyler L., Fangel J.U., Fagerström A.D., Steinwand M.A., Raab T.K., Willats W.G.T., Vogel J.P. Selection and phenotypic characterization of a core collection of *Brachypodium distachyon* inbred lines. BMC Plant Biology 2014; 14 (25): 1 – 15.

US Energy Information (EIA), US Government, US Department of Energy, EIA International Energy Statistics.

Vain P. Brachypodium as a model system for grass research. Journal of Cereal Science 2011; 54: 1 – 7.

Van der Schaar W., Alonso-Blanco C., Léon-Kloosterziel K.M., Jansen R.C., Van Ooijen J.W., Koornneef M. QTL analysis of seed dormancy in *Arabidopsis* using recombinant inbred lines and MQM mapping. Heredity 1997; 79: 190 – 200.

Vanholme R., Morreel K., Ralph J., Boerjan W. Lignin engineering. Current Opionion in Plant Biology 2008; 11: 278 – 285.

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

Vega-Sánchez M.E., Ronald P.C. Genetic and biotechnological approaches for biofuel crop improvement. Current Opinion in Biotechnology 2010; 21: 218 – 224.

Venu R.C., Ma J., Jia Y., Liu G., Jia M.H., Nobuta K., Sreerekha M.V., Moldenhauer K., McClung A.M., Meyers B.C., Wang G-L. Identification of candidate genes associated with positive and negative heterosis in rice. PLoS ONE 2014; 9(4): 1 – 15.

Vogel J. Unique aspects of the grass cell wall. Current Opinion in Plant Biology. 2008; 11: 301 – 307.

Wang B., Zhang Z., Fu Z., Liu Z., Hu Y., Tang J. Comparative QTL analysis of maize seed artificial aging between an immortalized F_2 population and its corresponding RILs. The Crop Journal 2016; 4(1): 30 – 39.

Wang M., Yuan D., Gao W., Li Y., Tan J., Zhang X. A comparative genome analysis of PME and PMEI families reveals the evolution of pectin metabolism in plant cell walls. PLoS One 2013; 8(8).

Wang Z., Gerstein M., Snyder M. RNA-Seq: A revolutionary tool for transcriptomics. Nature Reviews Genetics 2009; 10: 57 – 63.

Wu X., Zhang M., Wang L., Tu Y., Zhang J., Xie G., Xou W., Li F., Guo K., Li Q., Gao C., Peng L. Biomass digestibility is predominantly affected by three factors of wall polymer features distinctive in wheat accessions and rice mutants. Biotechnology for Biofuels 2013; 6(183): 1 – 14.

Xu N., Zhang W., Ren S., Liu F., Zhao C., Liao H., Xu Z., Huang J., Li Q., Tu Y., Yu B., Wang Y., Jiang J., Qin J., Peng L. Hemicelluloses negatively affect lignocellulose crystallinity for high biomass digestibility under NaOH and H_2SO_4 pretreatments in *Miscanthus*. Biotechnology for Biofuels 2012; 5(58): 1 – 12.

Xu X-Y., Bai G-H., Carver B.F., Shaner G.E., Hunger R.M. Mapping of QTLs prolonging the latent period of *Puccinia triticina* infection in wheat. Theoretical and Applied Genetics 2005; 110: 244 – 251.

Yin Y., Huang J., Xu Y. The cellulose synthase superfamily in fully sequenced plants and algae. BMC Plant Biology 2009; 9(99): 1 – 14.

Youngs H.L., Hamman T., Osborne E., Somerville C. Ch3: The cellulose synthase superfamily. Springer 2007; Cellulose: molecular and structural biology (Brown Jr R.M., Saxena I.M.)

Yu Q., Zhuang X., Lv S., He M., Zhang Y., Yuan Z., Qi W., Wang Q., Wang W., Tan X. Liquid hot water pretreatment of sugarcane bagasse and its comparison with chemical pretreatmet methods for the sugar recovery and structural changes. Bioresource Technology 2013; 129: 592 – 598.

Zhang W., Yi Z., Huang J., Li F., Hao B., Li M., Hong S., Lv Y., Sun W., Ragauskas A., Hu F., Peng J., Peng L. Three lignocellulose features that distinctively affect biomass enzymatic digestibility under NaOH and H2SO4 pretreatments in *Miscanthus*. Bioresource Technology 2013; 130: 30 – 37.

Zhang Z.M., Zhoa M.J., Ding H.P., Rong T.Z., Pan G.T. QTL mapping analysis of plant height and ear height of maize (*Zea mays* L.). Genetika 2