Improving *Brassica napus* straw for cellulosic ethanol production

Ian Paul Wood

Submitted for the Degree of Doctor of Philosophy

University of East Anglia

School of Biology

Institute of Food Research,

Norwich Research Park, Colney, Norwich, NR4 7UA

September 2014

Abstract

It is likely that a combination of process improvement and plant breeding will be needed to make efficient use of biomass. To achieve this, the main bottlenecks to decomposition must be identified and strategies developed to make the process more efficient.

Here, the likely process and genotypic variants relevant to saccharification efficiency of *Brassica napus* straw were investigated using steam explosion. Screening methods were developed to gather suitable data for association mapping. Areas of the *B. napus* transcriptome related to processing traits were highlighted and assessed.

The results show that autocatalytic pretreatments are effective, but commercial cellulases are poorly adapted to hydrolyse *Brassica napus* straw. Following pretreatment, saccharification was limited by uronic acids and xylose retention, which impeded cellulase performance. Surprisingly, lignin abundance correlated positively with the proportion of rapidly hydrolysable carbohydrate in the pretreated residues. Cultivars with glucan-rich straw did not necessarily produce higher saccharification or ethanol yields. Instead, variations in the non-cellulosic components were the most important differences between cultivars.

Smaller scale methods were developed and rational conditions selected for screening. In total, seventy-seven straw accessions were processed using four conditions which allowed processing/genotype interactions to be explored. Cultivars that produced more acetic and formic acid using suboptimal pretreatment conditions obtained higher glucose yields.

Associated areas of the transcriptome suggested that genes involved in sugar nucleotide provisioning, the endogenous hydrolysis of cellulose and non-cellulosic polysaccharide synthesis were related to saccharification yields. Candidate genes aid in the development of testable hypotheses related to biomass recalcitrance and provides specific targets to conduct experiments, but molecular markers were sensitive to agronomic conditions. *Arabidopsis* plants deficient in a selection of candidate genes revealed subtle changes in saccharification performance. Result from this work could be the first steps towards deducing the genetic determinants of biomass recalcitrance, paving the way for further research.

Acknowledgements

The research in this thesis would not have been possible without the financial support of the **Integrated Biorefining Research and Technology Club** (IBTI Club - BB/H004351/1).

I would like to express my sincere appreciation to my supervisors, **Keith Waldron** and **Ian Bancroft**, for their patience, support and guidance over the past four years.

I am also enormously grateful to everyone who helped me during the course of this Ph.D. Members of the IBTI Club **Andrea Harper**, **Sam Collins** and **Klaus Wellner** were integral to the success of the project and taught me so much. Special thanks to **Charlotte Miller**, for her help and encouragement at every stage of the project.

To my lab mates, **Adam Elliston**, **Xin Zhao**, **Peter Ryden**, **Graham Moates**, **David Wilson**, **Jim Robertson, Patricia Murciano, Susana de Val and Nicola Cook**, thank you for all your help. It has been a pleasure working with you.

I would also like to thank other vital members of staff at the Institute of Food Research and the John Innes Centre, all of which contributed to aspects of this thesis. Thank you to **Henri Tapp** for statistical advice, PCR guidance from **Rachel Wells**, tissue embedding and staining knowledge from **Nicola Stacey.** Thanks to **Barry Robertson and colleagues**, particularly **Damian Alger** and **Lesley Phillips** for looking after the *Arabidopsis* plants. Thank you to **Ian Roberts** for careful proofreading of manuscripts before submission.

Many thanks to **Peter Werner and colleagues** at KWS, **Martin Broadley**, **Neil Graham** and **Rory Hayden** at the University of Nottingham, and **Mark Nightingale and colleagues** at Elsoms Seeds for growing and threshing the OSR straw used in this study.

I am also very grateful to **my parents and family** for getting me this far. Finally, thanks to **Laura Draper** for her understanding and support over the years.

SECTIONS OF THIS THESIS HAVE BEEN COMMUNICATED IN THE FOLLOWING PUBLICATIONS AND PROCEEDINGS

PUBLICATIONS

I.P. Wood, A. Elliston, P. Ryden, I. Bancroft, I.N. Roberts, K.W. Waldron.
Rapid quantification of reducing sugars in biomass hydrolysates: Improving the speed and precision of the dinitrosalicylic acid assay. Biomass and Bioenergy, Volume 44, September 2012, Pages 117-121.

I.P. Wood, A. Elliston, S.R.Collins, D.Wilson, I.Bancroft, K.W. Waldron.
Steam explosion of oilseed rape straw: establishing key determinants of saccharification efficiency. Bioresource Technology, Volume 162, June 2014, Pages 175-183.

I.P. Wood, N. Wellner, A. Elliston, S.R. Collins, D.Wilson, I.Bancroft, K.W. Waldron.
Variations in product yields relevant to cellulosic ethanol production from Brassica napus straw after pilot scale steam explosion. *In prep.*

I.P. Wood, A. Harper, C. Miller, S. Collins, I.Bancroft, K.W. Waldron.
Probing the genetic basis of biomass recalcitrance using genome wide association mapping in *Brassica napus*. *In prep.*

PROCEEDINGS

I.P. Wood, I. Bancroft, K.W. Waldron.
Bioethanol from oilseed rape straw: moving towards commercial viability by exploiting cultivar variation. Set for Britain Competition. 2014.

I.P. Wood, I. Bancroft, K.W. Waldron.
Integrated Biorefining Research and Technology Club Dissemination Events 4-8 (May 2011 to Nov 2013).

CONTRIBUTIONS

IW planned, conducted, analysed and drafted manuscripts for publication, with technical assistance from AE (running HPLC), SRC (running GC), DW (running steam explosion apparatus) and NW (PLS modelling of FT-IR spectra). KW and IB conceived and guided the studies. All co-authors reviewed manuscripts before submission. Both IW and AE conducted parts of the development of screening techniques section (Chapter 6) outlined in Wood *et al.,* (2012).

Contents

I. Abbreviations

1G	First-generation
2FA	2-Furfuraldehyde
2G	Second-generation
AGI	Arabidopsis Genome Initiative
AGP	Arabinogalactan protein
AM	Association mapping
Api	Apiose
Ara	Arabinose
ATP	Adenosine triphosphate
ATR	Attenuated Total Reflectance
BCA	Bicinchoninic acid
βG	β-glucosidase
Ca.	Calculated
CAZy	Carbohydrate-active enzyme as defined by www.cazy.com
CesA	Cellulose synthase family A protein
Chr	Chromosome
CW	Cell wall
DAP	Days after planting
DNS	Dinitrosalicylic acid
DUF	Domain of unknown function
DWB	Dry weight basis
ER	Endoplasmic reticulum
F6P	Fructose-6-Phosphate
FPU	Filter Paper Units
FWB	Fresh weight basis
FT-IR	Fourier transform infrared spectroscopy
Fuc	Fucose
Gal	Galactose
GalA	Galacturonic acid
GC	Gas chromatography
GDP-	Guanosine diphosphate
GHG	Greenhouse Gas
Glc	Glucose
GlcA	Glucuronic acid
GM	Genetically modified
GOPOD	Glucose oxidase peroxidase assay
GWA	Genome wide association
HAs	Hydroxycinnamic acids
HG	Homogalacturonan
HMF	Hydroxymethylfurfural
HPLC	High performance liquid chromatography
HT	High throughput
IBTI	Integrated Biorefining Research and Technology Club
IFR	Institute of Food Research
IRX	Irregular Xylem

IR	Infra-red
JIC	John Innes Centre
KOR	Korrigan
LD	Linkage Disequilibrium
LHW	Liquid hot water
Man	Mannose
MYA	Million years ago
NCYC	National Collection of Yeast Cultures
NDP-	Nucleotide diphosphate
NREL	National Renewable Energy Laboratory
OSR	Oilseed rape (Brassica napus)
PCR	Polymerase chain reaction
PCW	Primary cell wall
PLS	Partial least squares regression
QTL	Quantitative trait loci
REML	Restricted maximum likelihood
RG	Rhamnogalacturonan
R_0	Reaction ordinate / pretreatment severity
RS	Reducing sugar
Rha	Rhamnose
SE	Steam Explosion
SNP	Single Nucleotide Polymorphism
Sp.	Species (singular)
Spp.	Species (plural)
Stdev	Standard Deviation
HMF	5-Hydroxymethylfurfural
SCW	Secondary cell wall
SNP	Single nucleotide polymorphism
SSF	Simultaneous saccharification and fermentation
TASSEL	Trait Analysis by aSSociation, Evolution and Linkage
T-DNA	Transfer DNA
UA	Uronic acid
UDP-	Uridine diphosphate
uNASC	European Arabidopsis Stock Centre
UV	Ultraviolet
WT	Wild-type
w/w	Weight/Weight
v/v	Volume/Volume
XG	Xyloglucan
XGA	Xylogalacturonan
Xyl	Xylose

II. List of figures

III. List of tables

IV. List of equations

1 Preface

The research presented in this thesis was conducted between Oct 2010 and Oct 2014, as part of a studentship funded by the Integrated Biorefining Research and Technology Club (IBTI Club) - BB/H004351/1. This Ph.D. studentship was associated with a larger project, concentrating on the "optimization of wheat and oilseed rape straw co-products for bio-alcohol production", involving researchers based at the John Innes Centre (Ian Bancroft's Lab) and the Institute of Food Research (Keith Waldron's Lab).

The project aimed to collect phenotype data relevant to bioethanol production from a range of wheat and oilseed rape cultivars including agronomic (Andrea Harper), stem mechanical (Charlotte Miller) and straw compositional data (Samuel Collins). The information gathered in this thesis adds a fourth set of phenotype data, showing the range in straw saccharification yields after pretreatment and enzymatic hydrolysis.

The goal of this project was to use association mapping to identify areas of the *Triticum aestivum L.* and *Brassica napus* L. (*B. napus*) transcriptomes that correlated each trait (Ian Bancroft's Lab). This could not only provide potential genetic targets for biomass improvement, but also the molecular markers needed for implementation.

This thesis focuses specifically on *B. napus* straw as a resource for bioethanol production. *B. napus* straw derived from a single cultivar was initially used to identify process-dependent changes in bio-matrix opening after pilot-scale steam explosion and saccharification. Data collected using optimum pretreatment conditions highlighted key differences in cell wall (CW) chemistry likely to control sugar and ethanol yields after pilot-scale processing. After exploring the main bottlenecks to saccharification yield at a pilot-scale, attention was drawn to exploring the genetic determinants of these traits. High throughput screening methodology was developed to collect phenotype data suitable for a genome-wide association study. Rational screening conditions were selected to produce various degrees of biomatrix opening, reflecting the flexibility in processing conditions available. This study gathered phenotype information relevant to bioethanol production from a range of cultivars. The data was used to detect areas of the *B. napus* transcriptome associated with the variation in product yields between cultivars by association mapping, with the aim of exposing likely genetic determinants of traits relevant to bioethanol production. The robustness of the associated molecular markers was assessed using material gathered from a second diversity panel and a small selection of potential candidate genes explored in greater depth using *Arabidopsis* plants deficient in particular candidate genes.

2 Introduction

2.1 Shifting baselines and recent milestones: A motivation for change

"The Stone Age did not end through lack of stone..."

Sheik Ahmed Zaki Yamani

Saudi Arabian Oil Minister, 1962-86

(Maass, 2005, Abramson *et al.*, 2010)

Fossil fuels have been the bedrock of human development for over 250 years, providing the raw materials needed to produce food, fuel and fibres. However, these resources are finite. Eventually the extraction of fossil fuels will be limited by availability and not by demand. Developing alternatives to fossil fuels will be necessary to maintain the basic provisions needed for human development in the long term.

The decline in fossil fuel availability is often cited as an underlying driver for research into fossil-fuel alternatives. Global peak oil production (the point in which extraction rate is limited by availability (Hubbert, 1956)), by conventional means, has either passed (Aleklett *et al.*, 2010), or will be reached in the next 20-30 years (Brandt *et al.*, 2013). However, when factoring in improvements in oil use efficiency (Brandt *et al.*, 2013) and the exploitation of non-conventional sources of fossil fuel, such as shale gas (Howarth *et al.*, 2011), this timeframe may extend considerably, to well beyond 2100. It is therefore likely that fossil fuels will be available for significant length of time.

However, fossil fuel use is causing unprecedented deterioration of our planet and fuelling anthropogenic climate change (IPCC, 2014). Degradation of natural ecosystems beyond repair would have wide-reaching impacts on human health and livelihoods. Historically, the progressive 'shifting of baselines' from one generation to the next, has limited our perception of ecosystem degradation (Hughes *et al.*, 2013). However, this can no longer be an excuse for inaction. Even medium-term records such as the atmospheric CO_2 concentration, continually monitored since 1958 (Keeling, 2008) and recorded temperature increases since the late 1970's (Hansen *et al.*, 2006) offer sufficient evidence to establish the prevailing trend and give reason for concern. It is now clear that changes in anthropogenic greenhouse gas (GHG) emissions (including CO_2) will continue to have a largely negative affect on humans and the ecosystems that they rely on (IPCC, 2014).

Indeed, the passing of recent ecological milestones illustrate the challenges we face and gives strong motivation for research into low-carbon fossil-fuel substitutes. One such milestone happened in May 2013, when the daily average CO_2 concentration rose above 400 ppm for the first time in *circa*. 2.5 million years (long before modern humans walked the earth) and continues to increase by approximately 2 ppm every year (Jones, 2013). Evidence to suggest that fossil fuels use is the cause for this change is now overwhelming (IPCC, 2014).

Despite the negatives of using fossil fuels, curbing their use will be increasingly difficult. For example, in 2013 the world's population exceeded 7.2 billion. By 2025, an additional 1 billion people will live on earth compared to today (United Nations World Population Division, 2013). An expanding population, particularly in the developing world, will only exacerbate current pressures on resources through increased demand.

Localisation of resources causes further bottlenecks in supply. For example, a particular concern in Europe is energy security. At present, the EU countries import over half of their energy from overseas (EuroStat, 2013). Importing fuel to meet this demand, gives political advantage to countries with large fossil fuel reserves. This provides an immediate and localised political motivation for the advancement of domestically sourced fossil-fuel substitutes.

Dependency on fossil fuels means that price volatility of a single resource can cause significant economic problems (Höök & Tang, 2013). For example, increases in oil prices have preceded 11 out of 12 post-war recessions in the USA (Hamilton, 2011; Höök & Tang, 2013). Therefore, governments actively support fossil fuel producers to ensure growth and prosperity. All economically developed countries give tax incentives to fossil fuel producers (in total, $55-90 billion USD collectively/year between 2005 and 2011) (OEDC, 2012), despite them being profitable businesses.

In 2011, the United Kingdom alone gave £4251 million in tax reductions to fossil fuel producers through: (i) A reduced rate in value added tax (VAT) (£3971 million), (ii) Petroleum Revenue Tax Oil Allowance (£240 million) and, (iii) Petroleum Revenue Tax Tariff Receipts Allowance (£40 million) (OEDC, 2012). For comparison, this is approximately the same amount as given to all nine publicly funded UK research councils and the two higher education funding councils over the same period, combined (Department for Business Innovation and Skills, 2010). The large amount of money currently spent on tax incentives for fossil fuel use, could be used to incentivise greener-technologies with more long-term prospects (OEDC, 2012).

Creating products and fuels from other sources allows countries to diversify their resource portfolios, therefore minimising the impacts of short-term variations in supply brought about by fossil fuel dependency.

Governments therefore have the incentive, capacity and responsibility to invoke changes in energy policy. Legislation at various levels has been developed to promote renewable energy and to act as a catalyst to change. Legislation applicable to the UK has primarily been implemented in response to directives outlined by the European parliament (Berti & Levidow, 2014).

The main legislative drivers for bioenergy in Europe are the EU 2020-20-20 targets. By 2020, the EU aims to increase energy efficiency by 20% and reduce greenhouse gas emissions by 20% compared to 1990 levels. Moreover, the European Commission set further binding targets that 10% of transportation fuel should be sustainably produced by 2020 (Lins, 2010). It is also likely that further targets will be set, beyond 2020 (European Commision, 2012). Legislation in the form of the Renewable Energy Directive, Energy Efficiency Directive and Fuel Quality Directive support these goals (Lins, 2010). The UK government has also implemented a number of policies to incentivise investment in renewable technology. Examples include the Renewables Obligation and the Feed-in-Tariff Scheme (Department of Energy & Climate Change, 2013).

2.2 The potential of biomass: Waste as a foundation for green biotechnology

Considering the challenges faced by the world in the next century, it is clear that the production of domestically sourced, low-carbon fossil-fuel alternatives would have many benefits. Fossil fuels can potentially be replaced by renewable alternatives such as: nuclear, wind, solar, hydro, ocean, geothermal and biomass (Lee *et al.*, 2007a). It is likely that most, if not all, of these technologies will be needed to meet the demand for energy currently met by fossil fuels.

Nevertheless, in addition to the benefits common amongst renewable technologies, such as being non-depletive, release no net emissions, reduce air pollution and create employment (Lee *et al.*, 2007a), biomass-based technologies have a number of additional advantages.

Technologies using biomass are unique as they can produce both energy and carbon-based products. Sugars derived from biomass can potentially produce many industrial chemicals that are currently petroleum-based (Bozell & Petersen 2010). Producing bio-based platform chemicals that intercept petroleum-sources early on in the supply chain may facilitate an easier transition to green technologies. In this way, bio-based

chemicals may augment, replace and eventually supersede products currently derived from fossil fuels.

Biomass-based technologies also have the potential to alleviate some of the problems associated with waste production. Much of the waste associated with food production (from agricultural residues to food trimmings) is biomass-based. Valorising these materials as feedstocks for biofuel production would be an excellent way of minimising the waste produced from these systems.

The most significant impact that biomass-based industries can make to reduce fossil fuel use in the near-term is to augment conventional transportation fuels. In 2011, the equivalent of 366 million tonnes of fossil fuel was used for transportation within the EU, with the vast majority of this (> 81%) being used by road vehicles (EuroStat, 2013). This provides an exceptionally large niche for biofuel companies to exploit. Consequently, the use of fuels derived from biological sources within the EU has seen unprecedented growth over the past 5 years (+134%) (EuroStat, 2013). Even with this rapid expansion, the market is far from saturation, with only 4.66% of the EUs transportation fuels are currently derived from biological sources (EuroStat, 2013).

2.3 Producing biofuels from biomass: Technological possibilities

At least two biotechnical routes are available for the production of transportation fuels from biomass. First-generation biofuels (1G) can be produced from easily accessible forms of plant compounds – such as storage sugars and plant oils. These compounds are often found in the edible parts of the plant. Minimal processing is required to create a useable product for fuel – simply pressing oil from the seed followed by transesterification, or fermenting readily available sugars. Alternatively, starch-based plant biomass can be used by adding amylase, fermenting the liberated sugars and distilling the product.

Using these technologies, a 1G-bioethanol sector has already been established in countries such as America and Brazil. Currently, almost all of the biofuels produced worldwide are derived using 1G methodologies. Almost 200 1G bioethanol plants are currently in operation in the USA, utilising the starch in maize kernels (Advanced Ethanol Council, 2012). In Brazil, the sucrose in sugar cane stems is used to create 1G bioethanol in nearly 400 plants (Barros, 2013). Therefore, production of bioalcohol using 1G technology is already well established in some countries.

However, the creation of ethanol from 1G sources is controversial for many reasons (Gasparatos *et al.*, 2013). Common debates over the use of 1G biofuels include: (i) conflicts in resource use between food and fuel, (ii) land use change (iii) pressure on

water supplies (iv) localised environmental impacts of intensive farming and (v) 1G ethanol is unlikely to reduce CO_2 emissions and may even increase them (Gasparatos *et al.*, 2013).

The production of bioethanol from second-generation (2G) sources circumvents many of the problems typically attributed to biomass exploitation. 2G biofuels are produced from more inaccessible, but more diverse, (ligno)cellulosic materials, derived from plant CWs. Plant biomass is a very abundant and a more evenly distributed resource compared to other energy sources. Moreover of the 150-170 Gt of plant CWs produced globally each year, only 2% currently is used by humans (Pauly & Keegstra, 2008). This demonstrates the sheer volume of unexploited, or poorly exploited, biological material, which could be used to create renewable products and fuels. Potential biomass sources include dedicated biomass crops and a variety of biomass waste streams, including municipal wastes and agricultural residues. Plant biomass can be burnt, anaerobically digested or converted to 2G bioethanol to create fuel.

Producing 2G biofuels from dedicated biomass crops can avoid some problems associated with biofuels such as direct food-fuel conflicts prevent some of the negative effects of intensive farming and can help lower GHG emissions. However, they cannot address problems associated with land-use change. Moreover, large volumes of easily collected biomass are already generated alongside other activities, but not utilised at present. Therefore unused biomass waste streams could be a readily available and cheap feedstock to target. The straw of oilseed rape (OSR) is a potentially an attractive agricultural residue suitable for producing 2G bioethanol.

2.4 Targeting OSR straw: An abundant and carbohydrate-rich waste stream

OSR is a very high yielding crop (≈1908 Kg rapeseed/ha worldwide)(FAOSTAT, 2014), primarily cultivated to produce seed-oil. Rapeseed is very attractive commercially; this has led to the proliferation in its use, particularly in China and Europe. In 2012, 65 million tonnes of rapeseed was harvested worldwide (FAOSTAT, 2014). Rapeseed oil is primarily used as an edible vegetable oil, for animal feed and as a 1G biofuel, particularly in central European countries (Zabaniotou *et al.*, 2008).

Currently, only the seed is taken for use, leaving a considerable amount (≈ 70%) of aboveground biomass remaining in the straw (Diepenbrock, 2000). Despite the abundance of OSR straw, it has very few uses. As a result, the majority of OSR straw is chipped and deposited on site, as a means of disposal (Glithero *et al.*, 2013). This residual straw contains relatively high glucose (Glc) content (27-37% DWB) and therefore regarded as a good potential lignocellulosic feedstock for 2G bioethanol production (Castro *et al.*, 2011, Diaz *et al.*, 2010, Jeong *et al.*, 2010, Jeong and Oh 2011 ,

Lu *et al.*, 2009, Lu *et al.*, 2011, Mathew *et al.*, 2011a and b; Petersson *et al.*, 2007; Wi *et al.*, 2011).

Accurate estimates of OSR straw availability are difficult to obtain, as it is rarely collected (Stoddart & Watts, 2012) but it is thought that ≈ 1.38 to 2.5 million tonnes of surplus OSR straw is produced in the UK each year (Copeland & Turley, 2008; Stoddart & Watts, 2012), with the majority of the OSR straw (80-90%) being produced in England (Copeland & Turley, 2008).

The combination of abundance, lack of commercial uses and comparatively high carbohydrate content makes OSR straw an ideal target for exploitation. Deriving value from agricultural residues could not only provide an abundant, renewable resource of for industry, but also increase whole-crop value. By exploiting waste biomass sources such as OSR straw, a low-value material may be converted into valuable commodity.

2.5 Straw to ethanol: can processing overcome biomass recalcitrance?

Despite the potential of 2G bioethanol production from lignocellulosic wastes such as OSR straw, many technical challenges must be addressed before commercial exploitation can commence. Plant CWs are fundamentally difficult to process. CWs have evolved over millennia, to resist biotic and abiotic stress (Sarkar *et al.*, 2009) and are therefore, very robust and resistant to degradation.

Cellulosic ethanol production from lignocellulosic biomass typically requires four consecutive procedures: pretreatment, hydrolysis, fermentation and separation (Waldron, 2010). Hydrolysis and fermentation can be conducted separately (separate hydrolysis and fermentation, SHF), or together under compromised conditions (simultaneous saccharification and fermentation, SSF). Consolidating these stages reduces capital costs and limits product inhibition of cellulases, therefore increasing yield (Olofsson *et al.*, 2008). Consequently, it is likely that in industrial applications, saccharification and fermentation stages may occur in the same environment (SSF). Here, hydrolysis and fermentation are discussed in separate sections for clarity; however, the fundamental principles remain the same for either method.

2.5.1 **Pretreatment**

"The only step more expensive than pretreatment is no pretreatment"

(Wyman, 2007).

Biomass is incredibly difficult to hydrolyse in its native form. For example, only 10-15% of the Glc available in OSR straw can be hydrolysed using commercial cellulases alone (Petersson *et al.*, 2007; Świątek *et al.*, 2014). Therefore, pretreatment before hydrolysis is an essential step in converting CW material into fermentable sugars.

The primary purpose of pretreatment is to disrupt the web of polysaccharides, aromatic compounds and proteins found in CWs leaving them more liable to hydrolysis. This makes the overall process more efficient and cost-effective (Moxley *et al.*, 2012). Pretreatments also fulfil a range of secondary roles such as densification and sterilisation of the material before exposure to subsequent, biologically sensitive, steps.

To overcome the native resistance of lignocellulose to degradation, known as 'biomass recalcitrance' (Himmel *et al.*, 2007), pretreatments are energy intensive, making them an expensive part of the overall process (Kumar & Murthy, 2011). This has generated considerable interest in developing cost-effective pretreatment methods using a plethora of mechanical, chemical and biological techniques (Conde-Mejía *et al.*, 2011; Hendriks & Zeeman, 2009; Kumar *et al.*, 2009; Mosier *et al.*, 2005; Pedersen & Meyer, 2010; Rollin *et al.*, 2011).

Lists of the most promising pretreatments can be found in a number of comprehensive reviews (Conde-Mejía *et al.*, 2011; Hendriks & Zeeman, 2009; Kumar *et al.*, 2009; Mosier *et al.*, 2005; Rollin *et al.*, 2011). However, pretreatments can be divided into groups, by their primary mode of action: mechanical, biological, solvent-based fractionation and pH-related.

Mechanical pretreatments aim to open the structure of lignocellulose by physically reducing particle size. This increases the surface area of biomass and sheers open the CWs allowing improved access during hydrolysis. However, the energy required to generate a particle size that is sufficiently small using milling, irradiation or electric pulses is very expensive (Conde-Mejía *et al.*, 2011; Hendriks & Zeeman, 2009). Moreover, it seems that mechanical pretreatments alone are relatively ineffective at increasing Glc yields beyond ≈25% (Hendriks & Zeeman, 2009). Nevertheless, all pretreatment are likely to require some particle-size reduction before processing.

Biomass can also be pretreated using chemical agents. Chemical pretreatments use water, acids or alkalis, often combined with heat, to open the structure of lignocellulose in various ways. Some of these pretreatments also involve some degree of physical particle size reduction, they are termed 'thermo-physical' (Mosier *et al.*, 2005) or 'physio-chemical' pretreatments (Olofsson *et al.*, 2008).

Generally, chemical pretreatments use pH and temperature to open the structure of biomass (Pedersen & Meyer, 2010)(Figure 1). Acidic pretreatment conditions (lower pH) generated by mineral or organic acids promote the deploymerisation of non-cellulosic components. Similarly, pretreating biomass under neutral conditions (also known as autocatalytic pretreatment) uses water or steam without adding other chemicals. Reactive ions released via autoionisation of water upon heating, disrupts the biomass in a similar way to acid pretreatment. In this thesis, two autocatalytic pretreatment methods are used: Steam explosion (SE) and Liquid hot water (LHW), which rely on the thermal depolymerisation of biomass (Jacquet *et al.*, 2011). Under these conditions, biomatrix opening is achieved using similar mechanisms to those experienced under low pH (acidic) conditions (Pedersen & Meyer, 2010).

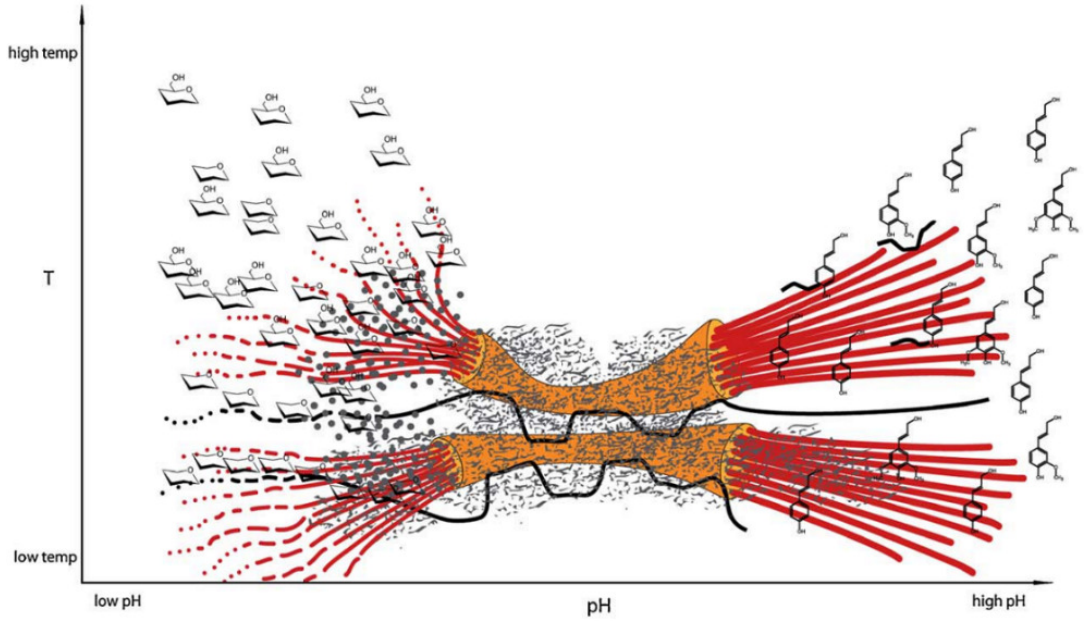


Figure 1 - Sketch illustrating how pretreatment pH and temperature results in bio-matrix opening - reproduced from Pedersen and Meyer (2010). At neutral pH, the main components of lignocellulose (cellulose, non-cellulosic carbohydrates and lignin) are organised with cellulose (orange and red) connected by non-cellulosic components (black line) perfused with lignin (grey points). Alkali conditions (high pH) favour the solubilisation of phenolic compounds, while acidic conditions primarily target carbohydrates components and cause the relocation of lignin into droplets (grey circles).

Conversely, alkali pretreatments (high pH) tend to target phenolic components, but can also cause the cleavage of glycosidic bonds and destructive alkali 'peeling' of sugars from the reducing ends of carbohydrates (Hendriks & Zeeman, 2009; Mosier *et al.*, 2005).

An important factor in determining the efficiency and cost-effectiveness of a pretreatment is the severity in which it is applied. Pretreatment severity for autocatalytic pretreatments (using heated steam or water) can be estimated using a single value called severity factor or reaction ordinate (R_o) (Overend & Chornet, 1987). The R_o of a particular pretreatment can be calculated using Equation 1, where "t" is the duration (min) and T is the temperature (°C) of the pretreatment. The value "14.75" is the activation energy constant – assuming biomass pretreatment follows first-order reaction kinetics (i.e. rate of depolymerisation is linear to reactant concentrations) (Overend & Chornet, 1987). The reaction ordinates combines pretreatment temperature and time into a single value (Figure 2).

$$R_0 = t * \exp [(T - 100)/14.75]$$

Equation 1

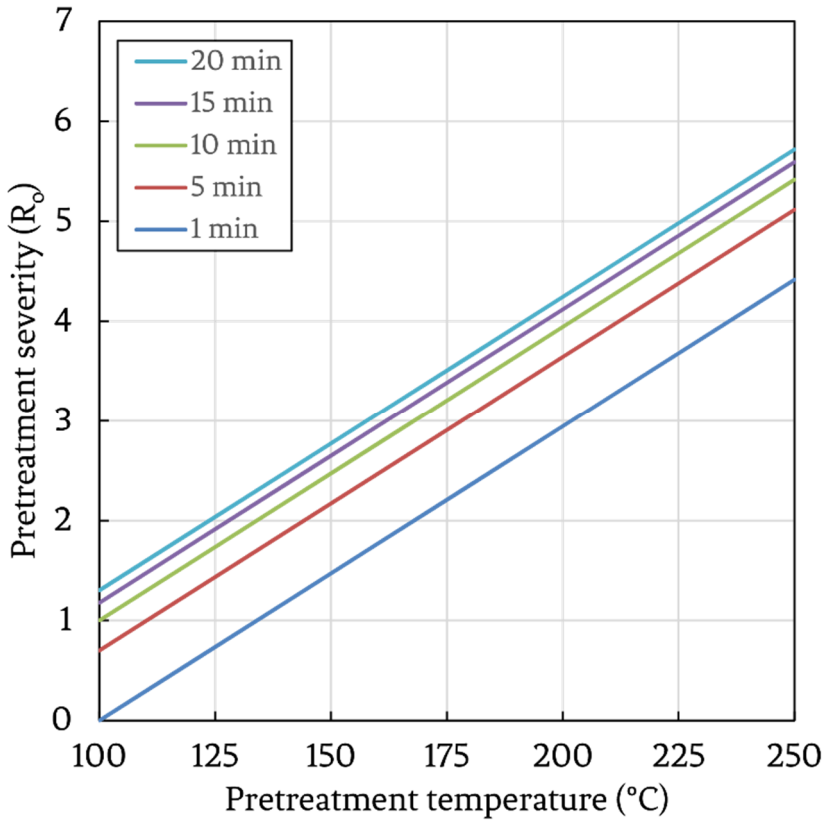


Figure 2 - Relationship between pretreatment severity (R_o), temperature and time.

Pretreating biomass under severe conditions (high R_o) often results in partial decomposition of potentially fermentable sugars into compounds which inhibit downstream processes (Figure 3). Common inhibitory compounds released from biomass during autocatalytic pretreatments include furfural derivatives, such as furfuraldehyde (2FA) and hydroxymethylfurfural (HMF) and aliphatic acids formic, acetic and levulinic acid. Some phenolic compounds are also released from lignin (Jonsson *et al.*, 2013; Rasmussen *et al.*, 2013). The dehydration of pentose and hexose sugars (typically from non-cellulosic components) under acidic pretreatment conditions results in the formation of 2FA and HMF respectively. HMF and acetic acid is formed from acetyl-groups found on non-cellulosic polysaccharides and lignin (Figure 3).

These compounds can inhibit both cellulases and fermentative organisms to varying extents (Palmqvist & Hahn-Hagerdal, 2000; Rasmussen *et al.*, 2013). These organic acids are also assumed to catalyse depolymerisation further, by creating a slightly acidic environment (Hendriks & Zeeman, 2009; Mosier *et al.*, 2005), but empirical evidence for this is difficult to obtain.

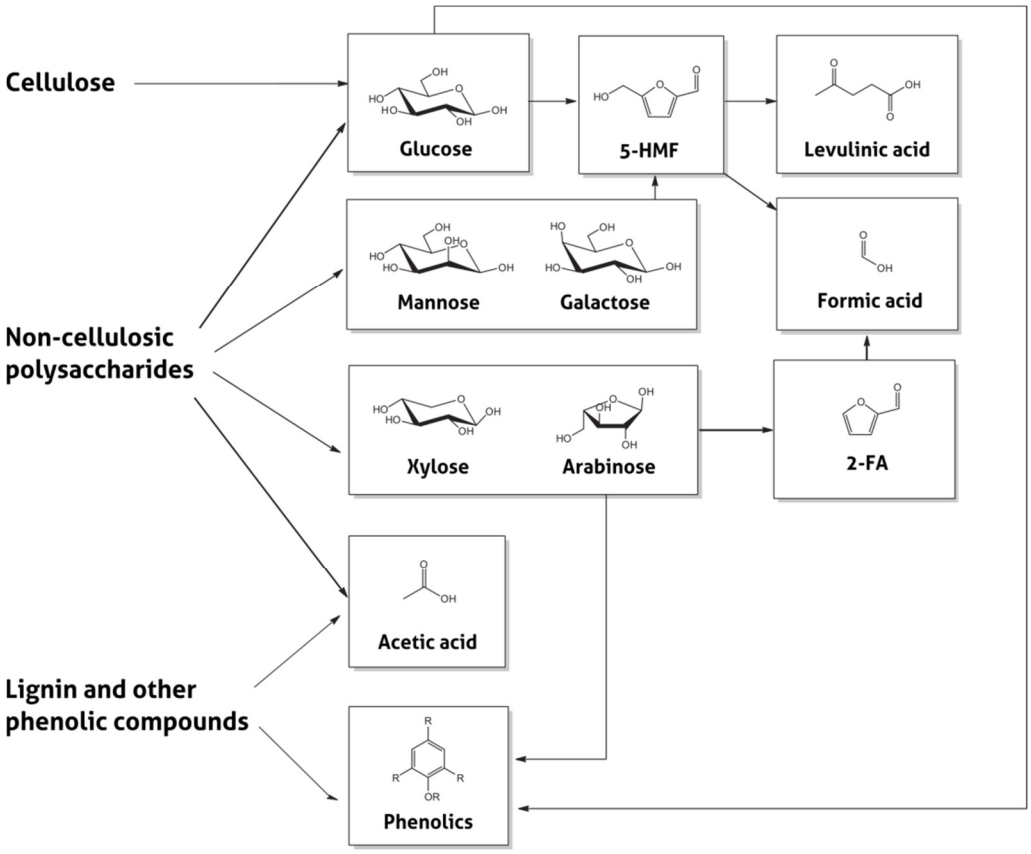


Figure 3 - Formation routes of common fermentation inhibitors during autocatalytic pretreatment redrawn from (Rasmussen *et al.*, 2013), using (Jonsson *et al.*, 2013).

Another consideration at the pretreatment stage is the effect of drying. Pretreated materials become more resistant to enzymatic hydrolysis when dried, possibly caused by pore collapse and increased hydrogen bonding between components (Zhao *et al.*, 2012). Therefore, biomass should remain hydrated between pretreatment and enzymatic hydrolysis to prevent the partial reversal of pretreatment.

Many pretreatment techniques have been explored using OSR straw as a feedstock including: wet oxidation (Arvaniti *et al.*, 2012; Petersson *et al.*, 2007), hydrothermal (Díaz *et al.*, 2010; López-Linares *et al.*, 2013), dilute acid (Castro *et al.*, 2011; Lu *et al.*, 2009; Mathew *et al.*, 2011b), microwave assisted dilute acid (Lu *et al.*, 2011), alkali (Mathew *et al.*, 2011a; Mathew *et al.*, 2014; Świątek *et al.*, 2014) and autocatalytic 'popping' (Wi *et al.*, 2011) and SE (Ryden *et al.*, 2014).

Assigning the relative effectiveness of these pretreatments would be useful to assess which pretreatment method would be best. However, insufficient evidence is available to do this. For example, the efficiency of different pretreatments may only be judged in the context of a whole-system, using commercial facilities which do not currently exist (Wyman, 2007).

Comparisons are often made using process conditions that are considered 'optimal', in terms of Glc yield. But it is unlikely that absolute maximum sugar yields will be optimal in terms of energy and costs, but instead be subject to the law of diminishing returns. Noting the effect of process permutations is therefore essential to gain comparable results that are relevant to application. This subject will be discussed in greater detail in Chapter 4, as motivation for a more comprehensive evaluation of the key determinants of saccharification efficiency of OSR straw, beyond final saccharification yield.

2.5.2 Saccharification

Further hydrolysis of pretreated materials is required to gain monomeric sugars suitable for fermentation. Monomeric sugars can be released from biomass via chemical (acid hydrolysis) or biochemical (cellulolytic enzymes) processes.

Acid hydrolysis is an effective way of breaking the glycosidic bonds that hold polysaccharides together. In fact, this method (72% H_2SO_4, followed by dilution to circa. 3% H_2SO_4 and heating to ≈100°C) is a common method used to liberate monosaccharides from biomass in the laboratory (Saeman *et al.*, 1945, Sluiter, 2004). However, acid hydrolysis is not ideal for commercial scale processing as it is relatively expensive, uses less environmentally friendly chemicals and requires high

temperatures. Consequently, the capital costs associated with acid hydrolysis are typically higher than enzymatic methods (Wyman, 2007).

Enzymatic hydrolysis is potentially a safer, cheaper and more efficient way of converting biomass to sugars. Cellulases can be produced from a range of bacterial and fungi (Lopez Ferreria et al., 2014). The most widely used cellulases found in commercial enzyme cocktails are derived from thermophilic fungal species, such as *Trichoderma reesei* (Table 1)(Lopez Ferreria *et al.*, 2014).

Cellulase name	**Company**	**Organism**
Cellulase AP30K	Amano Enzyme	*A. niger*
Novozymes 188	Novozymes	*A. niger*
Biocellulase TRI	Quest Intl.	*A. niger*
Alternafuel® 100P	Dyadic	*T. longibrachiatum*
Cellubrix (Celluclast)	Novozymes	*T. longibrachiatum & A. niger*
GC220	Danisco-Genencor	*T. reesei*
Accellerase® Series	Danisco-Genencor	*T. reesei*
Multifect CL	Genencor	*T. reesei*
Cellic® CTec Series	Novozymes	*T. reesei*
Celluclast® 1.5 L	Novozymes	*T. reesei*
Cellulyve	50L Lyven	*T. reesei* / *T. longibrachiatum*
Econase CE	Alko-EDC	*T. reesei* / *T. longibrachiatum*
Spezyme	Danisco-Genencor	*T. reesei* / *T. longibrachiatum*
Viscostar 150L	Dyadic	*T. reesei* / *T. longibrachiatum*
Cellulase 2000L	Rhodia-Danisco	*T. reesei* / *T. longibrachiatum*
Rohament CL	Rohm-AB Enzymes	*T. reesei* / *T. longibrachiatum*
Cellulase TRL	Solvay Enzymes	*T. reesei* / *T. longibrachiatum*
Ultra-Low Microbial	Iogen	*T. reesei* / *T. longibrachiatum*
Cellulase TAP106	Amano Enzyme	*T. viride*

Table 1 – Commercial cellulase cocktails used to produce 2G biofuels modified from (Lopez Ferreria *et al.*, 2014) and (Pandey et al., 2011), demonstrating the prevalence of *Trichoderma spp.* – and particularly *T. reesei* – in commercial cellulase preparations.

Commercial cellulases are generally multicomponent enzyme cocktails, derived from various modified strains of *T. reesei.* These fungi are likely secrete at least 300 different carbohydrate active (CAZy) enzymes (Lopez Ferreria *et al.*, 2014). However, at least three principle cellulase classes are required to hydrolyse biomass, namely: Endoglucanase, Cellobiohydrolase and β-glucosidase (βG). Endoglucanases dissect soluble glucan strands into smaller chains. Cellobiohydrolases remove disaccharides of cellobiose progressively from reducing (Cellobiohydrolase I) or non-reducing (Cellobiohydrolase II) ends. These disaccharides can be broken into monomeric Glc by βG. Cellulase proteins secreted by *T. reesei* are mainly Endoglucanases and Cellobiohydrolases (58-99% by weight) (Pandey *et al.*, 2011).

In addition to the core cellulases, "accessory" enzymes acting on non-cellulosic components can be added to improve performance (Banerjee *et al.*, 2010). These accessory enzymes include pectinases, xylanases, ligninases and ferulic acid esterases. Accessory enzymes act in synergy with cellulases to improve saccharification performance (Hu *et al.*, 2011), particularly when hydrolysing complex substrates. Optimisation of cellulase cocktails is an ongoing process (Banerjee *et al.*, 2010) and likely to require some substrate-specific tailoring to hydrolyse biomass with optimum efficiency (Hu *et al.*, 2013).

Just as pretreatment severity factor describes the intensity and therefore cost of pretreatment, the strength of a cellulase cocktail can be estimated by its cellulase activity. Cellulase activity can be quantified in many ways, with the most common being the filter paper unit (FPU) (Dashtban *et al.*, 2010). The FPU unit is based on the amount of cellulase required to hydrolyse 2 mg of reducing sugar (RS), as Glc from a standardised piece (1 x 6 cm) of Whatman 1 filter paper (a ubiquitous source of cellulose fibres) in 60 min using 500µL of diluted enzyme (Ghose, 1987). Each FPU produces one nmol of RS per minute. This gives an approximation of activities between cellulase batches, but the relationship between activity and dilution is not linear. Moreover, FPUs do not take into account the activity of accessory enzymes that act on non-cellulosic components.

Cellulase cocktail efficiency is ultimately dependent on substrate composition and cannot be predicted using standard assays of cellulase activity (Kabel *et al.*, 2006). Together, CW polymers interact to form an impenetrable network, resistant to attack. For example, cellulose hydrolysis rate is limited by the initial stages of amorphogenesis, as new areas of crystalline cellulose take time to solubilise. This is the primary reason why enzymatic hydrolysis is the rate-limiting step to bioethanol production (Arantes & Saddler, 2010). However, non-cellulosic components can also physically limit hydrolysis, by creating a biomatrix with a nanostructure that is too

small for enzymes to access (pore size < 0.8 nm) (Adani *et al.*, 2011). Similarly, after pretreatment, phenolic droplets deposited on the surface of biomass can impede the progressive movement of cellulases (Selig *et al.*, 2007), causing molecular "traffic jams" on the surface of microfibrils, limiting hydrolysis further (Igarashi *et al.*, 2011). In all these cases, cellulase-substrate interactions are the main limitations to enzymatic hydrolysis. Therefore, properties of the biomatrix after pretreatment are crucial to the performance of cellulases at the hydrolysis stage.

2.5.3 **Fermentation**

Sugars released during saccharification can be used as a 'platform chemicals' to produce a substrate for the microbial production of a range of chemicals, including fuel alcohols. Common organisms used to produce fuel ethanol are yeasts and bacteria species (Lin & Tanaka, 2006). The most common species yeast species used in producing ethanol is baker's yeast (*Saccharomyces cerevisiae*), as it is easy to culture and tolerant to a large range of conditions (Lin & Tanaka, 2006).

Yeast is also very responsive to selection pressures imposed in industrial environments so can be tailored to suit various process conditions (adaptive evolution) (Basso *et al.*, 2008). Having said that, yeast strains collected from extreme environments and have been exposed to particularly strong selection pressures in nature, have the ability to outperform strains bred specifically for ethanol production, in terms of temperature and fermentation inhibitor tolerances (Favaro *et al.*, 2014). This example illustrates that much of the genetic diversity required to make the bioethanol process more commercially competitive may already be present in many species. This is also why particularly robust naturally occurring, red-wine yeast, selected from the National Collection of Yeast Cultures, was used in this study.

A number of studies have fermented hydrolysates produced from OSR straw using *Saccharomyces cerevisiae* (Arvaniti *et al.*, 2012; Lopez-Linares *et al.*, 2013a and b; Mathew *et al.*, 2013; Mathew *et al.*, 2014; Thygesen *et al.*, 2012). Results from these studies would suggest that SSF of pretreated OSR straw at substrate concentrations of < 17% DM, will not be affected by fermentation inhibitors, as long a sufficient yeast inoculum is used (Thygesen *et al.*, 2012). However, in order to obtain ethanol concentrations suitable for fermentation (> 10% ethanol w/w) (see following section), hexose sugar concentrations in the beer must exceed ≈ 19.6%. Therefore substrate concentrations in excess of 40% DM (w/w) may be required (assuming substrate ≈50% matter of pretreatment material is fermentable). Inhibitory compounds in biomass at this concentration could have a negative effect on fermentation.

2.5.4 **Separation**

Fermentation of cellulosic sugars produces an alcoholic beer. Beer produced from fermentation typically contains 5-12% ethanol (w/w) that needs dehydrating to anhydrous ethanol (95-100% ethanol, w/w) to use as fuel (Huang *et al.*, 2010). To do this, the beer is first distilled using fractional distillation to ≈92.4% (w/w). Above 95.6% (w/w), ethanol and water form an azeotrope (the concentration of the two solvents in vapour is the same) making it impossible to concentrate the liquid further (Figure 4)(Katzen *et al.*, 1999). Above his concentration, azeotropic distillation is required. Another, more volatile solvent is added to the mixture (entrainer), forming a new lower-boiling azeotrope between ethanol and the main solvent. Ethanol can be concentrated further to > 99% and the solvent recycled (Huang *et al.*, 2010).

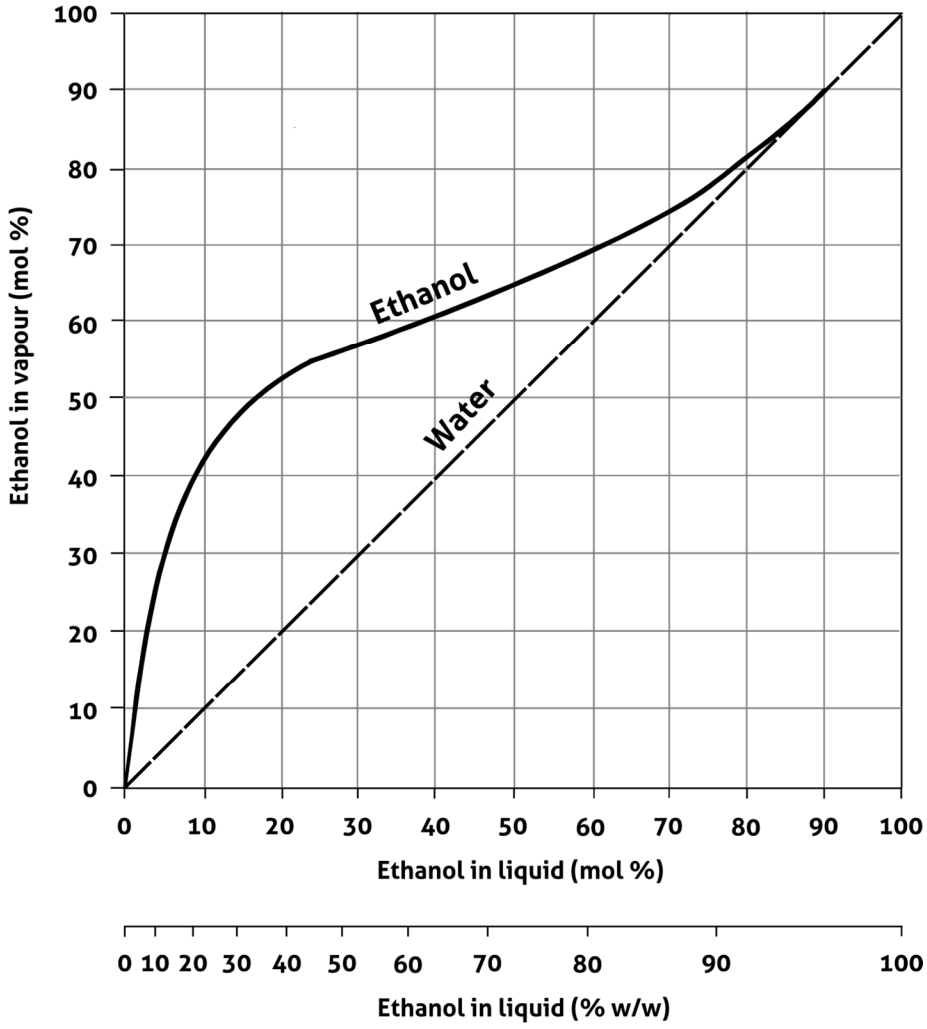


Figure 4 - Vapour/Liquid equilibrium of ethanol during distilation – drawn using data from Katzen *et al.*, (2003), adding an additional axis to show the concentration of ethanol in solution was added to express the concentrations in (w/w).

Distillation is very energy intensive, therefore significant cost reductions can be made by producing higher concentration beers. To achieve this earlier stages of the process must be improved to allow at higher solids content (Elliston *et al.*, 2013). Fermentation inhibitors produced during pretreatment will also be concentrated under these conditions and could be sufficiently high to reduce yields. This may require the improvement of pretreatments to minimise inhibitor release, tolerance of fermentative organisms, and altering feedstock composition to release fewer inhibitors.

2.5.5 Commercialisation of cellulosic ethanol

As can be seen above, the fundamental methodology for producing cellulosic bioethanol is well established, however further optimisation is required before commercial exploitation can commence. The majority of 2G ethanol plants currently in operation are limited to pilot or demonstration scale facilities (Balan *et al.*, 2013). There are few operational commercial-scale 2G ethanol plants (Table 2).

The limited exploitation of 2G bioethanol at commercial scales is an indication of the difficulty, or impossibility, of efficiently producing a product from a feedstock that has not been bred specifically for that purpose (Gressel, 2008). One of the best ways to improve profit margins, minimise operating costs and increase yields is to reduce the natural resistance of lignocellulose to saccharification (Wyman, 2007). Breeding crops to yield more sugar from their straw could be an excellent way of improving the cost-effectiveness of the process. Assessing the interaction between genetic origin and process conditions is also an important consideration for industry. However optimising a feedstock for a process that has not been established is difficult. A rational process must therefore be determined first so that key bottlenecks caused by the substrate can be assessed.

Company	Feedstock	Location	(MG/Y)	Start
Beta-renewables	Agri-residues & energy crops	Italy	20	2012
INEOS Bio	Vegetative & municipal waste	Florida, USA	8	2013
Granbio	Sugarcane straw	Alagoas, Brazil	21	2014
Fiberight	Municipal waste	Iowa, USA	6	2014
Enerkem	Municipal waste	Alberta, Canada	10	2014
POET-DSM	Corn stover residues	Iowa, USA	20	2014
Abengoa	Agri-residues & energy crops	Kansas, USA	25	2014
DuPont	Corn stover residues	Iowa, USA	30	2014
Fulcrum Bioenergy	Municipal waste	Nevada, USA	10	2015

Table 2 – Examples of commercial cellulosic ethanol plants currently in production or will be online imminently (> 5MG/Y, before 2015). Collecting information from current industry reports (Advanced Ethanol Council, 2012; Peplow, 2014) and (Balan *et al.*, 2013) - updated using company websites.

2.6 Composition of lignocellulose: the fundamentals of biomass recalcitrance

With a renewed interest in exploiting lignocellulose, resurgence in CW research for this purpose has been developed. As the field has grown, logical divisions have formed into applied and fundamental research areas (Lopez-Casado *et al.*, 2008). Generally, applied research views lignocellulose as a fixed, heterogeneous substrate and is concerned with deconstructing it efficiently. Conversely, fundamental research focusses on how CWs are constructed and function in living plant tissues (Lopez-Casado *et al.*, 2008). Crop breeding for the purpose of bioethanol production requires an appreciation of both these disciplines. Moreover, understanding how CWs come to form the final harvestable material is critically important for our understanding of biomass recalcitrance (Himmel *et al.*, 2007).

Therefore, the following sections give an overview of the structure and synthesis of plant CWs which make them resistant to degradation. In doing so, the potential genetic determinants of CW synthesis must be mentioned. In addition, outlining the known genetic determinants of CW synthesis will aid the interpretation of data presented in later chapters.

This section does not attempt to provide and exhaustive list of genes related to CW synthesis. Not only would such a list be impossible to write, as it would contain nearly 2700 genes, of which only ≈350 are annotated (Wang *et al.*, 2012), it would also be beyond the scope of this thesis. Moreover, a review if this kind would only add to the very many speculative reviews of this nature already in circulation (Abramson *et al.*, 2010, Bosch & Hazen, 2013, Burton & Fincher, 2014, Carpita, 2012, Farrokhi *et al.*, 2006, Himmel *et al.*, 2007, Pauly & Keegstra, 2008, Phitsuwan *et al.*, 2013, Yang *et al.*, 2013).

Instead, this section aims to give sufficient background on the composition and synthesis of plant CWs focussing, wherever possible, on information specifically relevant to the CWs of *B. napus* or close relatives. At the same time, potentially interesting bottlenecks to CW synthesis will be highlighted. Recent advances in our understanding of plant CW synthesis and manipulation will also be mentioned that could have important consequences for the application of this work.

The potential role of specific genes in determining biomass recalcitrance is reserved for later sections (Chapter 7), where speculation will be limited by either by genes located in close proximity to molecular markers associated with traits relevant to bioethanol production.

2.6.1 Composition of lignocellulose

Biomass contains around 35-40 cell types, each with different CW compositions tailored to specialised roles in the living plant (Cosgrove, 2005; Gorshkova *et al.*, 2013). However, all CWs begin as thin (0.1 µm), extendable, yet incredibly strong primary CWs (PCWs). In growing tissues, PCWs allow controlled cell expansion while maintaining cell adhesion and withstanding circa 2500 bar of internal turgor pressure (Carpita & Gibeaut, 1993). These selection pressures perhaps explain why plant CWs are so remarkably robust. When expansion ceases, a thicker secondary CW (SCW) can be deposited to give additional strength in selected tissues.

CW material contains an abundance of structural carbohydrates, including cellulose (41-51% total carbohydrate), non-cellulosic β-1,4-D-glycans (29-37% total carbohydrate) and small amounts of pectin. In certain tissues, these polysaccharides are embedded in an aromatic complex known as lignin (13-28% w/w). Lignin provides additional strength and waterproofing to these cells. Minor constituents such as protein (3-10%), lipids (1-5%) and other inorganic minerals (5-10% w/w) are also present (Pauly and Keegstra, 2008). These constituents combine to create a complex, heterogeneous matrix of polymers that has evolved to maintain the rigidity of plant cells (Figure 5) (Burton *et al.*, 2010).

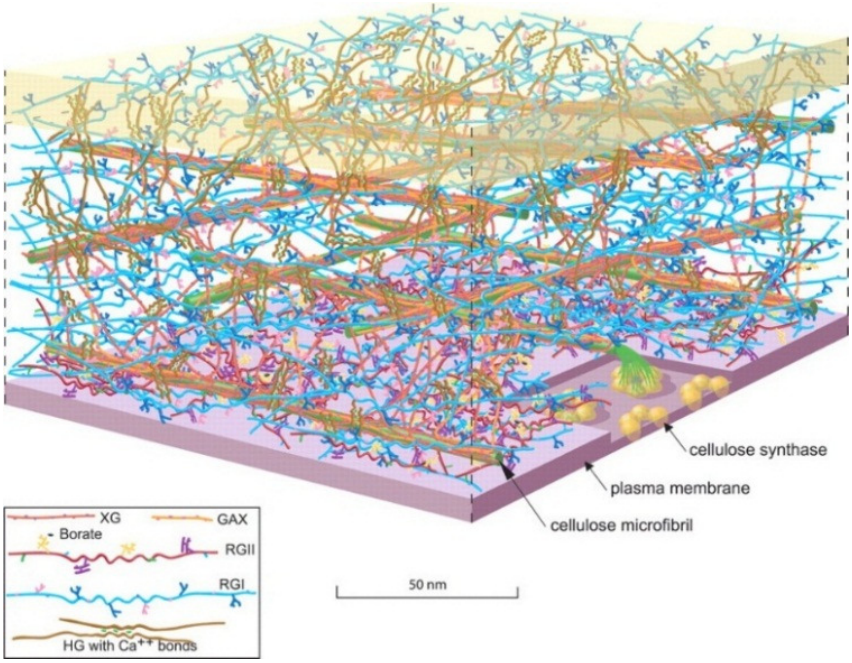


Figure 5 - A scale schematic of a dicot (*Arabidopsis*) primary cell wall taken from (Somerville *et al.*, 2004). For clarity, the abundance of cellulose is reduced. In real CWs, microfibrils are much denser, forming successive layers of thickening. The middle lamella is the uppermost region shaded orange. This is where pectin abundance is most concentrated and lignin deposition begins (Wi *et al.*, 2005).

Flowering plant species produce two CW classes which are both chemically (Carpita & Gibeaut, 1993) and genetically distinct (Yokoyama & Nishitani, 2004). Dicotyledonous (such as *B. napus*) and non-commelinoid monocotyledonous plant species produce "type I" CWs. Type I PCWs are typically rich in xyloglucans (XGs) and pectic polysaccharides. Their SCWs mainly comprise of glucuronoxylan (20-30%, w/w) with small amounts of glucomannan (2-5%, w/w), galactoglucomannan (0-3%) and traces of XG (Scheller & Ulvskov, 2010). Non-commelinoid monocotyledonous plants such as wheat, rice and barley have type II CWs. These contain less pectin and XG, but produce different polymers including glucuronoarabinoxylan and mixed-linked glycans. Type II CWs also form more phenolic-acid (particularly ferulic acid) cross-linking than type I walls (Yokoyama & Nishitani, 2004).

As *B. napus* is a dicotyledonous plant, its CWs are quite different in composition to most agricultural residues derived from cereal species. *B. napus* straw contains the following constituents, in approximately these proportions: Glc (33%), lignin (19%), Xyl (16%), UA (7%), acetyl groups (4%), ash (3%), Man (2%), protein (2%), Gal (2%), lipids (0.8%), Ara (0.8%) and other components (10%) DWB (Pronyk and Mazza 2012). Therefore, *B. napus* straw contains approximately three times more UA and is more acetylated than CW material derived from other (monocotyledonous) agricultural residues. Conversely the abundance of Ara in OSR straw is much lower (Pronyk & Mazza, 2012). OSR straw also contains a greater mixture of PCW and SCW materials derived from the rind and pith tissues.

2.6.2 Nucleotide-sugars: the building blocks of structural polysaccharides

CW synthesis begins with the production of nucleotide sugars (NDP-sugars); the building blocks for structural carbohydrate synthesis. Therefore, they are important determinants of CW functioning. Consequently, a considerable amount of research has aimed to understand the formation of these chemicals (Bonawitz & Chapple, 2010; Rennie & Scheller, 2014; Seifert, 2004; York & O'Neill, 2008).

Carbon fixed during photosynthesis is released from the Calvin-Benson–Bassham cycle as surplus glyceraldehyde-3-phosphate (G3P), which is initially converted to fructose-6-phosphate (F6P). The main precursors of all CW sugars, UDP-Glc and GDP-Man, are both synthesised from F6P (Figure 6). UDP-Gal, UDP-Rha and UDP-GlcA are produced from UDP-Glc. UDP-GlcA serves as a precursor for UDP-Xyl, UDP-Ara, UDP-GalA and UDP-Api synthesis. GDP-Man produced from F6P can give rise to GDP-Fuc and GDP-Gal (Figure 6). Alternatively, sucrose is synthesised from surplus F6P in photosynthesizing tissues and transported in the phloem to sugar-deficient cells (Bonawitz & Chapple, 2010; Rennie & Scheller, 2014; Seifert, 2004; York & O'Neill, 2008) (Figure 6).

Manipulation of key NDP-sugar genes has been shown to alter CW compositions (Seifert *et al.*, 2002), but it is not currently known to what extent NDP-sugars regulate plant glycan composition or abundances (Bar-Peled & O'Neill, 2011). Nevertheless, natural bottlenecks in UDP-sugar synthesis and interconversion exist as unidirectional conversions between NDP-sugar pools. These are likely to form natural control-points, where NDP-sugar abundances can be channelled to produce different polymers. For example, in order to produce Xyl-containing carbohydrates from photosynthetic carbon, at least two control points must be crossed - (1) UDP-Glc → UDP-GlcA and (2) UDP-Glc → UDP-Xyl (Figure 6). Enzymes at these junctions (in this case, UDP-Glc dehydrogenase and UDP-GlcA decarboxylase) are therefore likely to play important roles in the synthesis of Xyl and Ara, and presumably the polymers they produce. Moreover, recovery pathways converting sugar monomers back into usable sugar-nucleotides may be particularly important in sugar-deficient tissues.

NDP-sugars can be synthesised in both the cytosol and Golgi apparatus. This presents a further opportunity for regulation, by compartmentalising NDP-sugars between membranes (Bar-Peled & O'Neill, 2011). Although very few membrane-bound nucleotide sugar transporters have been categorised (Rautengarten *et al.*, 2014), they could play a particularly important role in making NDP-sugars available for the assembly of structural carbohydrates present in plant CWs.

In summary, the synthesis, transport and localisation of NDP-sugars provide multiple control-points that could regulate CW synthesis at the earliest stages. However, to what extent and under what circumstances these control points are used is not fully understood (Bar-Peled & O'Neill, 2011).

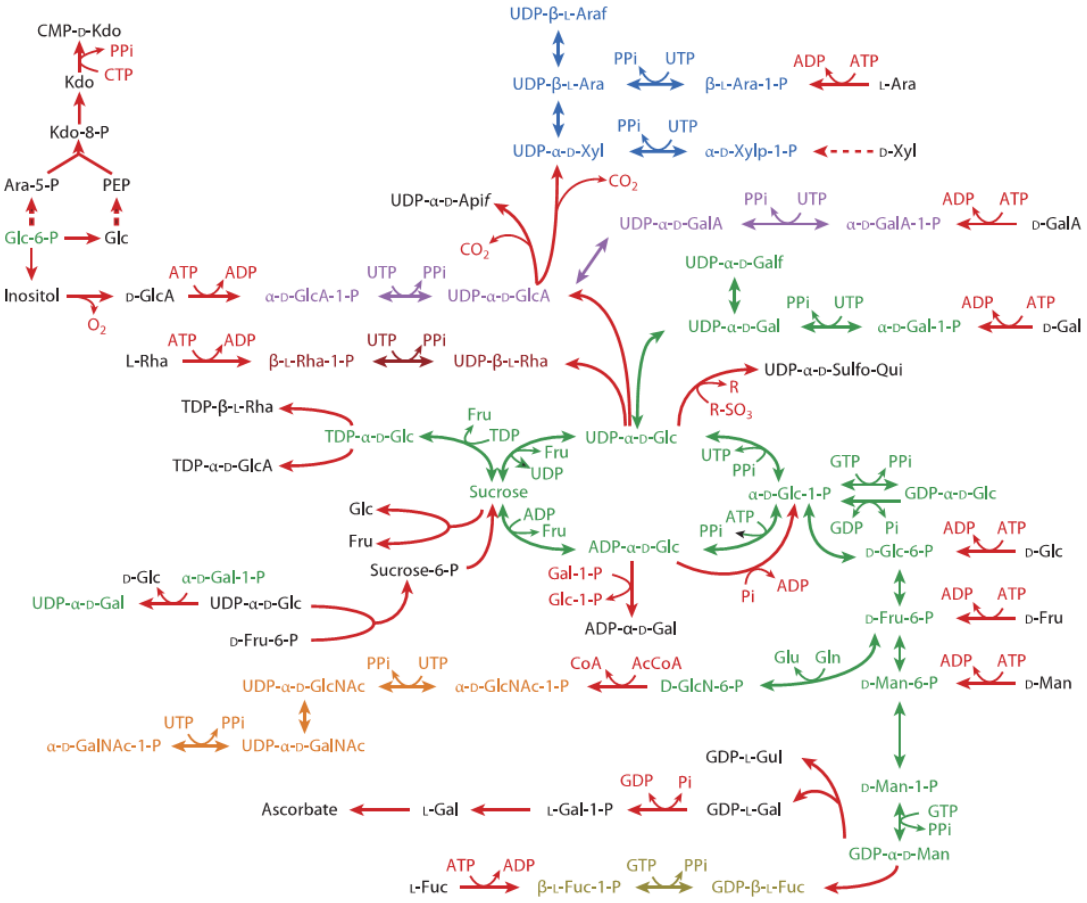


Figure 6 - An overview of the most common biosynthesis and salvage pathways for nucleotide sugars – the building blocks of CW carbohydrates - modified from Bar-Peled and O'Neill (2011). Colour has been added to the figure to identify key control points (red) allowing unidirectional conversion of sugar nucleotides between pools. The main pools synthesising various sugar nucleotides are noted in different colours: Glc, Gal, Man and glucosamine containing compounds (green); UA containing compounds (purple); Xyl and Ara containing compounds (blue); N-acetylated sugar amines (orange); Fuc containing compounds (khaki); and Rha containing compounds (brown). Dashed lines indicate three unresolved unidirectional pathways, namely: the salvage of D-Xyl to α-D-Xylp-1-P, the conversion of Glc-6-P to Ara-5-P and the conversion of Glc to PEP.

2.6.3 Cellulose structure and biosynthesis

Cellulose is the most abundant biological polymer on earth and the main load-bearing polymer in all plant CWs. As the main repository for Glc in the CW and the largest pool of readily fermentable sugar, cellulose is also a major target for industrial biotechnology. Understanding its structure and synthesis is therefore essential for those interested in the resistance of lignocellulose to degradation (Carpita, 2011; Endler & Persson, 2011; Li *et al.*, 2014; McFarlane *et al.*, 2014; Slabaugh *et al.*, 2014; Somerville, 2006).

Cellulose is a long, unbranched, linear polymer of β-1,4-linked Glc. When synthesised, these singular cellulose strands, aggregate with parallel chain producing a crystalline structure held together by hydrogen bonds between adjacent hydroxyl groups (Carpita, 2011). Each cellulose strand has a flat ribbon-like secondary structure, allowing adjacent strands to stack on top of each other, producing a crystalline microfibril (Gorshkova *et al.*, 2013). These microfibrils aggregate further to produce larger structures varying in size from 5 to 50 nm in diameter (Li *et al.*, 2014). Differences in hydrogen bonding patterns results in two forms of cellulose found in nature: cellulose Iα and Iβ. The crystalline β-form is most abundant in higher plants (Brett & Waldron, 1998; Sturcova *et al.*, 2004). Amorphous regions found within and at the surface of the microfibril can be enzymatically hydrolysed > 50 times faster than highly crystalline areas (Arantes & Saddler, 2011). Therefore, cellulose crystallinity has implications, not only to the how cellulose functions in the growing plant, but also to those wishing to exploit lignocellulose.

Cellulose is synthesised from UDP-Glc by specific glycosyltransferases embedded in the cell membrane known as cellulose synthase family A (CesA) proteins. CesA proteins arranged in hexagonal 'rosette' formations (Carpita, 2011), allowing the simultaneous synthesis of glucan chains in close proximity to each other, thereby promoting aggregation (Figure 7).

Although the core genes encoding the main components of the cellulose synthase complex have been identified, less is known about the regulation of key cellulose properties such as cellulose crystallinity, microfibril dimensions and orientation (Li *et al.*, 2014; McFarlane *et al.*, 2014). Identifying these genes would not only extend our biological understanding of cellulose synthesis, but also allow us to modify the properties of cellulose for biotechnology. Proteins currently thought to alter cellulose properties include members of the Korrigan (KOR), CesA-interactive protein (CSI), tracheary element differentiation-related (TED), chitinase-like (CTL), *COBRA* (COB),

COBRA-like (COBL), *POM-POM* and *KOBITO* gene families (Carpita, 2011; Endler & Persson, 2011; Li *et al.*, 2014; McFarlane *et al.*, 2014).

Other CW components are also likely to alter key properties of cellulose, through indirect interactions. Cellulose microfibrils are deposited into the CW either in a random or directional orientation. Non-cellulosic components already present in the CW tether adjacent microfibrils, generally restricting lateral movement. During expansion, cell turgor pressure is focussed longitudinally and the cell can elongate. As the cell elongates the microfibrils align through a process of 'passive reorientation' (Carpita & Gibeaut, 1993). Therefore, genes controlling the properties of non-cellulosic polysaccharides are also likely to alter microfibril properties.

For example, pectin methyl-esterification has profound effects on cellulose biosynthesis (Bischoff *et al.*, 2010). Likewise, acetylation of non-cellulosic polysaccharides causes changes to cellulose orientation, as observed by changes in trichome appearance under polarised light (Trichome Birefringence) (Bischoff *et al.*, 2010; Gille & Pauly, 2012b). Therefore, the compositions and synthesis of other non-cellulosic polysaccharides are also important in determining properties of the CW.

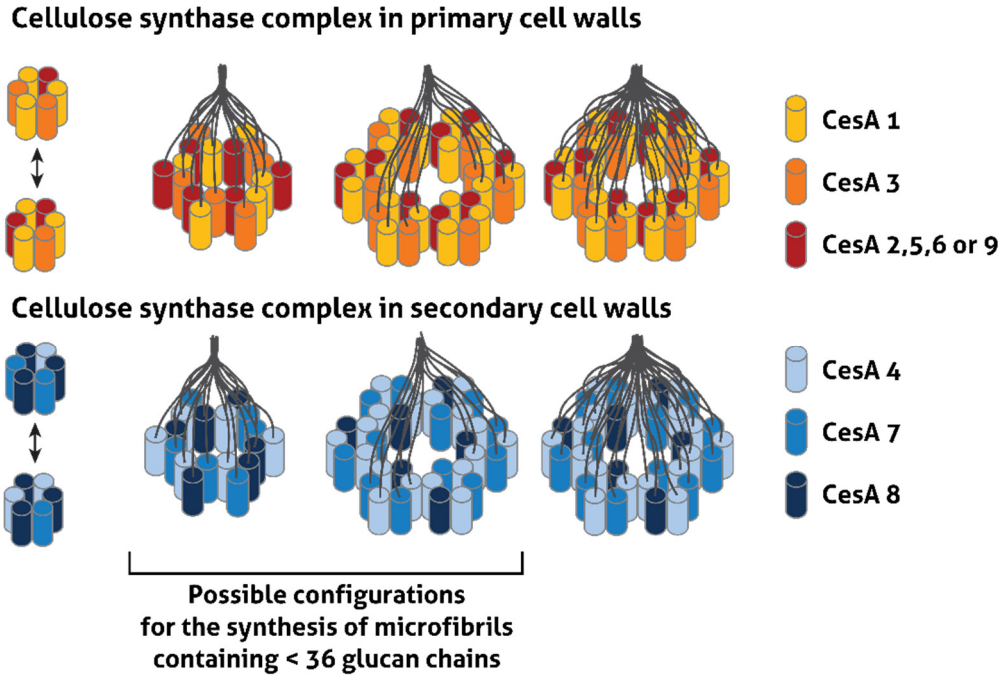


Figure 7 - Likely configurations for cellulose synthase complexes found in *Arabidopsis* primary and secondary cell walls - modified from McFarlane *et al.*, (2014). Cellulase synthase proteins form 'rosette' complexes that can synthesise microfibrils containing up to 36 glucan chains. The size of the cellulose synthase complex is not currently known (Endler & Persson, 2011; Li *et al.*, 2014). Microfibril size can differ, dependent on cell type. Whether this is achieved using complexes of different size, or that some CesA proteins do not contribute to the microfibril is not known (McFarlane *et al.*, 2014).

2.6.4 Non-cellulosic β-1,4-D-glycan structure and biosynthesis

Traditionally known as hemicelluloses, based on their extractability using strong chaotropic agents (disrupt hydrogen bonding) such as alkali (Pauly *et al.*, 2013), non-cellulosic β-1,4-D-glycans are a heterogeneous group of polysaccharides. Non-cellulosic β-1,4-D-glycans found in dicot PCWs are mostly XG (20-25% total CW), with small amounts of glucuronoarabinoxylan and glucomannan (5% and 3-5% respectively). By contrast, their SCWs contain mostly glucuronoxylan (20-30%), some gluco- and galactogluco-mannans (2-5% and 0-3% respectively) and small amounts of XG (Scheller & Ulvskov, 2010).

Non-cellulosic glycans are important as they act as a keystone polymer, tethering adjacent cellulose microfibrils, lignin and pectin (Pauly *et al.*, 2013). Non-cellulosic β-1,4-D-glycans therefore, have a strong influence on the accessibility of cellulose to hydrolysis (Rollin *et al.*, 2011). Many successful pretreatment methodologies target these compounds for removal, increasing access to cellulose (Mosier *et al.*, 2005). However, these carbohydrates are also an abundant source of non-cellulosic sugars, which could potentially be useful (Carpita, 2011; Pauly *et al.*, 2013; Scheller & Ulvskov, 2010). Therefore, genetic alterations that improve the extraction of these carbohydrates, under less-severe pretreatment conditions may improve the cost effectiveness of biorefining.

The structures and synthesis of non-cellulosic β-1,4-D-glycans have been recently reviewed, both together (Carpita, 2011; Pauly *et al.*, 2013; Scheller & Ulvskov, 2010) and separately as: XGs (Hayashi & Kaida, 2011; Zabotina, 2012); heteroxylans (Doering *et al.*, 2012; Rennie & Scheller, 2014; York & O'Neill, 2008) and heteromannans (Rodriguez-Gacio Mdel *et al.*, 2012) providing a comprehensive overview of current knowledge of β-1,4-D-glycan structures and synthesis.

1.3.4.1 Xyloglucan structure and biosynthesis.

XG is structurally similar cellulose, as it also contains a β-1,4-linked Glc backbone. Therefore, XGs also form flat ribbon-like structures that can stack closely with cellulose microfibrils (Gorshkova *et al.*, 2013). However, unlike cellulose, XG backbones are partially substituted with xylose (Xyl), followed by other decorations such as Gal, GalA and Ara. Gal residues can also be further acetylated and/or substituted with Fuc (Pauly *et al.*, 2013). Typically, dicot CWs contain XG decorated with α-L-Fuc-(1,2)-β-D-Gal-(1,2)-α-D-Xyl branches (Carpita, 2011). *Arabidopsis* and other closely related dicot CWs typically contain highly branched XG structures, with

three Xyl decorations for every four Glc residues (XXXG) (Pauly *et al.*, 2013). Therefore, this is the likely configuration of *B. napus* XGs.

It is likely that strong binding of XG to cellulose is limited to ca. 15% of the total XG found in dicot PCWs (Dick-Perez *et al.*, 2011) which are likely to hold adjacent microfibrils in close contact (Cosgrove & Jarvis, 2012; Park & Cosgrove, 2012). The potentially strong association between XG and cellulose means that their capacity to limit cellulase actions, even in SCWs, may be larger than their abundance would suggest.

Unlike cellulose, the main backbone of XG is synthesised in the Golgi using a glucan synthase complex containing least one member of the Cellulose synthase-like family C (CSLC) (Zabotina, 2012) and one or more xylosyltransferases (most likely XXT1,2 and 5)(Chou *et al.*, 2012; Pauly *et al.*, 2013). As the catalytic domains of CSLC proteins and XXTs appear are located either side of the Golgi membrane, it is likely that CSLC proteins use the same cytosolic source of UDP-Glc as cellulose synthase, while the main substrate for XXTs is UDP-Xyl residing in the Golgi lumen (Chou *et al.*, 2012). Therefore, transportation of UDP-Xyl or synthesis within the Golgi could alter the availability of UDP-Xyl for xylan synthesis.

Galactosyl- or galacturonosyl-transferases, some of which are known (XLT2, MUR3 and XUT1) decorate side groups with Gal and GalA. XLT2 and MUR3 specifically add Gal resides to either the 2nd or 3rd Xyl residue respectively while XUT1 is the only protein known to add GlcA to XGs (Pauly *et al.*, 2013). XG Fucosyltranferase 1 (FUT1) is an essential enzyme in adding Fuc residues to Gal and GalA and *O*-acetylation of XGs requires at least one member of the Trichome Birefringence-like family (AXY4)(Pauly *et al.*, 2013).

Acetylation of all CW polysaccharides occurs exclusively in the Golgi, presumably using acetyl-CoA as a donor molecule. Whether other acetyl carriers are used is not currently known (Scheller & Ulvskov, 2010). As with compartmentalised NDP-sugars, localised acetyl-group transporters may have an influence on CW polysaccharide acetylation but little is known about the genetic determinants of this process (Scheller & Ulvskov, 2010).

Growing tissues require the remodelling of PCWs to allow expansion. As XG is the main non-cellulosic glycan found in the dicot PCWs, numerous glycosylhydrolases (GHs) are required to allow post-synthesis modification of XG. Examples of these include endotransglucosylases (XTHs, XETs and XEH), xylosidases (XYL), galactosidases (e.g., BGAL10) and α-fucosidases (e.g., AXY8) (Pauly *et al.*, 2013). Considering the dynamic nature of PCWs, it is likely that post-synthesis modification

by GHs are likely to play an important role in CW glycan structure (Scheller & Ulvskov, 2010) and functioning (Hayashi & Kaida, 2011).

As commented in previous reviews, the process of synthesising trimming XG *in muro* appears very wasteful and reasoning for the evolution of this process is unclear (Scheller & Ulvskov, 2010). However, XGs could be synthesised and transported in a decorated (xylosylated, acetylated or methylated) form. If this were the case, exogenous GHs could determine the final properites of biomass for biorefining, rather than XG synthesis genes.

1.3.4.2 Heteroxylan structure and biosynthesis.

The main non-cellulosic β-1,4-D-glycan present in dicot SCWs are (hetero)xylans (Scheller & Ulvskov, 2010). All xylans have a common backbone of β-1,4-D-Xyl residues, substituted in dicots to varying degrees with GlcA, *O*-MeGlcA and *O*-acetyl groups and terminated at the reducing end with a β-D-Xyl*p*-(1,3)-α-L-Rha*p*-(1,2)-α-D-Gal*p*A-(1,4)-D-Xyl*p* motif (Pauly *et al.*, 2013; Rennie & Scheller, 2014; Scheller & Ulvskov, 2010).

Grass xylans tend to contain fewer acetyl-groups and Xyl residues substituted with Ara*f* instead of *O*-MeGlcA (Scheller & Ulvskov, 2010). Ara*f* residues can form covalent bonds with phenolic acids (either terminal *p*-coumaric acid, or cross-linking ferulate esters). This is why *B. napus* straw contains more UA and acetyl-groups and less Ara than most agricultural residues (primarily monocot species) (Pronyk & Mazza, 2012). Moreover, this also explains why saponification of *B. napus* straw releases very few phenolic compounds (Alexander *et al.*, 1987).

Unlike β-1,4-Glucans (cellulose and XG) the synthesis of β-1,4-Xylans is not likely to require a cellulose synthase-like complex (Scheller & Ulvskov, 2010). Instead, Golgi-localised NDP-sugars are polymerised using a suite of Golgi-membrane bound glycosyltransferases (Rennie & Scheller, 2014). Exactly how xylans are synthesised is not fully understood, but a number of potential synthesis routes have been suggested (York & O'Neill, 2008). Many mutants displaying modified xylan phenotypes have been identified – including *parvus,* fragile fibre (fra), irregular xylem (IRX and IRX-like), GAUT, members of the glycoside hydrolase (GH) family 75, glucuronoxylan *O*-methyltransferase (GXMT/GMX) and trichome birefringence-like 29 (TBL29)(Carpita, 2011; Rennie & Scheller, 2014; Scheller & Ulvskov, 2010; York & O'Neill, 2008) which are likely to be involved in xylan synthesis and modification.

It is likely that the reducing end motif is synthesised by IRX7/FRA8, IRX8, PARVUS and members of the GT8 and GT47 protein families (York & O'Neill, 2008). This may act as a primer or terminal molecule for the synthesis of the xylan backbone by xylotransferases (probably IRX9, 10 and 14 and IRX-L 9, 10 and 14)(Doering *et al.*, 2012; Rennie & Scheller, 2014). Xylan backbones can be substituted with GlcA spaced at regular intervals by GlcA substitution of xylan (GUX) proteins (Bromley *et al.*, 2013; Lee *et al.*, 2012a; Mortimer *et al.*, 2010). Glucuronoxylan methyl transferase (GXMT) methylate GlcA residues using S-adenosyl methionine (SAM) as a substrate. Similarly, Xyl residues can be acetylated by the combined actions of; reduced wall acetylation (RWA), trichome birefringence-like (TBL) and domain of unknown function 231 (DUF231) family members (Rennie & Scheller, 2014).

Heteroxylans are important for biotechnology as they are likely to crosslink both polysaccharides via hydrogen bonding and covalently bind phenolic components via GlcA esters, or Xyl and Ara ethers (Rennie & Scheller, 2014). These gene could therefore have major impacts on CW composition and therefore biomass recalcitrance.

1.3.4.3 Heteromannan structure and biosynthesis.

Heteromannans (particularly glucomannans) are a small but significant contributor to dicot CWs – with localised abundance in SCWs surrounding the xylem (Handford *et al.*, 2003). Indeed, *B. napus* straw contains ≈2% mannan, making a small but significant contribution to this biomass source (Pronyk & Mazza, 2012). Mannan backbones are synthesised from GDP-Man and GDP-Glc (Carpita, 2011) found within the Golgi lumen, producing polymers containing a chain of either β-1,4-linked Man (mannan) or β-1,4-linked Man and Glc (glucomannans)(Pauly *et al.*, 2013). Man residues in these polymers can be decorated further using UDP-Gal and UDP-Glc as a substrate – producing galactomannan and galactoglucomannans. Man residues are often *O*-acetylated (Scheller & Ulvskov, 2010).

The backbone of stem heteromannans are synthesised by Mannan synthases (ManS/CSLA 2, 3 and 9)(Goubet *et al.*, 2009), possibly facilitated by other proteins (Pauly *et al.*, 2013; Wang *et al.*, 2013b). Galactosyltransferases (GalT) add GalA to heteromannan chains. The *O*-acetyltransferase that acetylates heteromannans has not been definitively identified but likely candidates include TBL25 or 26 (Pauly *et al.*, 2013).

As with XGs, many genes are likely to alter mannan properties post-synthesis. These include mannosidases, glucosidases, galactosidases and acetyl-mannan esterases. However, in *Arabidopsis,* only three endo-β-mannases (MAN1, 3 and 5) are known to

be expressed in the vascular tissues (Rodriguez-Gacio Mdel *et al.*, 2012). The identity of further mannan-altering enzymes, which could alter biomass recalcitrance remains to be be discovered.

2.6.5 **Pectin structure and biosynthesis**

Pectins are perhaps the most diverse and dynamic group of CW polysaccharides, which are particularly abundant in dicot PCWs (up to 30% w/w)(Atmodjo *et al.*, 2013; Caffall & Mohnen, 2009). Dicot SCWs contain much less pectin than PCWs overall (≈7% w/w)(Ryden *et al.*, 2014), but pectin is concentrated in key areas – such as the middle lamella (Wi *et al.*, 2005) – are likely to play vital roles in SCW functioning and biomass recalcitrance (Xiao & Anderson, 2013). The middle lamella is particularly important as it controls cell adhesion and is the point where lignin deposition begins (Wi *et al.*, 2001).

Pectins are likely to comprise of several, covalently linked, carbohydrate domains with different chemical structures. The most common pectin domains are: homogalacturonan (HG), xylogalacturonan (XGA), rhamnogalacturonan I (RG-I), rhamnogalacturonan II (RG-II). HG, XGA and RGII share a linear α-D-1,4-GalA backbone, as RG-I contains an zigzag backbone of alternating 1,2-α-L-Rha 1,4- α-D-GalA (Yapo, 2011). HG, XGA and RG-II form 'crumpled ribbon' secondary structures, while it is likely that RG-I is helical (Gorshkova *et al.*, 2013). These moieties may interact to form more complex tertiary structures (Vincken *et al.*, 2003). Various branches from the main backbones determine the specificity of the polymers. How these polymers are arranged (either as side chains of RG-I, HG or both) is currently not known (Vincken *et al.*, 2003) and lacking empirical evidence. Extracting CW pectins is often destructive and therefore difficult to achieved (Yapo, 2011). Likewise, it is not known how pectin synthesis is orchestrated (Atmodjo *et al.*, 2013).

More evidence exists regarding the early stages of pectin synthesis in the Golgi apparatus (Driouich *et al.*, 2012). In order to produce such a diverse polymer, it has been estimated that at least 67 glycosyl-, methyl- and acetyl-transferases are needed (Mohnen, 2008). Comprehensive summaries of predicted and proven pectin synthesising enzymes can be found in Caffell and Mohnen (2009) and Atmodjo *et al.*, (2013). It is also likely that at least some enzymes modifying pectins are common to all Golgi-synthesised non-cellulosic glycans – including those with β-1,4-D-glycan backbones and glycoproteins (Atmodjo *et al.*, 2013; Driouich *et al.*, 2012; Mohnen, 2008).

Again, much of the pectin modification relevant to CW properties may be determined post-synthesis. It is likely that pectins are synthesised, transported and secreted in a highly methyl-esterified form (Xiao & Anderson, 2013). Exogenous pectin methylesterases (PMEs) methyl-esterify GlcA residues, allowing Ca^{+} (referred to as egg-box) crosslinks to bind adjacent COO^{-} HG domains (Caffall & Mohnen, 2009). Therefore, pectic properties relevant to bioethanol production could be determined *in muro* (Xiao & Anderson, 2013). This may be particularly true for agricultural residues such as OSR straw, which are relatively rich in pectin and contain a mixture of PCW and SCW material.

2.6.6 Lignin structure and biosynthesis

Some CWs also contain an aromatic polymer known as lignin. Like other CW components, lignin is traditionally defined by the methodologies used to extract it. Essentially, lignin is a phenolic polymer that is highly resistant to degradation which is impervious to extraction by H_2SO_4 (Klason), ball-milling (Bjorkman lignin), alkali or enzymes (Sun, 2010).

Biologically, lignin is a phenolic polymer that integrates with previously assembled polysaccharide components thus strengthening the CW. Lignin deposition begins in the middle lamellae at the intersection between cells (Wi et al., 2005). Lignin is hydrophobic, therefore coating polysaccharides surfaces strengthens hydrogen bonding between polysaccharides. Lignin is therefore found in tissues that require additional strength (such as SCWs) or waterproofing (PCWs in xylem) (Frankova & Fry, 2013).

Research into lignin and lignification has been conducted for decades to improve paper pulping, tensile strength of timber and nutritive value of forage crops (Gomez *et al.*, 2008). As a result, more is known about the evolution (Weng & Chapple, 2010) , genetics (Simmons *et al.*, 2010) and synthesis (Boerjan *et al.*, 2003) of lignin than any of the other CW macropolymer.

In dicotyledonous plants, a number of enzymes convert phenylalanine to monolignol lignin precursors (coniferyl and sinapyl alcohol) in the cytoplasm (Figure 8) (Simmons *et al.*, 2010). Most of the genes involved in this process are known, and have been extensively categorised in *Arabidopsis* (Van Acker *et al.*, 2013). Until recently, mechanisms of monolignols transport across the cell membrane was not known (Liu *et al.*, 2011), but we now know that some ABC transporters actively transport monolignols across cell membranes (Alejandro *et al.*, 2012; Sibout & Hofte, 2012).

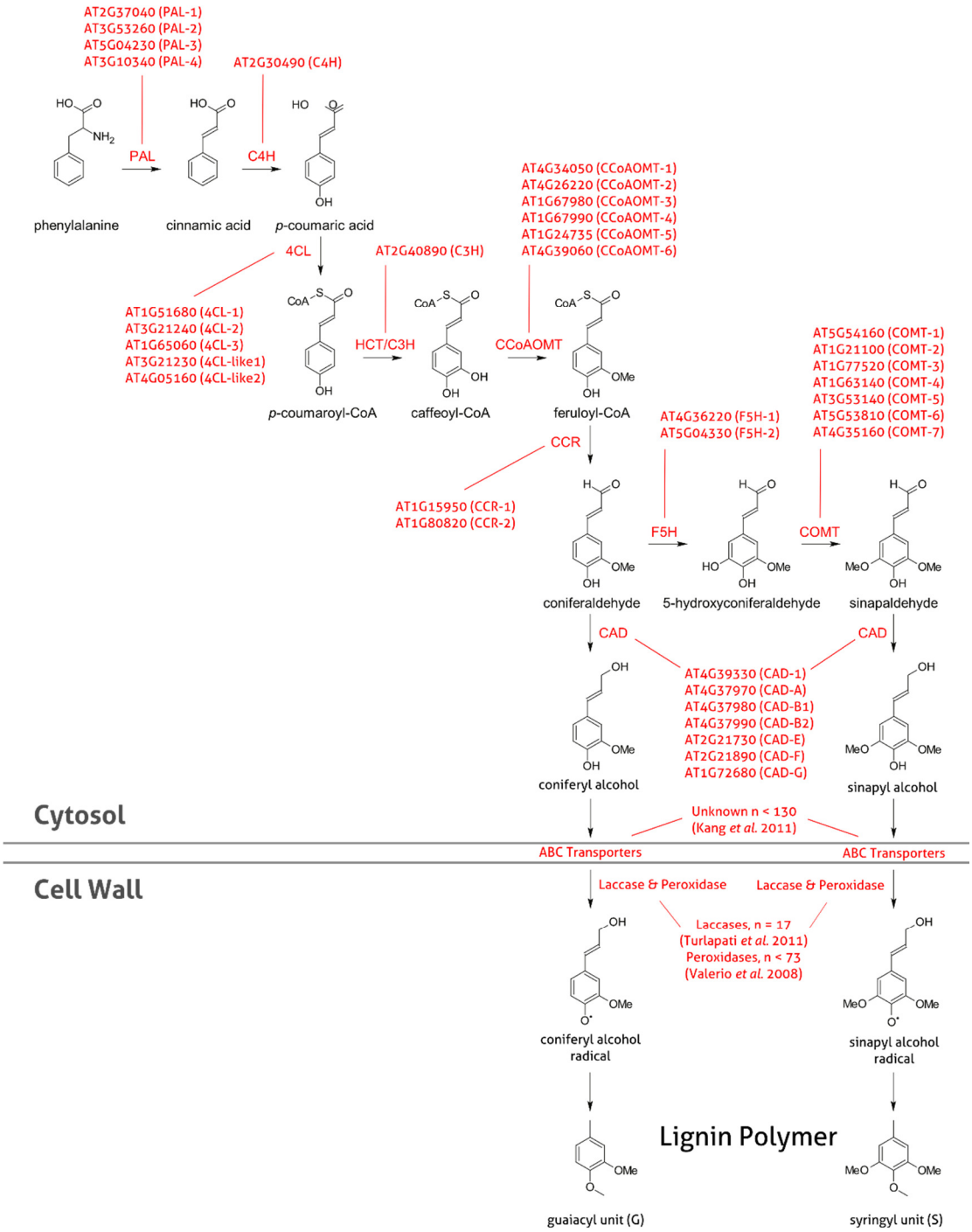


Figure 8 - Summary of the main lignin biosynthesis route in dicotyledonous plants, showing the majority of *Arabidopsis* genes included at each stage (red). Dicots also produce trace amounts of H-lignin (derived from *p*-coumaroyl alcohol), which has not been included. Figure drawn using information from: Goujon *et al.,* (2003), Kang *et al.,* (2011), Turlapati *et al.,* (2011), Valerio *et al.,* (2004) and Van Acker *et al.,* (2013).

Traditionally, it was thought that polymerisation of monolignols in the CW occurred by random, free radical polymerisation, generating a heterogeneous polymer resistant to enzymatic degradation (Gang *et al.*, 1999). However, evidence of repeating units in carefully extracted lignin (Banoub & Delmas, 2003) suggested that the endwise polymerisation of lignin may be directed by enzymatic actions. It likely that peroxidases and laccases are involved in oxidative radicalisation (Wang *et al.*, 2013a), while dirigent proteins addition the activated monolignols to the lignin polymer (Davin & Lewis, 2000).

Lignin is frequently quoted as a major hindrance to enzymatic hydrolysis as a motivation for further research (Akin, 2008). However, emerging evidence suggests that the role of lignin in determining CW recalcitrance is not so straightforward. Lignin abundance can have negative, neutral and even positive effects on biomass recalcitrance in particular circumstances. When surveying the role of delignification across many pretreatment methods, delignification alone can only explain an estimated 50% of the maximum Glc yield; which suggests that cellulose accessibility is more important than lignin quantity *per se* (Rollin *et al.*, 2011).

Mutations in key lignin genes which produce plants with improved saccharification yields, also have significant changes to their non-cellulosic polysaccharide compositions (Van Acker et al., 2013). This makes it very difficult to separate the true role of lignin deposition determining biomass recalcitrance. Recent examples have shown that enzymatically hydrolysing biomass from mutant plants containing higher lignin contents, can still obtain higher saccharification yields compared to wildtype plant tissues (Timpano *et al.*, 2014). The most likely reason for this is CW polysaccharide compositions are also altered, increasing biomatrix opening.

Paradoxically, the presence of lignin can have a positive influence on degradability of biomass in particular circumstances. For example, lignin can have a positive effect on sugar yields from xylan-extracted biomass, possibly by preventing the aggregation of cellulose microfibrils (Ishizawa *et al.*, 2009).

In addition to plant-synthesised lignin, high temperature pretreatments cause the destruction of carbohydrates into acid-insoluble phenolic components known as 'pseudo-lignin' (Kumar *et al.*, 2013). The production of pseudo-lignin is generally undesirable as it slows the hydrolysis of biomass and depletes carbohydrate sources available for hydrolysis (Kumar *et al.*, 2013). Lignin and pseudo-lignin coalesces into droplets that can be visualised with electron microscopy which cause moderate decreases in cellulase performance, but only during the initial stages of hydrolysis (Selig *et al.*, 2007).

Finally, non-specific binding of cellulases to lignin is also viewed as a hindrance to saccharification as non-productive binding of cellulases removes them from solution (Lu *et al.*, 2002). Cellulases derived from *T. reesei* bind strongly to lignin (Rahikainen *et al.* 2013). This would suggest that it is evolutionary advantageous for cellulases to bind to lignin in nature, but this may not be favourable in an industrial setting (Várnai *et al.* 2013).

2.6.7 Cell wall proteins

CW proteins are a minor but important constituent of plant CWs. CW proteins can be broadly divided into two groups, structural and non-structural. The majority of proteins found in *Arabidopsis* CWs (≈45%) are non-structural components that modify other CW components; examples include GHs, esterases, lyases, peroxidases and expansins (Jamet *et al.*, 2006), many of which have been mentioned in previous sections.

Plant cells coordinate the modification of CW compounds using a plethora of extracellular proteins. Examples include: Expansins that disrupt interactions between hydrogen-bonded carbohydrate components (cellulose and XG)(Braidwood *et al.*, 2014), GHs that selectively cleave, repair and produce bonds between CW carbohydrates (Frankova & Fry, 2013) and PMEs that control the properties of the pectic matrix (Jolie *et al.*, 2010). Any of these genes could potentially alter biomass recalcitrance to some degree, by altering the compositions of CW compounds *in muro*.

All plant CWs contain structural proteins which vary in composition depending on cell type (Albenne *et al.*, 2014; Keller, 1993). Structural proteins are most abundant in vascular tissues, suggesting that they may have some function in the lignification process of tissues (Showalter, 1993). Examples of structural CW proteins commonly found in dicot CWs are hydroproline-rich proteins (HPRPs), proline-rich proteins (PRPs), glycine-rich proteins (GRPs), lectins and arabinogalactan proteins (AGPs) (Showalter, 1993).

The name "CW protein" is somewhat of a misnomer when applied to structural CW proteins, as almost all of them (with the exception of GRPs) contain sugar moieties (Showalter, 1993), held together by a comparatively small protein backbone (Ellis *et al.*, 2010). For example, AGPs contain 90-98% glycan and only 2-10% protein (w/w) (Ellis *et al.*, 2010; Seifert & Roberts, 2007). Therefore, protein estimates for OSR straw of ≈ 2-3% (DWB) may underestimate the abundance and impact of these components (Pronyk & Mazza, 2012).

Structural glyco-proteins interact with CW carbohydrates and play a vital role in CW mechanics, which could have implications for biomass recalcitrance. The increased recruitment of structural proteins such as GRPs, extensins and PRPs upon tissue wounding, suggests that the deposition of these insoluble proteins may rapidly strengthen plant tissues (Showalter, 1993).

Recently, Tan *et al.,* (2013) demonstrated that AGPs form covalent cross-links between (arabino)xylans and pectins in *Arabidopsis* CWs. Therefore, AGPs could be vital polymers involved in tethering non-cellulosic components together (Tan *et al.*, 2013). AGPs are also involved in cell signalling (Ellis *et al.*, 2010). Therefore, it is possible that the breakage of these covalent links between macropolymers, deposited during development releases signalling molecules that initiates CW reinforcement. In this capacity, AGPs might play a direct role in CW integrity or an indirect role by controlling the response to CW breakage during development.

2.6.8 Other minor components of lignocellulose

Agricultural residues also contain small, but significant amounts of inorganic compounds known collectively as ash. In certain circumstances, OSR can contain up to 10% DW ash (Petersson *et al.*, 2007) but will differ depending on agronomic conditions. These compounds are vital for mineral nutrition in living plant and are used ion mediators in reactions (Marschner and Rimmington, 1996). Ions present in ash can therefore have stimulatory or inhibitory effects on processing cellulases (Yu & Chen, 2010).

Other minor compounds found in *B. napus* straw are cuticle waxes. Cuticular waxes are particularly important in *B. napus* as they improve water-use efficiency and disease resistance. Therefore, there is much interest in increasing and manipulating wax esters in *B. napus* to improve agronomic performance (Pu *et al.*, 2013). Waxes can hinder the hydrolysis of untreated straw (Hansen *et al.*, 2013). However, plant lipids have relatively low melting points (<100°C) and so do not limit the hydrolysis of straw pretreated under realistic conditions (Hansen *et al.*, 2013). This example illustrates the importance of using commercially relevant technologies to identify suitable targets for biomass-improvement.

2.6.9 Regulation of cell wall biosynthesis

Regulatory genes are likely alter feedstock quality at higher levels and could be used for biomass manipulation. For example, in particular feedstocks such as wheat straw, the proportions of leaf and stem tissue is likely influence overall straw digestability between cultivars (Zhang *et al.*, 2014). Therefore, regulatory genes governing tissue

differentiation are likely to have a strong effect on the recalcitrance of a particular cultviar to hydrolysis.

At lower spatial scales, genes that control the differentiation in tissue types such as growth hormones (auxin-, cytokinin- and brassinosteroid-related) may be important in determining PCW and SCW functioning. Indeed, results from forward genetic screens in *Arabidopsis* indicate that changes in auxin-related genes can influence saccharification yields after processing (Stamatiou *et al.,* 2013). This could have a knock-on effect on whole-straw compositions and recalcitrance (Braidwood *et al.*, 2014; Hussey *et al.*, 2013). Moreover, within each cell at least three tiers of transcription factors including members of the MYB, NST, VND, NAC, SND, BES, C3H, KNAT and KNOX gene families can control SCW formation (for comprehensive review see; Hussey *et al.*, 2013). Together, these regulatory networks may form control points for the CW synthesis at every stage of development.

2.7 Breeding biomass for improved saccharification qualities

2.7.1 Genetic determinants of biomass recalcitrance: Top-down approaches

Collecting information about the genetic determinants of biomass recalcitrance may be vital in improving our exploitation of these abundant but difficult to process materials. Scientifically, the discovery of key 'bioethanol genes' is of great interest as they not only enhance our exploitation of biomass, but also expand our understanding of the natural world (Akin, 2008; Bosch & Hazen, 2013; Himmel *et al.*, 2007; Torres *et al.*, 2013). Novel 'bioethanol genes' may also be of commercial interest, which has resulted in patents filed for their use (Lubieniechi e*t al.*, 2013).

Despite the strong incentive to identify genes relevant to bioethanol production, their discovery is rather challenging. Firstly, CW genes are very abundant, with over 2700 thought to be involved in CW metabolism (Endler & Persson, 2011). Furthermore, breeding targets may not only include genes directly related to the synthesis, but also a plethora of transcription factors that alter biomass properties (Hussey *et al.*, 2013). With such a huge selection of genes to choose from, finding those with the greatest potential is difficult.

The identification of key genes controlling CW synthesis has been led primarily by "top down" manipulation, observing the change in CW chemistry by reverse genetics. CW mutants inevitably exhibit changes in recalcitrance. This generates many potential avenues for biomass improvement (Abramson *et al.*, 2010, Bosch & Hazen, 2013, Burton & Fincher, 2014, Carpita, 2012, Farrokhi *et al.*, 2006, Himmel *et al.*, 2007, Pauly & Keegstra, 2008, Phitsuwan *et al.*, 2013, Yang *et al.*, 2013). However, top-down

methods of identification are limited to known gene families and subject to many practical difficulties.

It is difficult to explore CW genes using top-down approaches as altering one component of the CW often affects others. For example, mutations in known cellulose biosynthesis genes can result in altered lignin deposition (Endler & Persson, 2011) and altered lignin genes can cause changes in non-cellulosic carbohydrate compositions (Van Acker *et al.*, 2013).

Manipulation of important CW genes can have adverse pleiotropic effects on plant growth and functioning, which could not be used to improve crop waste streams. For example, manipulation of cellulose synthesis genes can result in a range of phenotypes including dwarfism, sterility, radial swelling and irregular (collapsed) xylem structures (Endler & Persson, 2011). The cause of these phenotypes could be either direct, or indirect through altered CW signalling (Seifert & Blaukopf, 2010). In some cases, the phenotype can be so strong that determining the exact function of the gene is difficult and application of these genes in a biomass feedstock seemed unlikely.

Conversely, CW mutants may not show a phenotype, despite playing important roles in CW functioning. This phenomenon is particularly prevalent in diverse carbohydrate-altering gene families, where partial functional redundancy occurs between family members (McCann and Carpita, 2005). Mutant plants can compensate for loss of one gene using other isoforms of the same family – resulting in no change in phenotype. This is an evolutionary advantage to the plant but an inconvenience to CW researchers as phenotypes may only be determined using multiple mutants of all members of a gene family (Manabe *et al.*, 2013).

Similarly, a number of strategies can be employed to 'rescue' CW mutants with very strong phenotypes and restore wild-type (WT) growth and development (Bonawitz *et al.*, 2012 and 2014, Petersen et al., 2012). However, these strategies can only overcome some of the immediate challenges associated with top-down manipulation of CW genes. Methods using top-down manipulation to elucidate gene function are very time consuming and labour intensive. Categorising all the genes relevant to CW synthesis using single mutants is likely to take decades, even before exploring multiple mutant combinations. Moreover, probing CW regulatory networks and interactions for improved biomass qualities may be even more challenging (Hussey *et al.*, 2013).

Finally, transferring information gathered using top-down methods, largely conducted in model organisms, into a crop species has additional difficulties. Genetically modified (GM) plants are excellent for elucidating the roles of CW genes and have many advantages, however, political and public resistance to GM crops makes it difficult to

envisage their use, particularly in Europe (Dunwell, 2014). Therefore, non-GM approaches that utilise the 'natural' diversity in biomass saccharification are more favourable (Burton & Fincher, 2014).

In conclusion, top-down manipulation of plant CW genes though reverse genetics is very useful in proving gene function, but is technically challenging, labour intensive and findings may be difficult to apply in crop species. Therefore, we must look outside of conventional approaches of gene discovery, to identify CW genes involved in bioethanol related traits. Ideally, methods that exploit the 'natural' diversity in CW recalcitrance within crop species would be favourable. Methods that narrow our focus on yield-critical genes would also be beneficial over those that broaden search to an ever-increasing number of candidates.

2.7.2 Genotypic variation in biomass recalcitrance: Bottom-up approaches

An alternative way to find key CW genes related to biomass recalcitrance is to take a "bottom-up" approach – undertaking genome-wide searches for genes involved in CW synthesis and recalcitrance. Genome-wide searches for novel CW genes using phylogenetic analysis of previously described genes has been one method used to identify new genes with similar function (Braidwood *et al.*, 2014). Although this strategy has been invaluable in discovering key carbohydrate synthesis genes (Goubet *et al.*, 2009), unfortunately, these strategies also do not help in narrowing our focus on key genes involved in CW synthesis. Rather they expand our search even further, to include more potential candidates requiring validation by reverse genetics.

Transcription profiling from tissues undergoing secondary thickening have also been used to highlight genes important in CW biosynthesis and potentially biomass recalcitrance (Hall & Ellis, 2013). However, this strategy focusses on genes integral to CW thickening and not necessarily ones that related to improved saccharification performance. Moreover, transcriptomic approaches of this kind are limited, as they do not take into account post-transcriptional changes in CW properties (Frankova & Fry, 2013).

Arguably, a more pragmatic approach would be to identify key bottlenecks to saccharification found within differing genotypes of a particular crop species using forward genetics.

Forward genetic screens, are led by phenotypic as opposed to genotypic information. These strategies target genes specifically associated with a particular trait. This can be achieved by collecting phenotypic information from mutant individuals and highlighting those with different CW compositions or saccharification properties

(Stamatiou *et al.*, 2013). However, these screens concentrate on model plant species, with the view to apply this knowledge in crop species in the future. A more directly applicable approach is to screen saccharification yields achieved from crop cultivars, to establish genotypic variations that are directly relevant to target traits, in a species in which the technology may be applied.

One of the few ways of exploring CW production, modification and deconstruction is by taking a genome-wide overview (Yokoyama & Nishitani, 2004). Moreover, when applying this to the challenge of biofuel production, some go further to say that genome wide association (GWA) studies are the only way to dissect CWs genetics for the purpose (Kintisch, 2008; Slavov *et al.*, 2013). However, phenotypic information and complementary genetic data must be gathered to achieve this.

2.7.3 Genome wide association studies: Linking phenotype to genotype

Identifying the genetic candidates responsible for improving biomass characteristics for 2G-bioethanol production is a new challenge. However, crop geneticists have investigated the relationships between phenotypes (quantitative differences in an observable trait) and genotypes (the genetic variation between individuals) for many decades. Therefore, the genetic tools required to address this question are already available, through association mapping (AM).

AM can identify molecular markers (identifiable differences in genetic code) that correlate with the variance in any quantitative trait (Rafalski, 2010). Areas of the genome that contain variants associate with a particular trait are called quantitative trait loci (QTL). AM identifies QTL by correlating the variation in molecular markers (indicators of genotypic differences) with the variance in phenotypes between individuals across a mapping population. AM takes advantage of linkage disequilibrium (LD), where genetic variants (alleles) positioned at neighbouring positions (loci) are more likely to be inherited together compared to those found further away (Rafalski, 2010) at a higher frequency than expected based on recombination (shuffling of the genetic code over time) (Hayward *et al.*, 2012).

Molecular markers for a particular QTL are valuable for two reasons. Firstly, they identify the likely position(s) of the gene(s) of interest. Secondly, they facilitate marker assisted breeding (identify superior germplasm based on molecular marker presence). Genotype data is likely to be much cheaper, easier and more rapid to collect than phenotype data. Therefore, marker assisted selection provides the most direct route to application in a crop species, without the need for genetic engineering.

If sufficient coverage of molecular markers and genomic annotations are available, association mapping can also be used for gene discovery. Consequently, GWA studies are becoming an increasingly popular way to elucidate the likely genetic basis of crop traits (Huang & Han, 2013; Zhu *et al.*, 2008).

GWA studies are inherently observational in nature (see: Lambert & Black, 2012), they seek areas of the genome that associate most strongly with the trait. These signals can be used to focus our efforts on critical areas of the genome connected with the variation in that trait. GWA studies do not attempt to give causation to identified markers or genes; but are the first step in the process of deduction to identify what genes are important for a specific trait (Lambert & Black, 2012). For the purpose of gene discovery, the ability to resolve true associations above the background level of associations is key. Phenotypic signal to noise of GWAs studies is determined by the genetic framework employed and the phenotyping strategy undertaken. By carefully considering these parameters, the chance of detecting pertinent genomic signals can be maximised.

Factors limiting the use of GWA mapping, particularly in crop species, include those that either diminish resolution by masking genetic associations, leading to false negatives (type II error) or those that confuse the pattern of LD, leading to false positives (type I errors) (Rafalski, 2010). The resolution of GWA studies can be influenced by (i) Molecular marker coverage, (ii) phenotypic diversity available in mapping population, (iii) rate of decay of LD and (iv) signal to noise of the trait in question (Huang & Han, 2013, Mackay & Powell, 2007; Rafalski, 2010; Zhu *et al.*, 2008).

High resolution is particularly important when trying to isolate genes associated with saccharification potential. Considering the abundance of CW-related genes found in the model plant *Arabidopsis* is > 2787 (≈10%) (Wang *et al.*, 2012) of the total 27,411 protein-coding genes (Lamesch *et al.*, 2012). The gene density of the *Arabidopsis* genome is ≈4.35 Kb/gene (Lamesch *et al.*, 2012). Therefore, one might expect in a typical search window of ±50 Kb either side of a high probability SNP (Wang *et al.*, 2013c), or ±10 unigenes (transcriptomic pseudogenes) in *B. napus* to contain at least 1-2 potential candidate genes, based on abundance alone. Therefore, very high marker densities are needed to pinpoint molecular markers within an acceptable distance from a single candidate gene.

Historically, molecular marker coverage and genotyping costs limited the use of association genetics as a tool for gene discovery (Zhu *et al.*, 2008). But recent advances in genetic sequencing allow the collections of affordable genotypic data (Snowdon &

Luy, 2012). Genetic sequences gathered from many individuals reveal abundant single nucleotide polymorphisms (SNPs) between genotypes (Hayward *et al.*, 2012). Therefore, SNP marker can provide a genetic platform with sufficient resolution to investigate complex traits (Zhu *et al.*, 2008).

Plant GWA studies are typically conducted using genome sequencing (Zhu *et al.*, 2008) but transcriptome sequencing can also be used (Stokes *et al.*, 2010, Harper *et al.*, 2012). Transcriptome sequencing is ideal for studying genetic variation in polyploid species, such as *B. napus*. Transcriptome sequencing is more cost effective compared to whole genome sequencing (fewer base reads) and focusses on parts of the genome with greatest functional significance (Bancroft *et al.*, 2011). This is particularly important when considering the genetic background of polyploid crop species such as *B. napus*.

2.7.4 Genetic background of *Brassica napus*

B. napus is not only a useful species to study at it produces a very abundant agricultural residue; it also has a useful genetic background for elucidating the genotypic determinants of crop traits in polyploid species. Polyploid species contain many duplicated genome copies. Having many copies of the same genome allows greater phenotypic plasticity than their parental species (Leitch & Leitch, 2008). This is why all *Brassica* spp. contain "morphotypes" displaying very different plant architectures, despite being members of the same species (Tang & Lyons, 2012).

Within the species *B. napus,* various morphotypes (displayed in different cultivars) have been bred to produce either an edible stem (Rutabagas/Swedes, Siberian & Japanese Kales and Fodder) or improved oilseed production (Winter, Spring and Semi-Winter) (Harper *et al.*, 2012). This morphological plasticity is indicative of the genetic diversity present in recently formed polyploids and subsequent artificial selection for various traits.

B. napus is one of six commercially important crop species in the genus *Brassica* – three of which are allotetraploids (*B. carotina, B. juncea & B. napus*) formed by the hybridisation of three diploid ancestor (*B. nigra, B. oleracea & B. rapa*) (Figure 9). The *B. napus* genome is large, containing 19 unique chromosomes ($2n$ = 38), recently inherited from two diploid ancestors – *B. rapa* (10 chromosomes) and *B. oleracea* (9 chromosomes). The hybridization of these diploid species is likely to have occurred during the co-cultivation of the two species by man at some point during the past 10,000 years (Iniguez-Luy & Federico, 2011). The genome of *B. napus* is so similar to those found in *B. rapa and B. oleracea*, they are labelled in respect to their ancestral origins ("A" and "C" genomes respectively). Genomic resources available for *B. napus* include the full genome sequence of the ancestral *B. rapa* genome (The Brassica rapa

Genome Sequencing Consortium, 2011) and large sections of the *B. oleracea* genome sequence are now available in various online repositories (Yu *et al.*, 2013).

Brassica species are the closest crop relatives to the first model plant to have its genome sequenced – *Arabidopsis thaliana* (*Arabidopsis*) (Tang & Lyons, 2012). Functional annotation of the *Arabidopsis* genome conducted over the past 13 years can be found in a freely available, online repository, known as The *Arabidopsis* Information Resource (TAIR, 2014). Both *Arabidopsis* and *Brassica spp.* belong to family (Brassicaceae), sharing a recent common ancestor as recent as 20-24 Mya (Ziolkowski *et al.*, 2006). Consequently, the genomes of *Arabidopsis and Brassica spp.* show strong collinearity (conservation of gene order) (Cheung *et al.*, 2009). This collinearity allows "Parkin blocks" (Tang & Lyons, 2012) showing how large areas of the *Arabidopsis* and *B. napus* genome might align (Parkin *et al.*, 2003 and 2005). Therefore, detailed functional annotation of the *Arabidopsis* genome can be used to provide potential functions of *B. napus* genes (Paterson *et al.*, 2001).

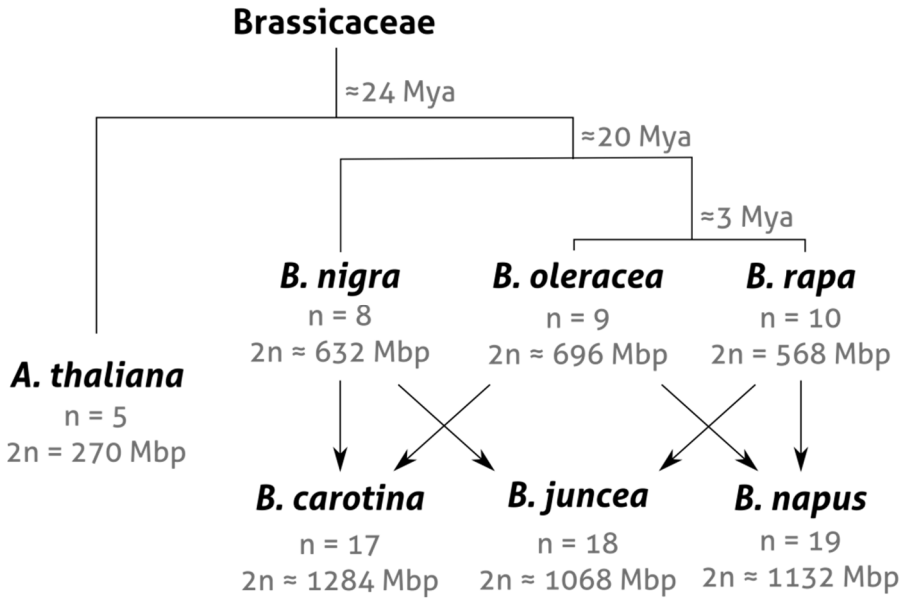


Figure 9 - Phylogenetic relationships between selected members of the Brassicacaceae. Likely divergence dates have been included following (Arias *et al.*, 2014) and estimated genome sizes collated from (Johnston *et al.*, 2005; TAIR, 2014; The *Brassica rapa* Genome Sequencing Project Consortium, 2011). Hybridisation can occur between the three diploid *Brassica* sp., indicated by arrows (Nagaharu, 1935). Hybridisation has occurred many times over the last ≈10,000 years, including backcrossing with diploid parental lines (Cheung *et al.*, 2009).

Although the synteny between the *Arabidopsis* and *B. napus* genomes is not as close as first thought (Snowdon & Luy, 2012), the progressive refinement of marker maps utilising genome (Cheung *et al.*, 2009) and transcriptome data (Bancroft *et al.*, 2011) as it has become available, gives us a good representation of the *B. napus* transcriptome. Association studies conducted in *Brassica* crops can therefore draw upon the comprehensive functional annotation available in *Arabidopsis* to assign likely gene functions to *Brassica* orthologues (similar gene sequence found in different species).

Recently, Harper *et al.,* (2012) demonstrated that GWA studies could be conducted in *B. napus* by combining phenotype data (seed glucosinolate measurements) collected from 53 *B. napus* cultivars, with complementary transcriptomic data (> 63,000 SNPs). A *HAG1* ortholog known to alter glucosinolate biosynthesis in *Arabidopsis* was located in an associated area of the *B. napus* genome (Harper *et al.*, 2012). This platform potentially allows association mapping to be conducted in a crop species with a large genome.

In conclusion, GWA studies can be conducted in *B. napus,* utlising cost-effective and high-resolution transcriptomic sequences (Harper *et al.*, 2012). Transcriptome sequencing captures the most pertinent genotypic variations between crop cultivars, without collecting full genome sequences. Likely candidate genes can be identified in a crop species, with a huge, unsequenced, unannotated genome by taking advantage of the close similarity and colinearity between *Arabidopsis* and *Brassica* genomes. Together with the importance of OSR straw as an abundant biomass waste stream, this makes *B. napus* an excelent species to explore the effects of straw genotype on traits relvant to bioethanol production. With GWA platforms now available, collecting complementary phenotypic data is the primary challenge for researchers (Cobb *et al.*, 2013). Although collecting data relevant for a GWA study using biomass is difficult (Decker *et al.,* 2009), these challenges must be overcome to take advantage of this potentially useful technology.

2.8 Aims and objectives

The ultimate aim of this research was to identify strategies to improve the use of *B. napus* straw for the purpose of cellulosic bioethanol production. It is likely that plant breeding for improved CW compositions will be required to meet this goal (Gressel, 2008; Wyman, 2007). However, traits relevant to bioethanol production can only be determined, and breeding targets identified, if a process is established. This study aimed to identify key bottlenecks limiting the saccharification of *B. napus* straw guided by more specific objectives as follows:

Objective 1 - Establish process-specific determinants of saccharification yield from *B. napus* straw after SE at a pilot-scale.

It is difficult to optimise a feedstock for a process that is not yet established (Wyman, 2007). Therefore a plausible process must first be determined to convert biomass to fermentable sugars. Economic models suggest that autocatalytic pretreatments such as liquid hot water and SE could be the most economically viable pretreatment methodologies available (Kumar & Murthy, 2011). However, our current understanding of how *B. napus* straw (and other biomass sources) respond to pretreatment is often limited to a single measure of saccharification potential, optimal Glc yield irrespective of processing conditions. Therefore, the initial objective of this thesis was to extend our understanding of how SE improves the enzymatic hydrolysis of *B. napus* straw, beyond final sugar yield, to include other measures of pretreatment effectiveness (Bansal *et al.*, 2009).

Objective 2 - Identifying cultivar-specific differences in biomass composition likely to influence product yields at a pilot-scale.

It is likely that a combination of process improvements and plant breeding will be required to make efficient use of *B. napus* straw and other lignocellulosic materials (Gressel, 2008; Wyman, 2007). Although it is known that cultivar selection can significantly alter saccharification yields in other species (Isci *et al.*, 2008, Larsen *et al.*, 2012, Lindedam *et al.*, 2010 and 2012, Matsuda *et al.*, 2011), pertinent intra-specific variations in biomass composition relevant to cellulosic ethanol production from *B. napus* are not currently known. Therefore, the aim of this chapter was to assess if variations in CW composition between *B. napus* cultivars influences saccharification and ethanol yields using pilot-scale steam explosion. If variation is observed, what are the key components preventing or improving saccharification?

Objective 3 - Develop high throughput methods and establish process parameters suitable for biomass screening.

Pilot-scale experiments can identify relevant pretreatment conditions and show the potential effect of genotype using a small selection of cultivars, under industrially relevant conditions. However, logistical difficulties limit the application of these methodologies for biomass screening (Rocha *et al.*, 2012). HT screening of lignocellulosic biomass opens up new avenues of research that would be previously impossible to achieve using conventional methods, but is technically challenging (Decker *et al.*, 2009). Therefore, this study aimed to establish a suitable process to collect quantitative phenotype data relevant to bioethanol production between crop cultivars. A number of existing techniques can be tailored specifically for this purpose.

Objective 4 – Collect data relevant to bioethanol production from a wider variety of *B. napus* cultivars using various processing conditions.

Screening platforms typically limit their processing regimes to those that yield near maximum Glc yields (Lindedam *et al.*, 2014). However, this potentially limits subsequent interpretation of the results as they may only reflect the intraspecific variations in straw composition related to specific conditions. This concept is crystallised in the axiom, "you get what you screen for" (Decker *et al.*, 2009). Alternatively, traits associated with biomass recalcitrance may be independent of screening conditions (Lindedam *et al.*, 2014). By collecting data from many cultivars, using disparate and complementary screening conditions further evidence can be gathered related to this debate. The data can also provide additional information regarding biomass properties related to saccharification. For example, organic acids released from biomass during pretreatment are likely to act as a catalysts to depolymerisation, under autocatalytic conditions (Hendriks & Zeeman, 2009). Using screening datasets, this hypothesis can be investigated in further detail.

Objective 5 – Use GWA mapping to highlight areas of the *B. napus* transcriptome associated with the range of product yields relevant to bioethanol production between cultivars.

A number of fundamental questions remain regarding the genetic determinants of biomass recalcitrance. Even basic questions regarding the main genetic bottlenecks to processing can only be partially answered using conjecture, based on current knowledge of CW properties and synthesis (Abramson *et al.*, 2010, Bosch & Hazen, 2013, Burton & Fincher, 2014, Carpita, 2012, Farrokhi *et al.*, 2006, Himmel *et al.*, 2007, Pauly & Keegstra, 2008, Phitsuwan *et al.*, 2013, Yang *et al.*, 2013). Ideally, informed decisions about how to deconstruct biomass would be based on genetic variants shown

to associate with saccharification yields, collected using industrially relevant conditions. GWA mapping is one of the few techniques with sufficient scope to give such an overview (Slavov *et al.*, 2013). Therefore, this study aimed to use GWA mapping to locate areas of the *B. napus* transcriptome that correlated most strongly with traits relevant to bioethanol production. Candidates genes found in close proximity to highly associated markers should permit the development of falsifiable hypotheses regarding the nature of biomass recalcitrance. Observations of this nature could potentially accelerate the identification of genes associated with these complex traits, paving the way for further research in this field.

Objective 6 – Assess the applicability and potential biological relevance of a selection of SNP markers and candidate genes.

The primary goal of any GWA study is to observe a previously unobservable connection between phenotype and genotype. In this regard, GWA studies are hypothesis generators (Lambert & Black, 2012), that focus our attention on key areas of the genome, which may harbour important genes related to the particular trait. However, potential breeding targets, identified using association mapping, require validation before implementation. Furthermore, candidate genes suggested from the patterns of association observed would also benefit from closer investigation. Therefore, the robustness of a selection of molecular markers were tested in *B. napus*, using phenotype data collected from an independent diversity panel. A handful of potential candidate genes were also investigated in *Arabidopsis* to give an example of how GWA studies can be used probe the nature of biomass recalcitrance. The result also give an indication as to the effect of particular genes on saccharificiation phenotypes.

2.9 Chapter contributions

This section outlines how each chapter contributes to achieving the objectives outlined above (**Figure 10**).

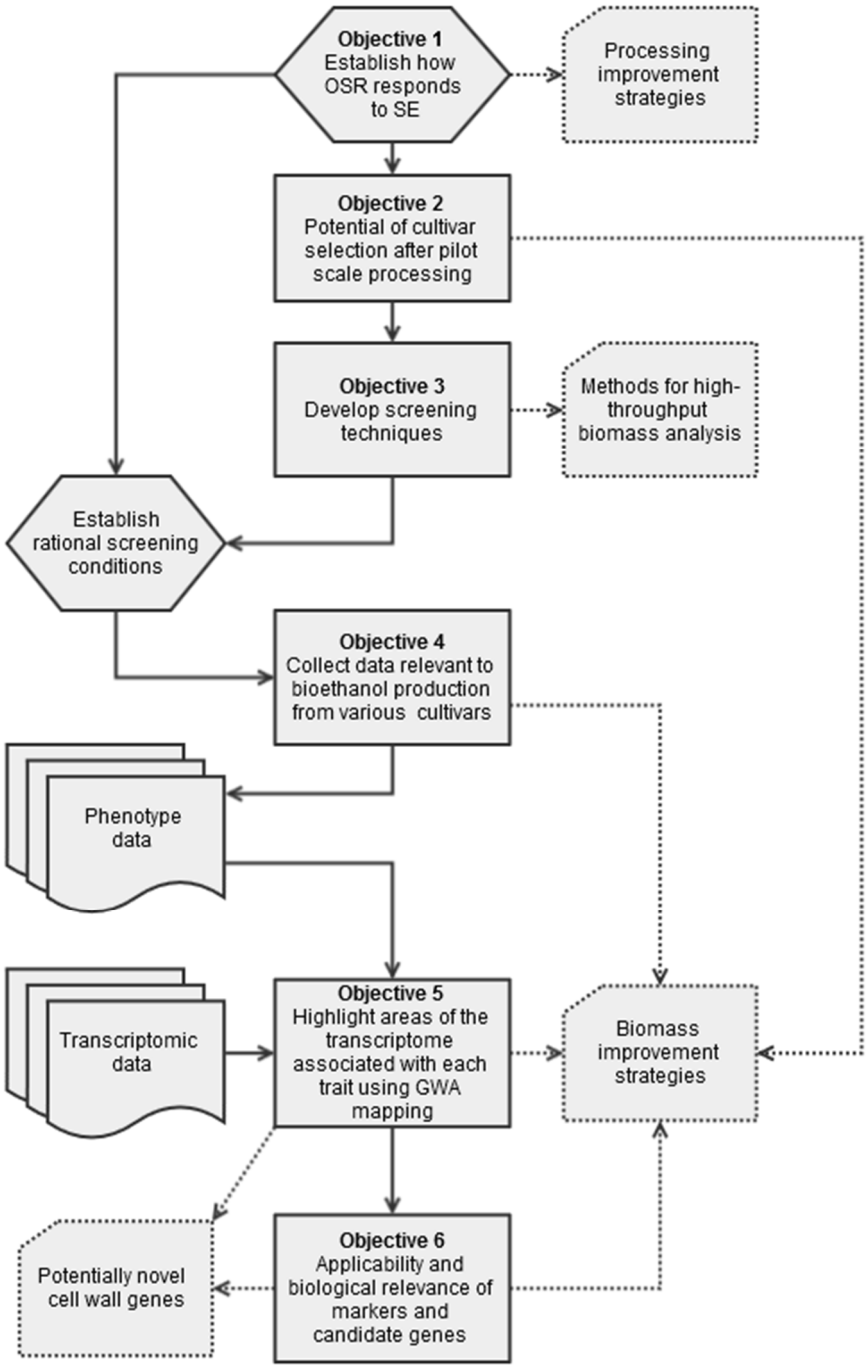


Figure 10 - Flow diagram illustrating the main objectives addressed in this study, and how they relate to each other. The desired outputs at each stage are displayed as boxes with dotted lines.

Steam explosion of oilseed rape straw: establishing key determinants of saccharification efficiency at a pilot-scale (Chapter 4, Objective 1)

- Gain an overview of how *B. napus* straw responds to autocatalytic pretreatment at varying pretreatment severities.
- Extend our understanding of how SE improves the enzymatic hydrolysis of *B. napus* straw through the impact of pretreatment, including changes in substrate-cellulase interactions.
- Highlight the key limitations to saccharification efficiency of *B. napus* straw at various stages of hydrolysis.

Variability in product yields relevant to cellulosic ethanol production from OSR cultivars after pilot-scale steam explosion (Chapter 5, Objective 2)

- Observe the differences in product yields relevant to 2G bioethanol production obtained from various cultivars processed at a pilot-scale.
- Identify the likely compositional differences between cultivars that will influence saccharification yields after pilot-scale SE.

Screening biomass for traits relevant to cellulosic ethanol production (Chapter 6, Objective 3 and 4)

- Tailor existing methods specifically to screen of biomass from various cultivars.
- Establish rational, complementary and contrasting screening conditions that will maximise the interpretation of later screening work.
- Obtain data relevant to 2G bioethanol production suitable for a pilot-scale GWA study.
- Evaluate to what extent genotypic variations in traits relevant to bioethanol production are related to process conditions.

Identifying areas of the *Brassica napus* transcriptome associated with cellulosic ethanol production (Chapter 7, Objective 5)

- Locate likely areas of the *B. napus* transcriptome that correlate most strongly with traits relevant to 2G bioethanol production – thereby highlighting potential genetic determinants of CW recalcitrance between cultivars.
- Suggest likely candidate genes using functional annotations in *Arabidopsis.*
- Use the identity of likely candidates to develop testable hypotheses that could reveal genetic determinants of biomass recalcitrance.

High-probability marker and candidate gene assessment in *B. napus* and *Arabidopsis thaliana* (Chapter 8, Objective 6)

- Test a selection of molecular markers using a second diversity panel to assess their suitability and robustness for marker-assisted selection.
- Select a number of interesting candidate genes likely to alter saccharification yields based on the observed association patterns and obtain *Arabidopsis* mutants deficient in those genes.
- Compare plants deficient in those genes, with WT plants in terms of growth and saccharification yields after processing under various conditions.

3 General methodology

3.1 Common chemicals, cellulases and equipment

Two commercially available cellulase cocktails were used in the following experiments, derived from *Trichoderma reesei.* Earlier work was carried out using Accellerase® 1500 (Genencor, UK) with a stock solution cellulase activity of 72.9 FPU/mL – calculated following Ghose *et al.,* (1987). Later work was conducted using Cellic® CTec 2 (Novozymes, Denmark) - with a stock cellulase activity of 180 FPU/mL.

Unless otherwise stated, all chemicals were purchased from Sigma-Aldrich®, UK.

3.2 Sample preparation techniques

Various partial size reduction techniques were used in this study to gain representative straw samples. Generally, straw was sequentially milled to the desired particle size – ranging from whole plant stems, to chipped material (3-4 cm), to milled material (< 2 mm), to a fine powder (40-85 nm). Stems and empty pods from whole plant were either supplied in a chipped form, or chipped into 3-4 cm sections using an industrial disk pin mill (D6450, Condux, Germany). If necessary, further size reduction was used, milling chipped straw to pass a 2 mm sieve using a laboratory sieve-mill (IKA® MF10 Microfine grinder drive, equipped with a cutting-grinding head – IKA-WORKS Inc., USA). In cases where finer material was required, a representative sample of milled straw (≈3 g) was cryogenically milled. Samples contained in polycarbonate centre cylinders, with stainless steel impactors and end plugs (SPEX SamplePrep, USA) were submerged in liquid nitrogen and milled for 3 min at maximum impact frequency (Spex Freezer/Mill® 6700, SPEX SamplePrep, USA).

3.3 Determining the chemical composition of solids

To gain detailed information as to the compositions of solid biomass, a number of techniques were used. Together, these techniques quantify the majority of components present in solid biomass including water content, neutral sugar composition, UA, Klason lignin and ash content.

3.3.1 Moisture content

The moisture content of the pretreated biomass was established using an infrared drying balance (Mettler LP16, Mettler-Toledo, Belgium) typically drying duplicate samples (0.5 g) at 105°C to constant mass.

3.3.2 Sugar composition of solid materials by gas chromatography (GC)

Determining the abundance of constituent sugar monomers in solid samples was achieved by hydrolysing 3 mg of dry sample in 200 µL of 72% H_2SO_4 (w/w) for 3 h at room temperature, before dilution to 1 M H_2SO_4 by adding 2.2 mL of ddH_2O. The diluted solutions were incubated for a further 2.5 h at 100°C to gain monomeric sugars in solution. This is commonly known as Saeman hydrolysis (Saeman *et al.*, 1945). At this point, an internal standard was added to all samples and standards (2-deoxy Glc, 200 µg) and a 1 mL sample of the supernatant taken forward to the next step.

This hydrolysis methodology differs from commonly used protocols such as the National Renewable Energy Laboratory (NREL) (Sluiter, 2004), who use slightly different hydrolysis conditions to those used here (IFR). Using the NREL method, the sample is initially hydrolysed (1:10, sample to 72% H_2SO_4, 30 °C, 60 min before dilution to 4% H_2SO_4, 0.4 M, 121 °C, 1 h). Sugar recovery standards are used to account for the loss in sugars during hydrolysis (Sluiter, 2004).

To clarify the differences between these two methods, sugar recovery standards (Glc and Xyl) were taken through both NREL and IFR hydrolysis protocols and analysed by HPLC (section 3.6.2. pg. 80) to compare losses during hydrolysis (n = 3, injected in duplicate). The NREL protocol resulted in the loss of Glc (93.5 ± 1.7% recovery) and Xyl (84.9 ± 0.8% recovery). However, no significant loss in sugar was observed using the IFR protocol – Glc (103.5 ± 1.2% recovery), Xyl (99.8 ± 1.2% recovery) after analysis. Therefore, sugar recovery standards are required for accurate quantification using the NREL protocol, but they are not needed when using the IFR hydrolysis conditions.

Monomeric sugar concentration of each sample was typically established by gas chromatography (GC). To prepare samples for GC, acid-hydrolysed-sugars were converted to their aditol-acetate derivatives following Blakeney *et al.*, (1983). Briefly, the hydrolysed sugars were first reduced to their respective aditols using 3M NH_3 containing 150 g/L $NaBH_4$ (100 µL, 1h, 30°C) - followed by neutralisation (200 µL acetic acid). A 300 µL sample was acetylated using 3 mL acetic anhydride in the presence of a catalyst (450 µL 1-methylimidazole) (30°C, 30 min). The derivatised aditol-acetates were diluted (adding 3.5 mL ddH_2O) and solvent extracted twice, using dichloromethane (3 mL DCM). Both extracted organic layers (DCM) were combined, backwashed with 3 mL H_2O and the aqueous layer removed by aspiration. The remaining DCM was evaporated (40°C, under nitrogen), reconstituted in acetone (1 mL) and loaded into 100 µL glass vials (0.1 mL Sci-Vi Crimp Top Vial, Chromacol, UK).

The prepared samples (2 µL) were passed down a Restek™ RTX™-225 Capillary Column (15 m length, 0.32 mm inside diameter, 0.25µm film thickness) held in a

PerkinElmer AutoSystem XL Gas Chromatograph. The injected sample was volatilised at 140 °C for 5 min followed by ramping 2.5°C / min to 210°C, hold for 17 min. Products were detected by flame ionisation.

GC is preferable to other methods of quantifying neutral sugars, such as HPLC because all neutral sugars (Rha, Fuc, Ara, Xyl, Glc, Man, Glc) can be quantified and are baseline resolved (Figure 11). Precise quantification of all components can therefore be achieved. One disadvantage of this methodology is that derivatisation of hydrolysed sugars is labour intensive and cannot detect the presence of UAs. Similarly, sugar oligomers (particularly cellobiose) that could indicate incomplete hydrolysis, cannot be detected by this method.

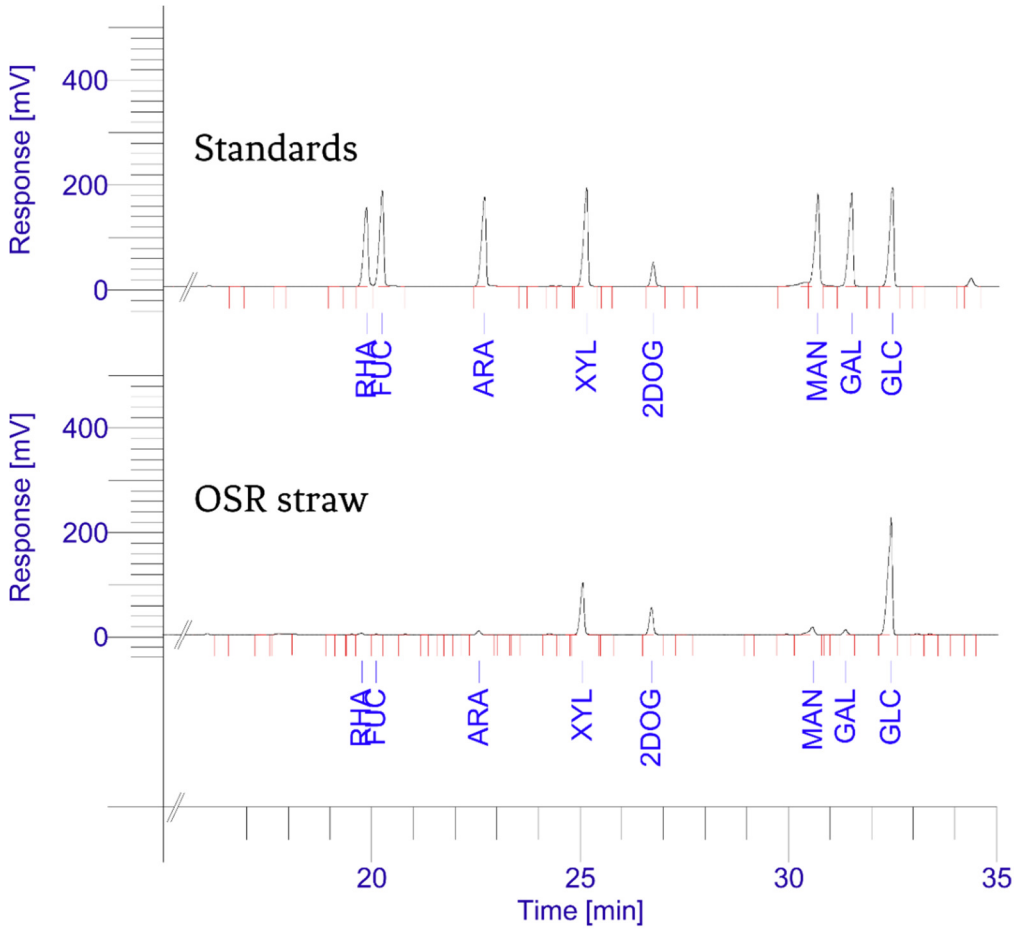


Figure 11 - Typical chromatograms quantifying the abundance of derivatised sugar standards (Standards) and sugars hydrolysed from OSR straw by GC and detected by flame ionisation. Plant sugars (Rha, Fuc, Ara, Xyl, Man, Gal, Glc) were quantified relative to an internal standard (2-Deoxy-D-glucose - 2DOG)

3.3.3 Uronic acid content in solid materials

UA content of a solid sample was established using milder acidic conditions and products quantified established colorimetrically (Blumenkrantz & Asboe-hansen, 1973).

Briefly, 3 mg of sample was acid-hydrolysed (72% H_2SO_4, 20 °C, 3 h → dilution to 1M, 100 °C, 1 h) and a sample of the supernatant (500 µL) transferred to acid-washed test tubes. The acid-hydrolysed solutions were treated with 3 mL of 25 mM $Na_2B_4O_7$ in conc. H_2SO_4. The solutions were boiled in a water bath (100°C, 10 min) before cooling on ice. A 50 µL dose of 3-phenyl phenol (0.15%) in 0.5% NaOH was added to each of the solutions and vortexed. Solutions were incubated in darkness for 30 min and 200 µL of each solution transferred to a flat-bottomed reader plate. The absorbance of the product was measured at 490 nm and compared to a set of GalA standards (0-40 µg/mL, n = 6). Substrate controls (substituting 3-phenyl phenol solution for 0.5% NaOH) were included with each plate and subtracted from the total. A key limitation of this method is that it does not discriminate between GalA and GlcA. Moreover, the response factor (absorbance/mass) of GalA is slightly lower than GlcA (Blumenkrantz & Asboe-hansen, 1973). Therefore, this method approximates the total UAs found in the sample.

3.3.4 Klason lignin content

Klason lignin content was determined gravimetrically after acid hydrolysis (100 mg sample). The residue remaining after 72% H_2SO_4, 20 °C, 3 h followed by dilution to 1M, 100 °C, 2.5 h that remained after filtering through a No. 4 sintered glass filter was quantified. The remaining acid-insoluble material (Klason lignin) was dried to a constant weight at 105 °C overnight and weighed. The remaining acid insoluble residue was ashed using the conditions outlined below and the residual material (acid-insoluble ash) subtracted from the total.

3.3.5 Ash content

Ash content was calculated by charring samples in dry, pre-weighed crucibles a muffle furnace (120 °C, 2 h followed by 250 °C, 4 h followed by 500 °C, 24 h; ramping at 5 °C/min, 2 °C/min and 5 °C/min respectively). The mass of the remaining solid was compared to the mass of the original material.

3.3.6 Fourier transform infra-red spectroscopy (FT-IR)

Fourier Transform Infra-red (FT-IR) spectra were collected in the 800–4000 cm^{-1} region for each homogenised sample using a dynamic alignment FT-IR spectrophotometer (Bio-Rad FTS 175C, Bio-Rad Laboratories, Cambridge, USA), speed 10 kHz, filter 5, UDR 2, resolution 2cm^{-1}, sensitivity 1, 64 scans. To do this, the sample was trapped against a diamond element in a Golden Gate™ diamond attenuated total reflectance (ATR) accessory (Specac, Slough, UK). The diamond was cleaned using 100% ethanol, ensuring all moisture evaporated between samples. Typically, five spectra were taken for each material. Solid samples were analysed using air as a reference medium, while liquid samples were compared to ddH_2O. Spectra were truncated to the 800-1800 cm^{-1} fingerprint region, baseline anchored to 1800 cm^{-1} and areas normalised before averaging.

3.4 Pretreatment methods

3.4.1 Lab-scale – Liquid hot water (LHW)

Small scale pretreatment (< 1g biomass) were conducted using a Biotage® Initiator^{+} Microwave Synthesizer (Biotage®, UK). Freeze-milled straw from each cultivar was made into 15 mL slurries containing 5% substrate (FWB) in ultrapure H_2O (Milli-Q, Millipore, USA), a magnetic flea was added to ensure continual movement of the substrate (900RPM) during pretreatment and the contents sealed in 10-20 mL Biotage Microwave Vials. By using this machine, a desired temperature of the liquid could be rapidly (2-5 °C/sec), consistently and automatically achieved by focussing a 400W microwave beam onto the vial. The sample was held at the desired temperature ± 2°C for the duration (typically 10 min). Compressed air cooled each vial to <50°C.

3.4.2 Pilot-scale – Steam explosion (SE)

A bespoke Cambi™ Steam Explosion Pilot Plant (Cambi, Asker, Norway) was used to pretreat materials at a larger scale (1 kg) (Figure 12). After pre-heating the boiler at the desired temperature for approximately 1 h, air-dried and chipped (< 5 cm) OSR straw (1 Kg) was introduced into the pressure chamber (30 L volume). The chamber was sealed and the temperature increased to 180-230 °C (9.01 to 26.94 bar) using steam and held for 10 min. The pressure chamber was rapidly opened, expelling the contents into the cyclone chamber. A cyclone of water jets separated steam from the solid residue. The majority of the pretreated solid was deposited into the collection hopper, containing 6.6L of water. The chamber was cleared twice by applying 2-3 bar of pressure (30 sec) and releasing. The contents of the hopper were collected and filtered through a 100 µm nylon mesh bag in a low speed centrifuge while still warm. The two fractions were collected in their entirety before taking representative samples.

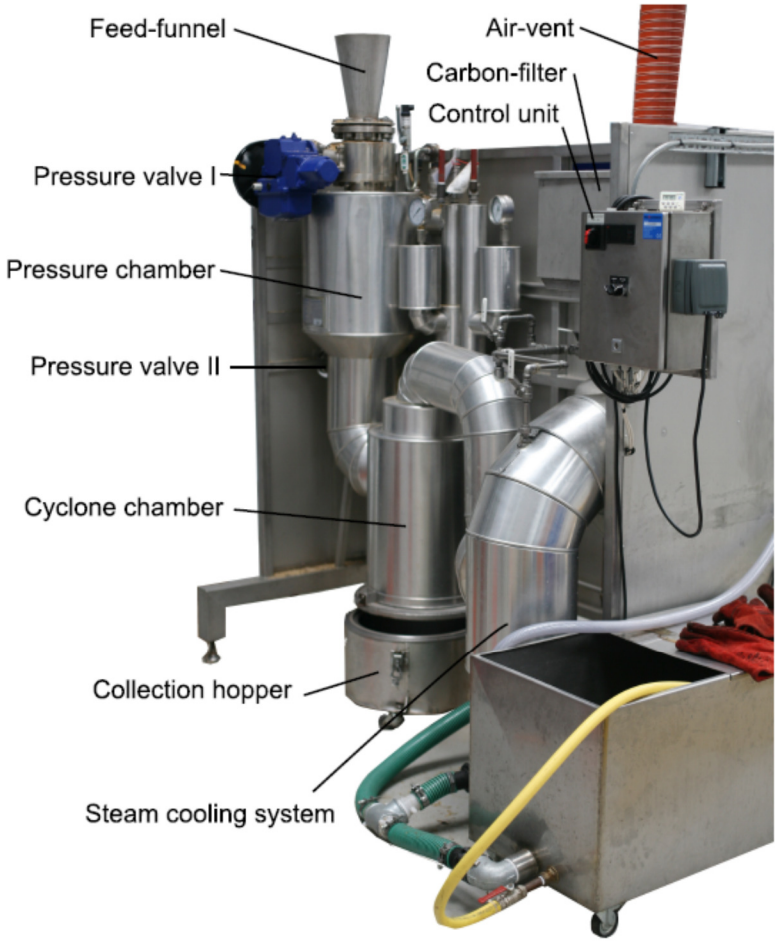


Figure 12 - Cambi™ Steam Explosion Pilot Plant located at The NRP Biorefinery Centre.

3.5 Saccharification and fermentation methods

3.5.1 Enzymatic saccharification

Enzyme digests (25 mL in total) were typically conducted using a desired quantity of substrate, adjusted to the selected substrate concentration in sodium acetate / acetic acid buffer (0.1M, pH 5, 0.01% thiomersal) held in 30 mL Universal containers (Sterlin, UK). Vials were continually agitated (30-50 °C, 150 RPM) in a shaker plate incubator (Thermoshake, Gerhardt, Germany). Cellulase was added to temperature-equilibrated solutions and digestions typically conducted in three independent replicates. If time-course data was taken from the same digest, the quantity of sample taken over the time-course was kept to < 1% of the total volume (< 250 µL). If larger samples were required, parallel digests were conducted and samples taken across several vials. All comparable hydrolyses were conducted at the same time, under identical conditions. The enzyme cocktails used in this study contained significant amounts of RS (Ctec2 - 332 ± 52 g/L), most of which was Glc (Ctec2 - 305±22 g/L) when 10x diluted samples

were analysed by DNS and GOPOD respectively (See section 3.6, below). Enzyme controls containing only buffer and cellulase were included and subtracted from the total to gain a more accurate quantification of biomass-derived sugars.

3.5.2 Simultaneous saccharification and fermentation (SSF)

A sample of each pretreated substrate was suspended in 10 mL solution with a final concentration of 5% substrate in nitrogen base (Formedium, Hunstanton, UK) and held in 20 mL screw-topped glass vials. Both cellulase (36 FPU cellulase/g substrate) and a concentrated yeast inoculum were added to each vial and incubated for 96 h, 40 °C. The yeast inoculum used (*Saccharomyces cerevisiae,* strain NCYC 2826, National Collection of Yeast Cultures, Norwich, UK) was grown from a slope culture, inoculating 1 L of Yeast Mould broth (3 d, 25°C) before centrifuging, discarding the supernatant and partially reconstituting the yeast in nitrogen base. The number of viable cells / mL when inoculated was calculated using a NucleoCounter® YC-100™ (ChemoMetec, Denmark). SSFs were conducted as three independent replicates and the ethanol released from a cellulase + yeast control was subtracted from each sample.

3.6 Quantification of hydrolysis and fermentation products

3.6.1 Colorimetric methods (DNS, GOPOD)

3.6.1.1 Reducing Sugars: Dinitrosalicylic acid (DNS) assay

The DNS method involves mixing dinitrosalicylic acid (DNS) reagent with a sample, heating to catalyse the reaction (Equation 2) and measuring the visible absorbance of the products (particularly 3-amino-5-nitrosalicylic acid).

3,5-dinitrosalicylic acid + reducing sugar → 3-amino-5-nitrosalicylic acid + oxidised sugar

Equation 2

DNS reagent contains 3,5-dinitrosalicylic acid (10 g/L), sodium potassium tartrate (300 g/L) and NaOH (16 g/L) and was stored in darkness at room temperature.

Scaled assays (optimised in Chapter 6) were typically conducted using 9 µL of sample and 171 µL DNS solution (1:20, sample: DNS reagent). DNS reagent was added to the sample in 0.34 mL, 'tall-chimney', 96-well plates (Fisherbrand®, UK) closed with TPE PCR sealing mats (BRAND, at Fisher, UK). The resulting solutions were heated in a thermocycler (Biometra® T-Gradient, Germany) typically at 100 °C, 1 min, held for 2 min at 20 °C to cool. A smaller aliquot (90 µL) of the solution was transferred to a 96-well flat-bottomed microtitre plate (Nunc, Denmark) and analysed using a microplate spectrophotometer (Benchmark Plus, BioRad, CA, USA) within 3 min.

3.6.1.2 Glucose: Glucose oxidase/peroxidase (GOPOD) assay

Glc concentrations in biomass hydrolysates were calculated using a Glucose oxidase/peroxidase (GOPOD) assay kit (Megazyme, IRL). Initially, glucose oxidase converts Glc to gluconate and hydrogen peroxide (Equation 3). The hydrogen peroxide released (Equation 3) activates a quinoneimine dye, catalysed using a peroxidase (Equation 4). Glc abundance is quantified by comparing the visible absorbance of the sample to a set of Glc standards.

$$\text{D-glucose} + H_2O + O_2 \rightarrow \text{D-gluconate} + H_2O_2$$

Equation 3

$$2\ H_2O_2 + p\text{-hydroxybenzoic acid} + \text{4-aminoantipyrine} \rightarrow \text{Quinoneimine dye} + 4\ H_2O$$

Equation 4

It is important to note that the absorbance of the GOPOD reagent is linear at concentrations < 2 mg/mL and can cause inaccuracies in quantification if not considered (Lindedam *et al.*, 2014). Therefore, extensive dilution of biomass hydrolysates is often required to gain accurate quantification and highlights the importance of flanking standards, particularly when using colorimetric methods.

In this thesis, scaled versions of the GOPOD method were used, adding 195 µL GOPOD reagent to 5 µL of diluted sample in a 96-well flat-bottomed microtitre plate with lid (Nunc, Denmark). Plates were covered and incubated (50 °C, 20-30 min) before reading the absorbance (510 nm) using a microplate spectrophotometer (Benchmark Plus, BioRad, CA, USA).

3.6.2 High performance liquid chromatography (HPLC)

Hydrolysates (300 µL) were filtered to remove any particulates before analysis either individually (syringe filtered, 0.22 µm, Millex®, Millipore, USA) or collectively in 96-well filter plates (AcroPrep™, 0.2 µm PTFE, Pall Corporation, NY).

A Flexar® FX-10 UHPLC instrument (Perkin Elmer, UK) equipped with an Aminex HPX-87P carbohydrate analysis column (Bio-Rad Laboratories Ltd, UK) (85 °C, mobile phase Milli-Q water, flow rate 0.6 mL/min) and refractive index detector was used to quantify sugar and ethanol in biomass hydrolysates (Cellobiose, Glc, Xyl, Gal and Ethanol)(Figure 13). The main strength of this method is that hydrolysates can be directly analysed, to establish their component sugars in a single run. This avoids the need for lengthy derivatisation of sugars, as with GC. Also, the detection of cellobiose

can reveal incomplete hydrolysis which would not be detected using other methods. However, the co-elution of minor sugars (e.g., Gal with Rha; Ara with Fuc), means that quantification is less precise.

Organic acids and furfural derivatives (acetic acid, formic acid, 2FA, HMF) present in pretreatment liquors were quantified using the same HPLC system, fitted with an Aminex HPX-87H organic acid analysis column (Bio-Rad Laboratories Ltd, UK) (65 °C, mobile phase 5 mM H_2SO_4, flow rate 0.5 mL/min)(Figure 14). Organic acids were quantified using an RI detector and furfural derivatives quantified using a photodiode array detector (PDA) at 210 nm. The principle caveat of this method is the poor base-line resolution of products, particularly formic and acetic acid (Figure 14).

3.7 **Robotic liquid handling**

Typically, calibration standards and samples were loaded into 96-well 2 mL 'deep-well' polypropylene plates (Nunc®, Roskilde, Denmark) using hand-held pipettes. In cases where many samples were analysed simultaneously, a Tecan Freedom Evo® 100 workstation fitted with a 96-well MultiChannel Arm (MCA96) with disposable tips (Tecan, Männedorf, Switzerland) moved liquids between multiple plates. Samples and reagents were added to each well of a 96-well plate simultaneously. Pipetting speeds were slowed to 250 µL/s to avoid liquid loss upon aspiration and docking sites were arranged to minimise travelling distance.

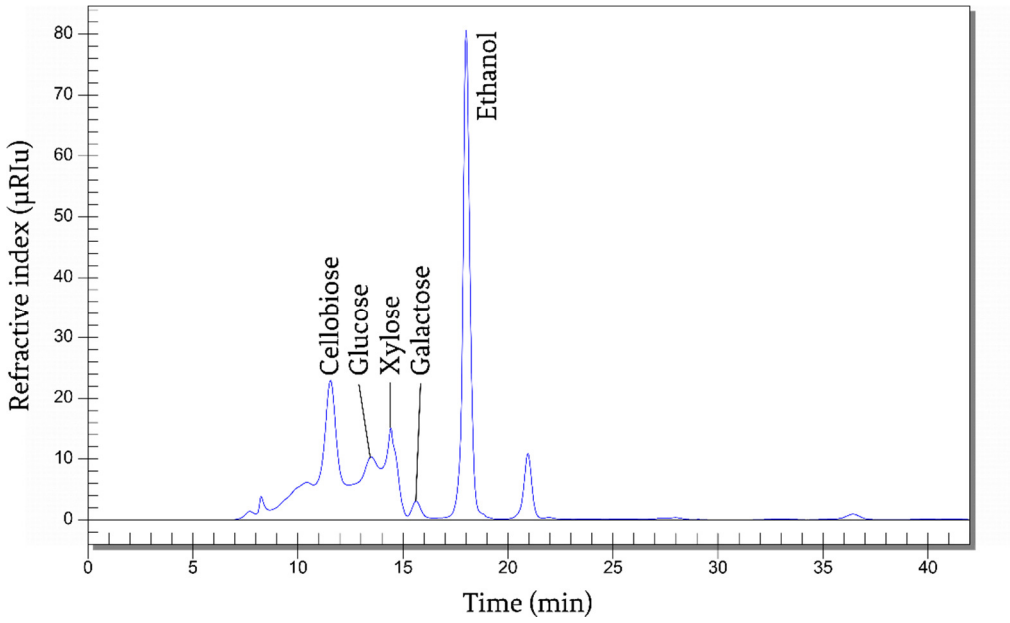


Figure 13 – Example chromatogram quantifying the abundance of sugars and ethanol in partially hydrolysed steam exploded OSR straw by HPLC using an Aminex HPX-87P column.

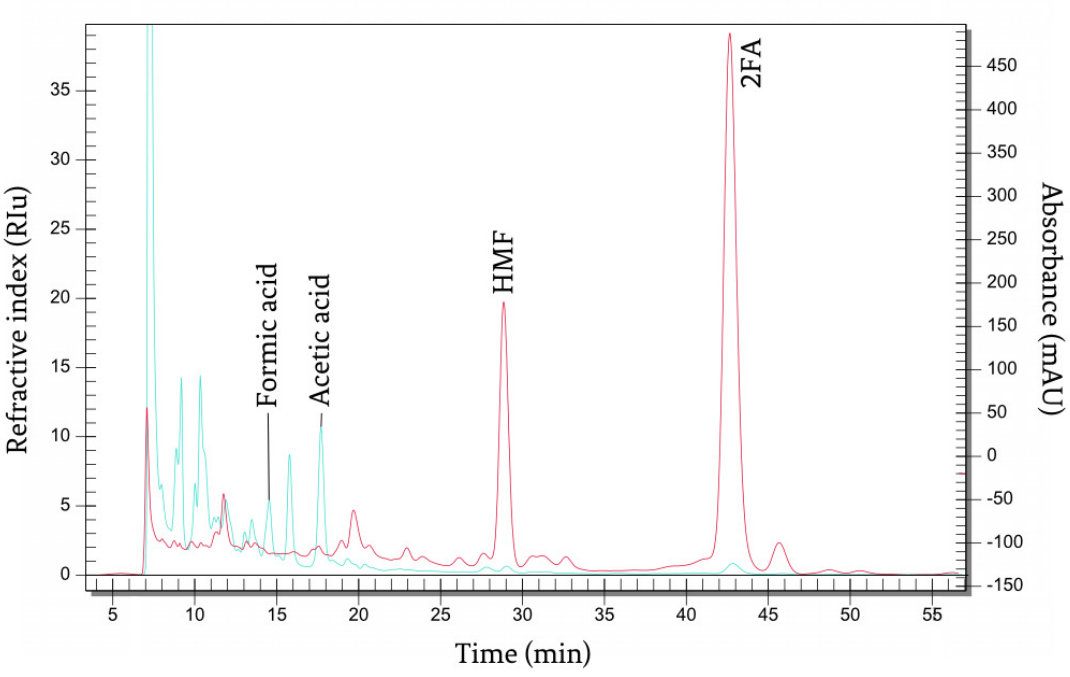


Figure 14 – Typical chromatograms quantifying the abundance of organic acids released from pretreated OSR straw (210°C) using an Aminex HPX-87H column, quantifying using a RI and PDA (210nm) detectors.

4 Steam explosion of oilseed rape straw: establishing key determinants of saccharification efficiency at a pilot-scale

4.1 Chapter outline

Results from this chapter establish the likely determinants of biomass recalcitrance when processing of OSR straw at a pilot-scale. SE primarily targeted non-cellulosic components, resulting in a more readily hydrolysable substrate. Initial hydrolysis rate was limited by the amount of UA remaining in the substrate. The proportion of rapidly hydrolysable carbohydrate was most closely and positively related to Klason lignin abundance, which could not be explained only by the cumulative loss of non-cellulosic components. Final sugar yield most closely related to xylan removal from the substrate. Further experiments comparing milled and un-milled steam exploded straw demonstrated the influence of physical structure on hydrolysis rates and yields, particularly when using lower pretreatment severities. The results from this section provide a more detailed insight into the key determinants of saccharification efficiency of OSR straw than previously known and form a solid foundation for further work.

4.2 Introduction

A number of pretreatment techniques have been explored using OSR straw as a feedstock, including microwave (Lu *et al.*, 2011), wet oxidation (Petersson *et al.*, 2007), hydrothermal (Diaz *et al.*, 2010), alkali (Mathew *et al.*, 2011a), dilute acid (Castro *et al.*, 2011; Jeong & Oh, 2011a; Jeong *et al.*, 2010b; Lu *et al.*, 2009; Mathew *et al.*, 2011b) and autocatalytic 'popping' (Wi *et al.*, 2011). However, little is known about how OSR straw responds to SE.

SE has many potential advantages over other pretreatment methodologies including: lower capital costs, higher water efficiencies, high substrate loadings, better overall energy use and consequently, a lower potential product price (Conde-Mejía *et al.*, 2011; Kumar & Murthy, 2011). Therefore this methodology presents a potentially viable route for converting OSR straw into fuel ethanol, worthy of further investigation.

Recent work using SE as a pretreatment method demonstrated the effect of SE on the main polymer groups of OSR straw (Ryden *et al.*, 2014). Ryden *et al.,* (2014) concluded that SE primarily targets non-cellulosic components, resulting in a substrate enriched in cellulose, lignin and some residual pectic components. The results showed the effect this has on hydrolysis yield after 22 h of digestion using an excess cellulase dose. In this and similar studies, pretreatment effectiveness is determined under a set of predefined 'optimum' conditions (Conde-Mejía *et al.*, 2011; Kumar & Murthy, 2011).

However, truly optimal pretreatment conditions will be a compromise between processing costs and product yields (Mathew *et al.,* 2011a and b).

Pretreatment not only increases the overall saccharification yield, but also the efficiency of the cellulase in producing monomeric sugars (Arantes and Saddler, 2011). Although pretreatment is well known to have large effects on biomass composition, the exact mechanisms by which hydrolysis is improved are not well understood.

Lignocellulose is a heterogeneous composite made from carbohydrate and phenolic components and can be hydrolysed using multi-component cellulase cocktails. As a result, biomass hydrolysis does not follow typical enzyme kinetics. Instead, the final sugar yield is determined by a number of hydrolysis stages, limited by various interactions between multiple cellulases and the solid substrate (Bansal *et al.*, 2009).

The aim of this chapter was to understand the relationships between pretreatment severity, changes in substrate chemical composition, and the rate, extent and efficiency of cellulase binding and enzymatic hydrolysis. In doing so, we hoped to:

- Gain an overview of how *B. napus* straw responds to autocatalytic pretreatment at varying pretreatment severities.
- Extend our understanding of how SE improves the enzymatic hydrolysis of *B. napus* straw through the impact of pretreatment, including changes in substrate-cellulase interactions.
- Highlight the key limitations to saccharification efficiency of *B. napus* straw at various stages of hydrolysis.

To achieve these aims, OSR straw pretreated at different severities was assessed for key differences in initial cellulase binding and hydrolysis yields produced for varying cellulase doses. Rate-dependent disparities in hydrolysis were investigated by comparing the hydrolysis yields achieved at various points during saccharification. These parameters were then compared to the composition of the pretreated materials to gain a greater insight into the chemical determinants of saccharification performance.

4.3 Materials and methods

4.3.1 Materials

OSR straw (stems and empty pods) was sourced from Hemp Technology Ltd. Suffolk, UK (52°21'15.7"N 1°30'35.7"E). Straw was harvested in 2011, chipped into < 350 mm pieces, dust extracted and baled. The straw was stored as a 20 Kg bale in a dry, unheated room before analysis (< 1 year). No further size-reduction was conducted before SE.

The cellulase cocktail used in this chapter was Accelerase® 1500 (Genencor, UK) with a stock solution cellulase activity of 72.9 FPU/mL – calculated following Ghose *et al.,* (1987) and the protein concentration of 12.92 g/L calculated using a colorimetric bicinchoninic acid (BCA) assay (Section 4.3.5, below).

4.3.2 Steam explosion of OSR straw

A sample of OSR straw (1 Kg FWB) was steam exploded into hot water (6.6 L) at a range of pretreatment severities (9.01 to 26.94 bar, 180-230 °C, 10 min, R_o = 3.36-4.83) using a Cambi™ Steam Explosion Pilot Plant, as outlined in chapter 3. The pretreated biomass was collected and immediately filtered through a 100 µm nylon mesh bag in a low speed centrifuge. Both the liquor and insoluble pretreated solid were collected in their entirety and frozen immediately to prevent microbial growth (-40 °C). The SE unit was extensively rinsed between runs to prevent any carry-over of material to subsequent pretreatments.

4.3.3 Chemical composition of the pretreated solids and liquids

The water and ash content of the water insoluble residues was quantified following standard protocols outlined in Chapter 3. Klason lignin, sugar (including UA) compositions of freeze-dried steam-exploded materials were also determined following standard protocols (see Chapter 3).

The pH of the steam exploded liquors was determined directly after collection, using a digital pH meter (Jenway 3020, Jenway Ltd., UK). The liquor was syringe filtered (0.22 µm, Millex®, Millipore, USA) before monomeric sugar and degradation product concentrations were quantified by HPLC.

FT-IR spectra were also collected for each pretreated solid and liquor using methodology outlined in Chapter 3. The solid fraction was analysed using air as the reference medium and liquid fractions were analysed against ultrapure water (MilliQ).

4.3.4 Biomass preparation before saccharification assessment

Water-insoluble steam exploded biomass was homogenised by cryogenic milling (maximum impact frequency for 3 min, SPEX 6700 freezer/mill, Spex Industries, NJ) so consistent suspension of the solids could be achieved. All substrates, irrespective of pretreatment severity were reduced to sub-cellular fractions of similar size (mea*n* = 41-73 nm, media*n* = 39-85 nm) after freeze-milling - assayed using a LS 13 320 Laser Diffraction Particle Size Analyzer fitted with a universal liquid module (Beckman Coulter™).

The milled biomass was made into a slurry of known concentration by diluting with an appropriate amount of sodium acetate / acetic acid buffer (0.1M, pH 5) and hydrated in a sealed, continually stirred beaker overnight. Microbial activity was prevented using thiomersal (0.01% w/w). The slurry (1 mL aliquots) was transferred to individual 1.2 mL polypropylene cluster tubes held in a 96 well format (Corning, NY) using a 12-channel multipipette equipped with 200 µL, graduated, 'wide orifice' tips. This allowed simultaneous and even distribution of solids to each well (rows A-G). The final three columns (10-12) contained pre-weighed microtubes that were dried (105°C, overnight) and weighed again to determine the precise amount of dry biomass transferred to each well (typically ± 1.7% CV, n = 96). All plates included cellulase (row H) and substrate (column 1) controls. Plates were sealed using polyethylene cluster tube cap strips (Corning, NY), fixed firmly in place with adhesive tape and held in a plate shaking incubator (either 30 or 50 °C, Camlab Microtherm, 100 RPM).

4.3.5 Quantifying cellulase protein absorbance

The pretreated substrates held in a 96-well plate were rinsed six times using a Tecan Freedom Evo® 100 workstation (outlined in Chapter 3) – replacing 750 µL of the supernatant after centrifugation (2000 RPM, 2 min) with fresh buffer to dilute any soluble protein derived from the pretreated biomass.

Protein adsorption isotherms were created by adding varying concentrations (0-60 µL/ mL) of vacuum filtered (GF/C, Whatman, UK) cellulase (Accellerase 1500) to 1 mL slurry containing ca. 20 g/L rinsed substrate (DWB) at < 4 °C on ice. After 2 h equilibration at 4°C the slurries held in a 96-well format were centrifuged (2000 RPM, 2 min) and a sample of the supernatant (10 µL) was dispersed in 200 µL of BCA reagent (400 µL Copper (II) sulfate solution (4%) in 20 mL bicinchoninic acid solution). Sample absorbance (562 nm) was quantified after incubation (1 h, RT) - comparing samples to a set of bovine albumin standards (0-1 mg/mL, n = 12). The amount of protein adsorbed to the substrate was calculated by subtracting soluble protein concentration from the amount present in a cellulase control (cellulase solution without substrate).

4.3.6 Time-course enzymatic saccharification of steam exploded material

Time-course data was collected by hydrolysing 1% D.W. equivalent of un-milled or freeze-milled pretreated solid in 20 mL sodium acetate/acetic acid buffer (0.1M, pH 5, 0.01% thiomersal) held in a shaker plate incubator (50 °C, 150 RPM). Accellerase® 1500 (Genencor®, NY, USA) was added to the equilibrated solutions at a cellulase dose of 0.5 mL/g substrate (ca. 36 FPU/g). Digestions were conducted in three independent replicates, quantifying the amount of RS released at several time-points during incubation (removing 9 µL, or < 0.05% v/v, at each time point to prevent a significant change in volume through sampling).

The abundance of RS in each biomass hydrolysate was quantified using a multiplexed DNS method (optimised in Chapter 6). In this case, most accurate quantification was gained by adding 144 µL of DNS reagent to 36 µL of sample. Products were quantified against a series of Glc standards (0-10 g/L, n = 12).

Racked tube-plates were centrifuged (2000 RPM, 2 min) and a sample of the supernatant boiled (100 µL) in a sealed PCR plate to denature the cellulase (100 °C, 5 min). A sample of the supernatant was diluted to within a readable range (0-2 g/L) and 5 µL was dispersed in GOPOD reagent (195 µL, Megazyme International Ltd., Ireland) in a microtitre plate. The amount of Glc in each hydrolysate was quantified after 20 min of incubation (50°C) by comparing product absorbance against a Glc dilution series (0 – 2 g/L, n = 8).

4.3.7 Statistical analysis

All statistical analysis was conducted using Genstat v. 13 (VSN International, Ltd.) Cellulase adsorption capacities were calculated using Equation 3 where ‘β’ and ‘ρ’ are constants that describe the rate of change in cellulase absorbance as they reach the maximum adsorption capacity of the substrate ‘α’.

$$y = \alpha + \beta\rho^{x}$$

Equation 2

Time course data was fitted to Equation 4 where ‘A’ is the plateau value (% g/g) of the initial, exponent part of the curve and ‘t’ is hydrolysis time. The parameter ‘R’ describes the rate of decay of the initial part of the curve, beginning at an offset ‘B’. The gradient of the second, linear section is described by parameter C. From these parameters, the initial rate of saccharification (d_y/d_x) was estimated (Equation 5) – derived with the

help from Dr Henri Tapp. These parameters quantitatively describe three biologically relevant parameters.

$$y = A + BR^t + Ct$$

Equation 3

$$d_y/d_x = BlnR + C$$

Equation 4

Unless otherwise stated, all experiments were conducted in triplicate and means presented ± 1 S.D.

4.4 Results and discussion

4.4.1 Steam explosion recoveries

After steam exploding OSR straw (1 kg) into hot water (6.6 L) at a range of severities, the resulting material was fractionated into solid and liquid phases. SE of OSR straw into water at increasing severities reduced the fibrous structure, which retained less water producing denser water insoluble solids (Table 3) as is typical when pretreating biomass (Díaz *et al.*, 2010; Kabel *et al.*, 2007; Lu *et al.*, 2009; Mathew *et al.*, 2011b). SE of OSR straw at increasing severities also caused a darkening of the pretreated material (Figure 15).

The mass of solids recovered after SE generally decreased at higher pretreatment severities as an increasing proportion of the biomass was liquefied, particularly at higher temperatures (-13-46% DM, Table 3). Although, some of this reduction in mass can be attributed to the liquefaction of non-cellulosic materials, imperfect separation of steam and solid in the SE apparatus caused a loss of some material within the machine. This is an unfortunate but unavoidable difficulty in using pilot-scale methodology to conduct discrete experiments (Rocha *et al.*, 2012). Under continuous operation, significant losses of material would not occur.

Pretreatment severity	Solid matter recoveries		
	Wet biomass (g/kg original)	Dry matter (% PT)	DM recovery (% original)
180°C	2207	34.5	86.4
190°C	2066	31.1	73.9
200°C	1710	33.5	65.1
210°C	1484	42.1	71.9
220°C	1321	42.0	63.0
230°C	1295	40.5	60.0

Table 3 - Recovery of solid material after SE (1 Kg, FWB) at varying severities for 10 min.

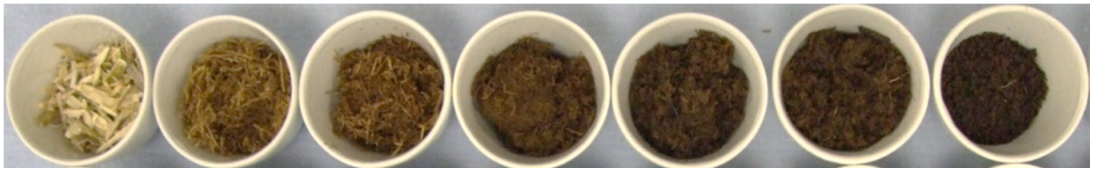


Figure 15 - Changes in physical appearance between water-insoluble fractions of OSR straw steam exploded under varying conditions. From left to right: Untreated, 180, 190, 200, 210, 220 and 230 °C. All substrates were pretreated for 10 min temperature.

4.4.2 Chemical composition of the pretreated solids

The untreated straw used in this study contained Glc (28.72 ± 1.42%), Klason lignin (20.1 ± 3.54%), Xyl (12.64 ± 0.69%), ash (4.9 ± 0.02%), UA (4.7 ± 0.16%), Man (1.8 ± 0.10%), Gal (0.9 ± 0.06%), Ara (0.7 ± 0.10%), Rha (0.5 ± 0.02%) and Fuc (0.1 ± 0.02%) DWB, in keeping with earlier reports (Castro *et al.*, 2011; Díaz *et al.*, 2010; Jeong & Oh, 2011b; Lu *et al.*, 2011; Mathew *et al.*, 2011a; Mathew *et al.*, 2011b; Petersen *et al.*, 2009; Wi *et al.*, 2011). The moisture content of the untreated straw was 12.0 ± 1.46% FW.

In accordance with previous work, SE of OSR straw resulted in the removal of non-cellulosic polysaccharides from the solid fraction (Ryden *et al.*, 2014). Xyl, UA, Man, Gal, Ara and Fuc were all removed with diminishing effectiveness as pretreatment severity increased. At the most severe pretreatment (230°C, 10 min), only ≈ 12 g/Kg original straw of non-cellulolytic sugars remained in the solid fraction, compared to 195 g/kg in the untreated straw (FWB). More robust CW components, cellulose and lignin, were concentrated in the insoluble fraction (Figure 16). The abundance of Klason lignin found in the residue not only increased as a proportion of the insoluble residue, but also in overall abundance, suggesting pseudo-lignin is produced from carbohydrate components at severities exceeding 200°C (Kumar *et al.*, 2013).

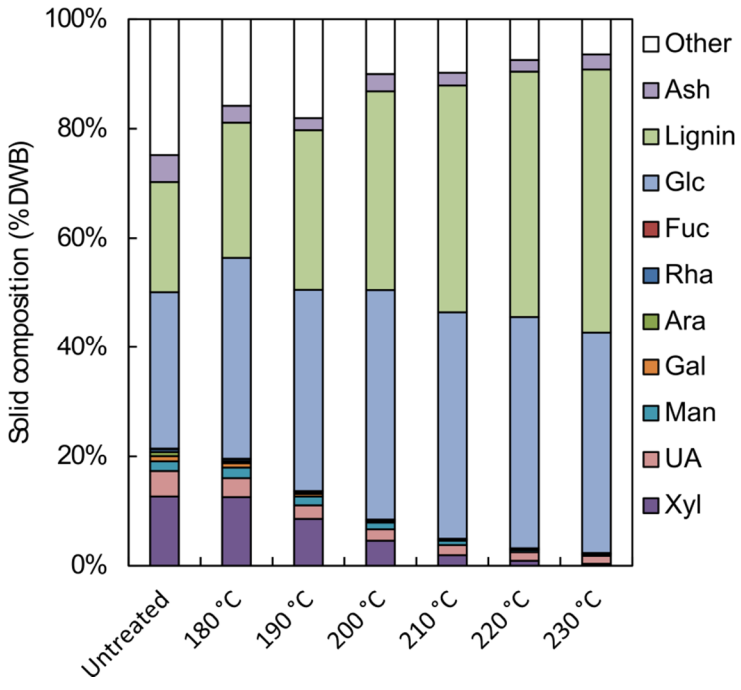


Figure 16 - Composition of OSR straw steam exploded at varying severities (180-230 °C, 10 min). Components are listed in descending order – demonstrating the loss of non-cellulosic sugars (bottom) and unquantified components (uppermost). These losses resulted in a concentration of Glc and Klason lignin in the SE material.

4.4.3 Chemical analysis of pretreated liquor

Ryden *et al.,* (2014), did not quantify the amount of fermentation inhibitors released during SE. However, a potential caveat of SE is that comparatively high amounts of fermentation inhibitors could be produced during pretreatment (Kumar & Murthy, 2011). Therefore the SE liquors were analysed in this study. The pH of the SE liquors decreased with increasing pretreatment severity, but remained at approximately pH 4 at after pretreatment above 200°C. The pH of the liquors generated by pretreatment at between 180-230 °C were: 4.75, 4.33, 4.06, 4.01, 4.00, 4.00 and 3.93 respectively. This is likely to reflect the comparatively high organic acid concentrations in the liquid fraction, which contained predominantly sugar degradation products and few monomeric sugars (Table 4). Although the concentrations of organic acids and furfural derivatives in the pretreated liquors produced in this study are unlikely to inhibit fermentative organisms, degradation products could have a negative effect on downstream processes if conducted at high substrate concentrations (Rasmussen *et al.*, 2013). It was therefore important that these compounds were quantified.

	180°C	190°C	200°C	210°C	220°C	230°C
Organic acids and furfural derivatives						
Formic acid	7.9	17.2	23.9	28.5	27.5	27.2
Acetic acid	6.4	17.4	27.8	35.9	37.1	38.5
5HMF	0.2	0.4	0.8	1.1	1.2	1.3
2FA	0.7	1.7	3.4	3.9	3.4	2.3
Monomeric sugars						
Glucose	0.8	2.5	2.8	2.7	2.7	3.5
Cellobiose	2.5	3.3	3.3	2.3	2.1	2.1
Xylose	4.4	4.8	5.2	3.7	2.4	3.9
Galactose	2.5	8.3	3.8	1.0	1.0	1.0
Arabinose	3.0	3.0	5.2	2.8	3.3	3.0
Mannose	1.2	1.0	0.9	1.3	1.0	0.9
Totals						
Monomeric sugars	14.4	22.9	21.2	13.7	12.5	14.4
Fermentation inhibitors	15.2	36.7	55.9	69.4	69.2	69.3
Total	29.6	59.6	77.1	83.1	81.7	83.7

Table 4 - Mass of soluble products in the pretreatment liquor (g/kg original straw) after SE at varying severities for 10 min. All compositional data was measured with an RSD ± < 7%.

The most abundant compounds detected in the liquors were acetic and formic acid, which increased with pretreatment temperature up to approximately 210°C (Table 4). The abundance of acetic and formic acid produced using SE is comparable to those generated when hydrothermally pretreating OSR straw (Díaz *et al.*, 2010). Dilute acid pretreatment generally produces higher acetic acid yields – 44-61 g / Kg original straw depending on pretreatment conditions (Castro *et al.*, 2011; Lu *et al.*, 2009). In this study, maximum 2-furfuraldehyde (2FA) concentrations resulted after pretreatment at 210°C, but were reduced at higher severities. Steam exploding OSR into hot water resulted in significantly less 2FA in the pretreatment liquor than other comparable pretreatments (Díaz *et al.*, 2010; Jeong *et al.*, 2010a). Vaporisation of these compounds at high pretreatment temperatures is the most likely reason for this. The concentration of 5-hydroxymethylfurfural (HMF) retained in the SE liquors increased with pretreatment severity but was limited to < 1.29 g / Kg original straw. This is similar to the amount produced via acid-catalysed hydrothermal pretreatment (Jeong & Oh, 2011a) but considerably less than other dilute acid or uncatalysed hydrothermal pretreatments where ca. 25 g of HMF is produced/kg of original OSR straw (Castro *et al.*, 2011; Díaz *et al.*, 2010). Whether the quantity of fermentation inhibitors produced during SE is acceptable will largely depend on subsequent process-parameters and the sensitivity of the organism selected for fermentation (Rasmussen *et al.*, 2013).

Only small quantities of intact monomeric sugars were present in the hydrolysates (< 3.4 g/L in total) (Table 4). However, a substantial fraction of solubilised oligomeric sugars may be present in the liquors thereby evading quantification by HPLC (Kabel *et al.*, 2007).

4.4.4 Fourier transform infrared spectra of pretreated solids and liquids

Ryden *et al.,* (2014) showed that steam exploding OSR straw at increasing severities resulted in spectral changes consistent with the loss of non-cellulosic polysaccharides, particularly xylans, which are the principle target of this pretreatment methodology. Another significant mechanism of the pretreatment was to remove some, but not all of the pectic components, even after SE under severe conditions (Ryden *et al.*, 2014). The transfer of soluble components to the pretreatment liquor was not explored by Ryden *et al.,* (2014). Therefore, FT-IR spectra were collected from all pretreated solid and liquid fractions to provide further information on chemical composition not highlighted using more specific techniques (Figure 17). FT-IR was specifically selected as spectral changes allow the detection of a range of soluble components, including oligomers of varying chemistries, without *a priori* knowledge of the sample contents.

Spectral changes related to the solid component were similar to those observed by Ryden *et al.,* (2014) (Figure 17). Spectra collected from the SE liquors showed an initial increase in the abundance of solubilised carbohydrates (1000–1150 cm^{-1}) which declined at higher severities. The intensity of lignin associated peaks (1250, 1510 and 1590 cm^{-1}) also increased with pretreatment severity, suggesting a release of water soluble phenolic compounds to the liquors (Figure 17).

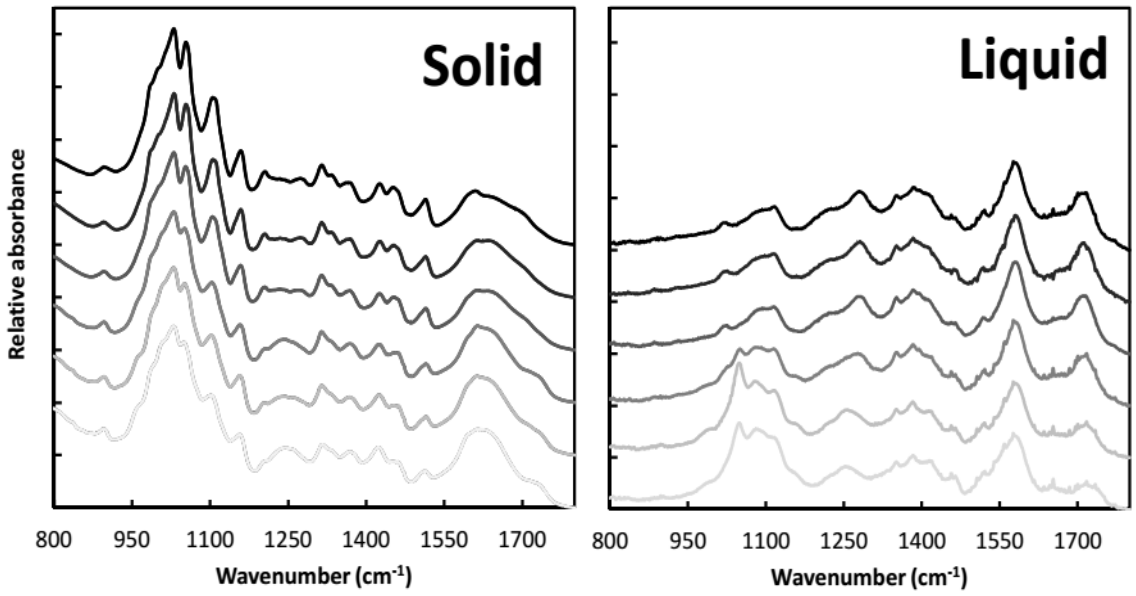


Figure 17 - FT-IR spectra for solid and liquid fractions of OSR straw steam exploded into hot water at a range of pretreatment severities from 180 °C, 10 min (light grey) to 230 °C, 10 min (black), increasing in 10 °C increments.

4.4.5 Differences in sugar yields following pretreatment and hydrolysis

Pretreated and washed substrates were saccharified for 96 h using Accellerase 1500 (36 FPU/g substrate) (Figure 18). Circa 100% of the Glc available for hydrolysis could be released from the pretreated material using conditions > 220°C, 10 min (Figure 18). A maximum of 86.1% of the total Glc present in the original material could be recovered after hydrolysis of OSR straw after pretreatment at 210°C, 10 min. The optimal yields are comparable to those obtained after pretreating with dilute acid (81%; Mathew *et al.*, 2011a) or alkali (100%; Mathew *et al.*, 2011b).

SE also produces slightly higher yields than other autocatalytic pretreatments – such as hydrothermal pretreatment, which can achieve overall yields of 67.6% Glc conversion under optimum conditions (Díaz *et al.*, 2010). The addition of dilute acids may reduce the residence time of hydrothermal treatments, but could result in a slight lowering of yields (61.5-63.7%)(Castro *et al.*, 2011; Lu *et al.*, 2009). High yields have been achieved using a similar method to SE, 'popping' water-soaked OSR straw by heating to 220°C in

closed vessel, rapidly releasing the pressure and finely milling the residue (Wi *et al.*, 2011).

When considering the effectiveness of pretreatments, in terms of energy use or processing costs, optimal conditions in regard to glucan conversion are often compared (Conde-Mejía *et al.*, 2011; Kumar & Murthy, 2011). However, optimal pretreatment conditions will be a compromise between processing costs and product yields (Mathew *et al.*, 2011a and b). To compare results between pretreatments, beyond total Glc conversion, more detailed information is needed that reflects the flexibility in available pretreatment conditions.

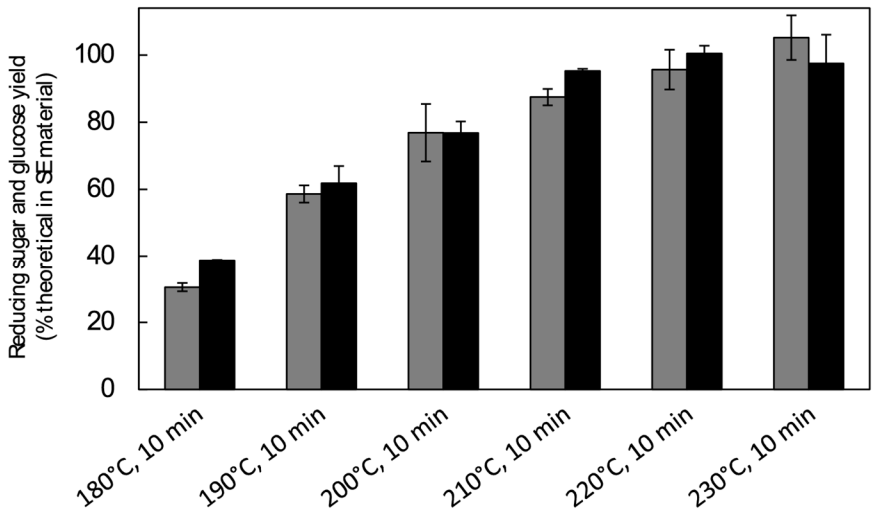


Figure 18 - Variations in Glc and RS yields (grey and black respectively) released from OSR straw, steam exploded at varying conditions

4.4.6 **Effects of steam explosion on cellulase absorbance.**

The initial cellulase binding capacity of a material is thought to be one of the most important parameters describing saccharification efficiency (Bansal *et al.*, 2009). In order to determine the initial cellulase absorption capacity of steam exploded OSR straw, varying concentrations of cellulase protein (Accelerase 1500) was added to 18.2 ± 0.70 g/L (DWB) of milled substrate solutions and equilibrated at 4 °C, 2 h. The amount of cellulase protein that bound to the substrate was quantified by measuring the loss of protein from the liquor. Under these conditions, saccharification of the substrates was limited to < 1.65% (data not shown), allowing accurate assessment of the relative initial binding capacity of each substrate (Figure 19).

SE significantly increased the initial cellulase protein binding capacity of OSR straw (α) from ≈10 mg protein/g untreated straw to ≈90mg/g pretreated biomass (Table 5).

Increasing pretreatment severity from 180 to 190 °C, 10 min produced the greatest increase in adsorption capacity. OSR straw steam exploded at between 200 °C to 220 °C achieved similar binding capacities (ca. 70-75 mg/g), therefore a similar initial rate of hydrolysis for all substrates steam exploded at these temperatures when > 70 mg/g of cellulase was available to bind to the substrate. The absorption capacity of the substrates only increased further by using very high temperatures in excess of 230 °C.

The cellulase adsorption capacities of steam exploded OSR straw pretreated at the highest severity are similar to those observed from lignocellulosic materials from monocotyledonous plants such as corn stover (Kumar & Wyman, 2009) and switchgrass (Shi *et al.*, 2011) pretreated using various methods (typically 90-170 mg/g substrate). These studies show that different pretreatment methods produce substrates with varying cellulase adsorption capacities (Kumar & Wyman, 2009; Shi *et al.*, 2011). As one might expect, results from this study show that initial cellulase adsorption capacity is also dependent on pretreatment severity.

All of the substrates pretreated at severities higher than 190 °C achieved similar degrees of cellulase absorbance per cellulase added (ρ), showing that the absorbance of cellulase below 70 mg cellulase/g substrate is limited by cellulase availability and not substrate-specific properties (Figure 19, Table 5).

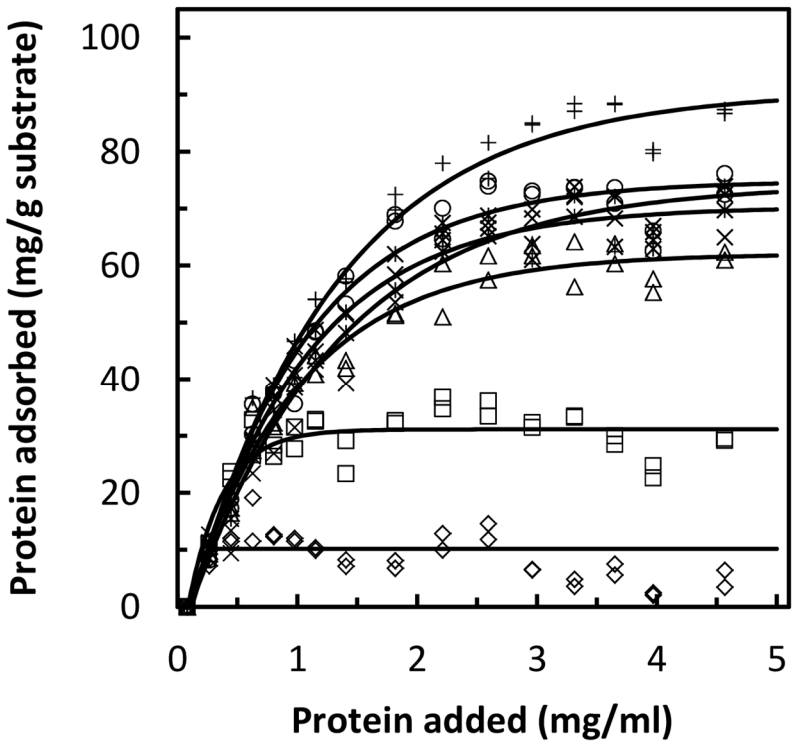


Figure 19 - Adsorption isotherms of cellulase protein from Accellerase 1500 to the water insoluble fraction of steam exploded OSR straw pretreated at 180 °C (□), 190 °C (△), 200 °C (x), 210 °C (✳), 220 °C (○) or 230 °C (+). Non-steam exploded material was also included (◇).

Pretreatment conditions	$\rho \times 10^{-2}$	Absorption capacities (α)	
		mg/g substrate	mg/g cellulose
Untreated	-	9 ± 0.7	31 ± 1.8
180 °C, 10 min	17 ± 4.9	33 ± 0.9	90 ± 2.6
190 °C, 10 min	51 ± 3.0	70 ± 2.2	192 ± 6.1
200 °C, 10 min	58 ± 2.7	84 ± 2.9	198 ± 7.8
210 °C, 10 min	51 ± 2.7	77 ± 2.2	184 ± 5.9
220 °C, 10 min	50 ± 2.4	82 ± 2.1	196 ± 5.9
230 °C, 10 min	60 ± 2.2	108 ± 3.3	265 ± 9.0

Table 5 – Differing adsorption parameters adding varying concentrations of Accellerase 1500 to OSR straw steam exploded at a range of pretreatment severities. Adsorption data was fitted to the following equation: $y = \alpha + \beta\rho^x$ and parameters expressed ± 1 Standard error

4.4.7 Hydrolysis yields at varying cellulase doses

To directly determine the cellulase dose required to optimally hydrolyse steam exploded OSR straw, each substrate (≈ 3% DW) was hydrolysed at a range of cellulase concentrations (0-1 mL or 0-73 FPU /g substrate, n = 19) at 30 °C and 50 °C. These temperatures were chosen as 50°C is the optimum temperature for cellulase activity and 30°C is a more suitable temperature for simultaneous saccharification and fermentation (Olofsson *et al.*, 2008). The RS concentrations of the supernatant were quantified after 24 h of digestion (Figure 20).

Increased severity resulted in a decrease in cellulase required to achieve the same yield, but reasonable yields (> 70%) were only achieved after pretreatment at 200-220 °C, 10 min at cellulase doses > 30 FPU/g (Figure 20). Interestingly, this cellulase dose is similar to the quantity required to optimally hydrolyse hydrothermally pretreated OSR straw (López-Linares *et al.*, 2013b). Together these studies indicate that commercial cellulases, which typically require <20 FPU/g to hydrolyses SE wheat straw (Horn *et al.*, 2011), are poorly adapted to hydrolysing OSR straw.

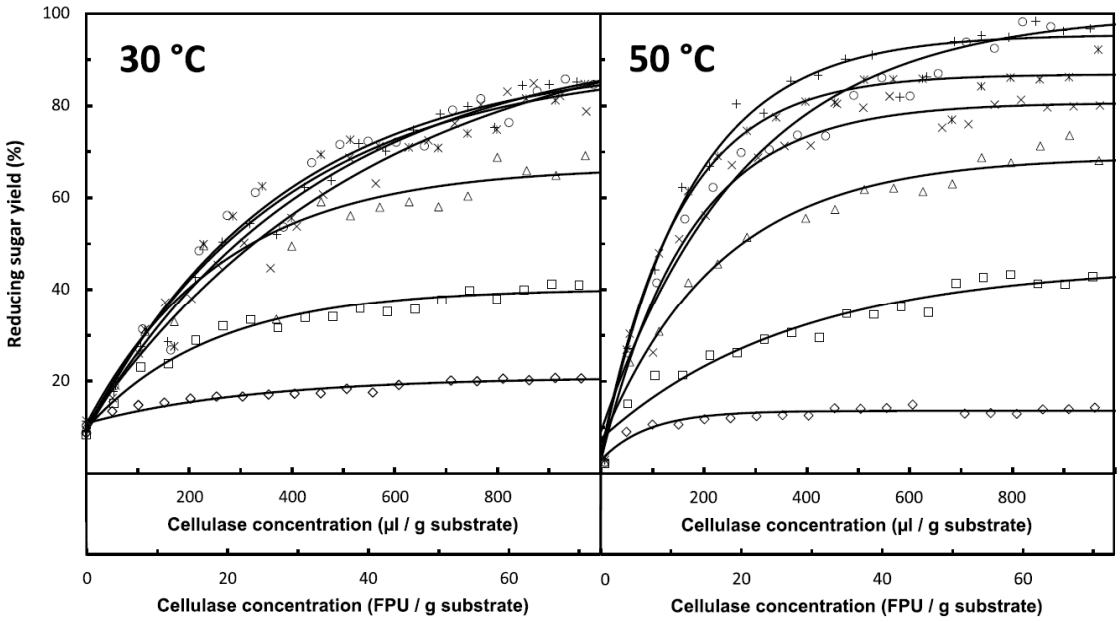


Figure 20 – Saccharification yields of untreated (◇) and OSR straw, SE at varying severities; 180 °C (□), 190 °C (△), 200 °C (x), 210 °C (✳), 220 °C (○) or 230 °C (+) and hydrolysed using varying cellulase concentrations (0-73 FPU/g substrate) at two temperatures (30 °C or 50 °C) for 24 h.

4.4.8 Time-dependent hydrolysis parameters

The rate of fermentable sugar production from lignocellulose also differs depending on hydrolysis time - with a rapid initial phase of hydrolysis followed by a slower phase until the maximum sugar yield is reached (Arantes & Saddler, 2011; Mansfield *et al.*, 1999).

Pretreated OSR straw was enzymatically hydrolysed (20 mL, 1% substrate DW, 50 °C, ca. 36 FPU/g Accellerase 1500) for up to 48 h (Figure 21). The time-course data was fitted to a right-handed, line plus exponential curve to establish three biologically relevant parameters - (1) the initial rate of hydrolysis (d_y/d_x); (2) the proportion of rapidly hydrolysable carbohydrate (A) and (3) the rate of hydrolysis of slowly-hydrolysable carbohydrate (C) (Table 6). This established quantitative, time-dependent differences in saccharification between the substrates (Table 6).

The initial rate of hydrolysis (d_y/d_x) generally increased with pretreatment severity until ≈ 190°C, after which similar hydrolysis rates were achieved (20-25%/h). After the initial hydrolysis period, the rate of hydrolysis slowed considerably to approximately 0.1-0.9%/h until the maximum hydrolysis yield was achieved (Table 6). The majority of hydrolysis was completed during the first 12-24 h of incubation.

Despite the similarity in initial hydrolysis rates of straw steam exploded at > 190 °C, the proportion of rapidly hydrolysable material (A) differs more notably between the substrates produced. For example, although un-milled substrates steam exploded at 200°C and 210°C have similar initial hydrolysis rates (22-25%/h), the proportion of the carbohydrates found in the pretreated material that could be hydrolysed at this rate (A) is almost double for material pretreated at 210°C (Table 6). Therefore, the proportion of rapidly hydrolysable material (A) is an important time-dependent determinant of pretreatment effectiveness.

The initial rates of hydrolysis (dy/dx) calculated for each substrate correlated with their corresponding initial cellulase binding capacities (Figure 22, A). This suggests that most of initial cellulase binding is productive, binding to hydrolysable components. Previous work has shown that corn stover pretreated using different methodologies also gave different initial cellulase binding capacities, correlated well with the hydrolysis yield achieved after 24 h (r^2 = 0.75 – 0.9) (Kumar & Wyman, 2009). The results presented here show that the relationship between initial cellulase adsorption capacity and initial hydrolysis rate also applies to OSR straw pretreated using the same methodology, but at varying pretreatment severities (Figure 22, A).

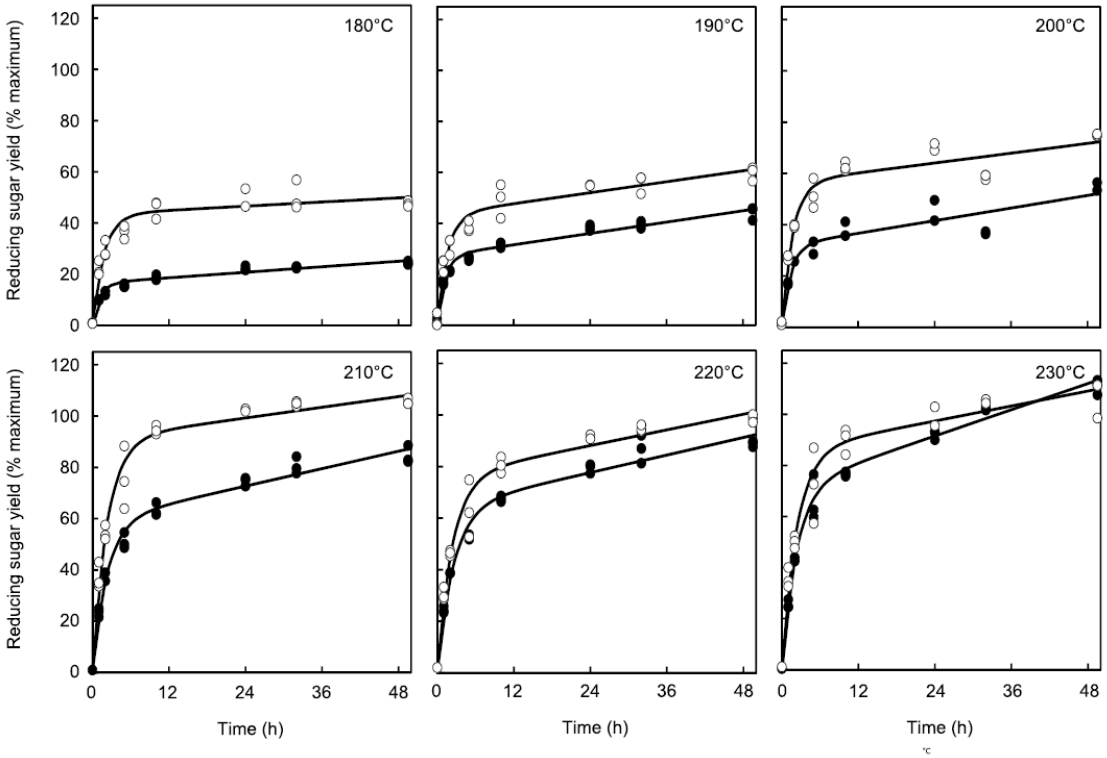


Figure 21 - Comparison of un-milled (●) and freeze-milled (○), steam exploded OSR straw hydrolysed using 1% substrate (DW equivalent) at 36 FPU/g. Yields are expressed as a percentage of the theoretical maximum reducing sugar that could be released from the insoluble pretreated material.

Pretreatment conditions	Hydrolysis parameters					
	Initial rate - d_y/d_x (%RS/h)		Initial plateau - A (%RS)		Secondary rate - C (%RS/h)	
	UM	FM	UM	FM	UM	FM
Untreated	4.8 ± 1.4	9.2 ± 1.4	7 ± 2.3	15 ± 2.2	0.14 ± 0.08	0.17 ± 0.07
180 °C, 10 min	13.3 ± 1.4	22.6 ± 0.6	16 ± 2.1	43 ± 2.5	0.18 ± 0.07	0.14 ± 0.08
190 °C, 10 min	20.5 ± 1.1	24.4 ± 0.7	27 ± 2.1	43 ± 2.4	0.37 ± 0.07	0.36 ± 0.08
200 °C, 10 min	22.1 ± 1.1	30.9 ± 0.5	32 ± 2.7	56 ± 2.5	0.41 ± 0.09	0.32 ± 0.08
210 °C, 10 min	25.1 ± 0.4	38.5 ± 0.3	59 ± 2.8	91 ± 2.9	0.57 ± 0.09	0.35 ± 0.09
220 °C, 10 min	24.7 ± 0.4	30.8 ± 0.4	64 ± 3.0	76 ± 2.9	0.57 ± 0.09	0.35 ± 0.09
230 °C, 10 min	31.4 ± 0.4	35.2 ± 0.3	71 ± 2.8	86 ± 2.9	0.86 ± 0.09	0.49 ± 0.09

Table 6 - Time course parameters describing the saccharification of unmilled (UM) and freeze-milled (FM) steam exploded OSR straw. Hydrolysis data was fitted to the equation: y = A+B*Rx+Cx and parameters expressed ± 1 standard error.

4.4.9 Physical hindrance to enzymatic hydrolysis following steam explosion

To establish the effect of physical structure on the digestibility of the steam exploded substrates, steam exploded materials were also freeze-milled and hydrolysed under the same conditions (Figure 21). Milling increased the initial hydrolysis rate (d_y/d_x) and the portion of easily hydrolysable material (A) of all steam exploded materials, however milling caused greater increases in these parameters when lower pretreatment severities were used (Table 6). Reducing the particle size of steam exploded substrates pretreated at > 220 °C only increased the proportion of easily hydrolysable material by ca. 20% compared to 53% when material steam exploded at 210°C was finely milled.

Finely milling lignocellulosic materials tends to produce higher saccharification yields, as the surface area accessible to cellulase in increased (Lindedam *et al.*, 2010). The findings presented here concur with Zeng *et al.,* (2007) who found that finely milling corn stover significantly increased Glc yields from untreated materials, but not from biomass sufficiently pretreated using liquid hot water (190°C, 15-20 min)(Zeng *et al.*, 2007). Particle size reduction increases hydrolysis yields from insufficiently pretreated OSR straw, but the effect of further particle size reduction is negligible when biomass pretreated at high severities (> 220 °C). This confirms that particle size after pilot-scale pretreatment is an important factor in determining the hydrolysis efficiency of the resulting substrate (Lindedam *et al.*, 2010).

4.4.10 Relationship between substrate chemistry and saccharification parameters

The importance of substrate composition in determining hydrolysis efficiency was investigated by comparing time-dependent hydrolysis parameters with the abundance of CW components (Figure 22, B-D). Firstly, the results indicate that SE of OSR straw increases the accessibility of cellulose to hydrolysis (initial cellulase binding capacity) and consequently the initial rate of hydrolysis primarily by the removal of UA-containing components which are found in close association with more abundant structural carbohydrates. This is indicated by a strong linear negative correlation between UA abundance in the pretreated material and initial hydrolysis rate (Figure 22, B). This observation is consistent with the most recent dicot CW models where pectins are the predominant polysaccharide coating cellulose microfibrils (Cosgrove & Jarvis, 2012). Galacturonans also form close associations with XGs – the other main non-cellulosic polysaccharide closely associated with cellulose microfibrils (Marcus *et al.*, 2008). Moreover, de-methyl-esterified HG presence is known to increase the recalcitrance of other dicotyledonous plant material – resulting in lower Glc yields after processing (Lionetti *et al.*, 2010).

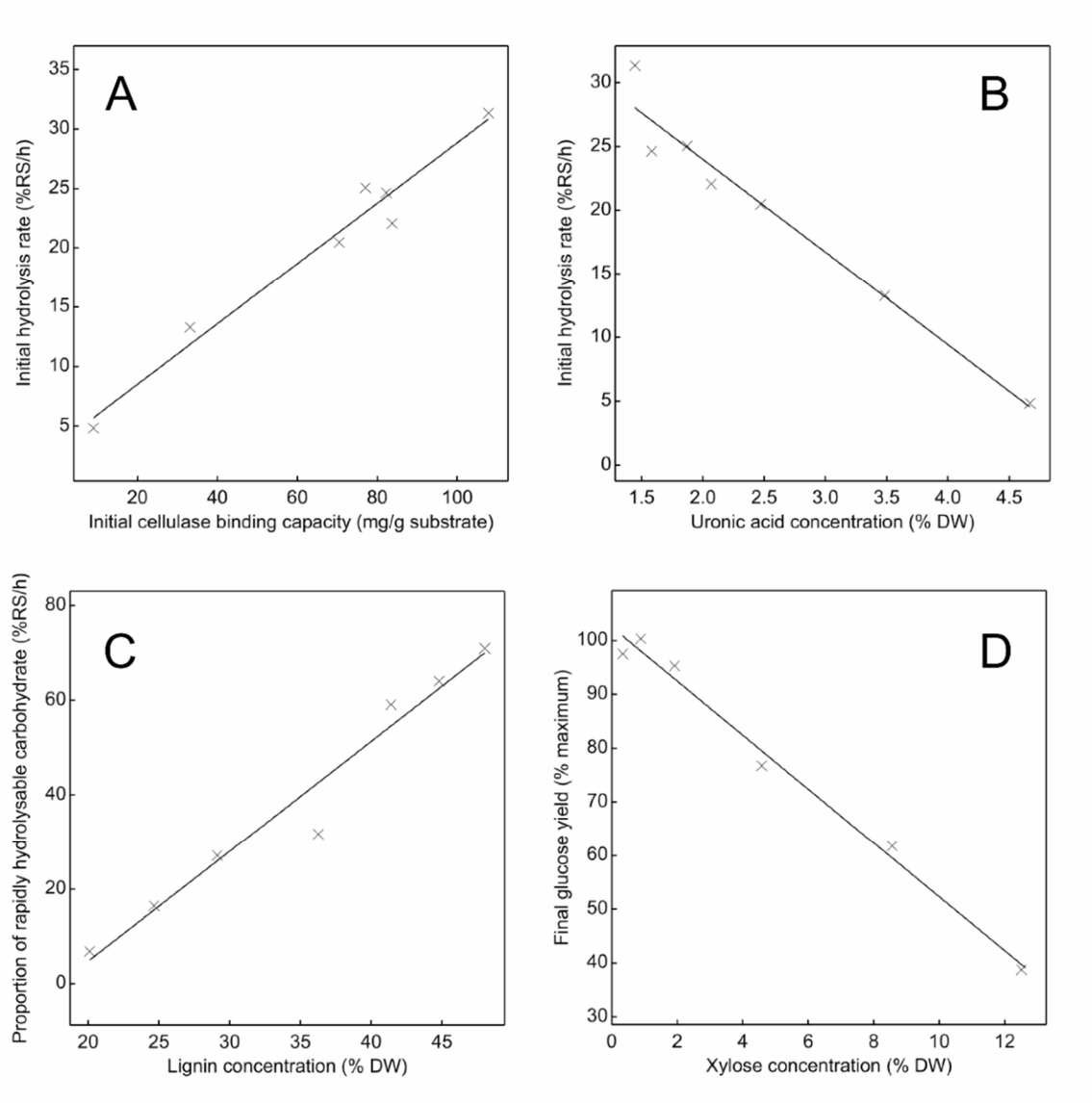


Figure 22 - Linear correlations between key hydrolysis parameters and the chemical composition of the pretreated material for OSR straw pretreated at different severities. Strong correlations were observed between the cellulase binding capacity and initial hydrolysis rate of a substrate (A). Chemical components that correlated most strongly with the initial hydrolysis rate (B), proportion of rapidly hydrolysable carbohydrate (C) and final glucose yield (D) are also displayed. The R^2 of regression lines A, B, C and D were 0.97, 0.96, 0.96 and 0.98 respectively.

In SCW materials, Glucuronoxylan also contain GlcA. However, Xyl is almost completely removed following SE while a significant portion of pectic components remain in the pretreated residue (Ryden *et al.*, 2014). It is therefore likely that the observed correlation between initial cellulase binding sites and UA abundance relates to the retention of pectic components in the insoluble residue.

The second hydrolysis parameter explored was the proportion of rapidly hydrolysable carbohydrate (A). This parameter quantifies the amount of carbohydrate (mostly cellulose) that can be hydrolysed at a rate approaching the initial rate - typically complete during the first 8 h. Interestingly, a strong positive linear correlation was observed between Klason lignin abundance and proportion of rapidly hydrolysable material in the substrate - parameter 'A' (Figure 22, C), which could not be explained by the cumulative loss of other CW components.

The parameter 'A' describes the point at which the hydrolysis rate changes from one close to the initial rate (≈ 20% RS/h), to a significantly slower hydrolysis phase (≈ 0.5% RS/h). This change in hydrolysis rate is likely to coincide with the near complete removal of rapidly hydrolysable CW components. After this point, the remaining cellulose microfibrils may begin to aggregate - therefore increasing the resistance of the remaining substrate to saccharification (Ishizawa *et al.*, 2009). The positive correlation between lignin abundance and this parameter suggests that the more lignin or pseudo-lignin present in the substrate could prevent this aggregation – allowing rapid hydrolysis to continue for longer. This is consistent with data from other researchers that have noted that delignification of substrates that have already had xylan removed, result in poorer hydrolysis performance and therefore lower Glc yields (Ishizawa *et al.*, 2009).

Finally, the maximum Glc yields achieved from each substrate were correlated against CW component abundance in the pretreated residues. In this case, Xyl, Gal and UA abundance in the pretreated substrate negatively correlated with maximum Glc yields after enzymatic hydrolysis, with Xyl abundance correlating most strongly (Figure 22, D) This concurs with previous work demonstrating that non-cellulosic β-glycan (hemicellulose) removal is important in gaining maximum yields from lignocellulose (Rollin *et al.*, 2011; Zhu *et al.*, 2011). The abundance of hemicellulose remaining in the material is likely to limit accessibility to the residual carbohydrate in the partially hydrolysed sample, therefore determining the final yield.

When using autocatalytic pretreatments (heating without the addition of chemical catalysts) such as SE, enzymatic digestibility is closely linked to xylan removal. This relationship is approximately linear to 80% conversion (Moxley *et al.*, 2012). Moxley *et al.,* (2012) suggested that saccharification yield is closely related to hemicellulose removal until completion. After this point, chemical modifications to lignin become critical in increasing enzyme digestibility further (Moxley *et al.*, 2012). One might argue that pectin, hemicellulose and lignin all play important roles in maintaining cellulase access to cellulose at various stages of hydrolysis. Achieving optimum hydrolysis performance is therefore a compromise between the degree of biomass

opening and substrate integrity during hydrolysis. Further work is required to establish strategies of hydrolysing lignocellulose completely. Maintaining the structural integrity of the biomass during hydrolysis may be vital in achieving this goal.

Although the cellulase cocktail used in this study is likely to contain small quantities of pectinases and xylanases, the results would suggest that further additions of these accessory enzymes may enhance hydrolysis of OSR straw, particularly when low pretreatment severities are used (Banerjee *et al.*, 2011). Modelling time-dependent limitations to hydrolysis using specific cellulase cocktails identifies this need and provides targets for cellulase cocktail improvement.

4.5 Conclusions

SE of OSR straw caused extensive removal of non-cellulosic polysaccharides from the pretreated solid. Consequently, glucan and lignin were concentrated in water insoluble solid. Solubilised components were generally degraded to furfural derivatives and formic acid, with few monomeric sugars retained in the pretreatment liquors.

The optimum pretreatment severity in terms of Glc yield was ≈ 210 °C, 10 min. Under these conditions, ca. 247 g of Glc per Kg of untreated OSR straw, an equivalent of 95% g/g Glc in pretreated substrate or ≈ 86% g/g Glc in original substrate, can be achieved with a high cellulase dose (36 FPU/g, 96 h). Hydrolysing the pretreated material using various cellulase doses showed a relatively high cellulase dose was needed for optimum hydrolysis (30-40 FPU/g substrate). This suggests that commercial cellulases are poorly adapted to hydrolyse SE OSR straw.

Here, three quantitative and biologically relevant hydrolysis parameters were determined for each substrate, reflecting temporal differences in saccharification potential between substrates. The initial rate of hydrolysis closely correlated with the initial cellulase binding capacity of the substrates. Typically, SE OSR straw was initially hydrolysed rapidly (20-30%/h). After these rapidly hydrolysable carbohydrates had been saccharified, the hydrolysis rate slowed significantly (typically 0.3%/h) until the maximum yield was reached. The point at which this change in rate occurred indicated the proportion of rapidly hydrolysable carbohydrate.

Cellulase binding capacity and initial hydrolysis rate correlated most strongly with UA abundance, implicating pectic components in reducing initial hydrolysis rate. The proportion of carbohydrate that could be hydrolysed at a rate close to the initial rate was positively correlated with lignin abundance, suggesting a role in maintaining biomass accessibility. The maximum hydrolysis yield correlated most strongly with

Xyl removal, suggesting that residual xylan limits access to the remaining material. Further experiments comparing milled and un-milled steam exploded straw demonstrated a physical constraint to hydrolysis - particularly when lower pretreatment severities are used.

Ultimately, the aim of the experiments conducted in this chapter was to gain an understanding of how *B. napus* straw responds to pretreatment at a pilot-scale. Variations in process parameters established in this section, were used to inform later experiments and aid in the selection of industrially relevant conditions with which to screen many cultivars. In the following chapter, near-optimum pretreatment conditions (210°C, 10 min), were applied to straw collected from a selection of cultivars, to explore the likely influence of cultivar selection on product yields relevant to bioethanol production after pilot-scale processing.

5 Variability in product yields relevant to cellulosic ethanol production from OSR cultivars after pilot-scale steam explosion

5.1 Chapter outline

Straw derived from a selection of *Brassica napus* cultivars was steam exploded at a pilot-scale using near-optimal conditions (210°C, 10 min) to establish the potential for genotypic variations in cellulosic ethanol traits. Despite being grown and processed in the same manner, straw from various cultivars produced different yields after processing. The abundance of fermentation inhibitors released by SE also varied between genotypes. Cultivars with glucan-rich straw did not necessarily produce higher saccharification or ethanol yields after processing. Instead, substrate quality was determined by the abundance and characteristics of non-cellulosic components – where the majority of the variation between genotypes was found. This variation reflected in the quality of the pretreated material, the amount of hydrolysable glucan, and therefore final ethanol yields after processing. The abundance of Gal, and potentially Ara, containing carbohydrates and properties of the pectic matrix (such as methyl-esterification) appeared to be particularly important in governing ethanol yields between straw genotypes.

5.2 Introduction

Although process dependent differences have been explored using a straw from a single genotype, little is known about the effect that cultivar variation may have on the yields produced after pilot-scale processing. The previous chapter highlighted the likely determinants of saccharification efficiency governed by pretreatment and hydrolysis conditions. This information, combined with previous work on the SE of OSR straw provides a detailed overview of how OSR straw responds to pilot-scale processing (Ryden *et al.*, 2014; Vivekanand *et al.*, 2012). This creates a solid foundation in which to conduct further work.

Genotype variation is an important consideration for industry for many reasons. If sufficient variation exists between cultivars, crop breeders could exploit this variation to improve feedstock quality (Larsen *et al.*, 2012; Lindedam *et al.*, 2012). On the other hand, variation may be undesirable to Biorefinery operators, who may want uniform and predictable yields regardless of the biomass source. Therefore, variation in CW structures between cultivars is an interesting parameter to explore.

Other researchers have observed variations in fermentable sugar yields between cultivars, using other agricultural residues such as wheat straw (Larsen *et al.*, 2012; Lindedam *et al.*, 2012; Lindedam *et al.*, 2014), rice straw (Matsuda *et al.*, 2011), corn

stover (Isci *et al.*, 2008; Lewis *et al.*, 2010) and maize stover (Lorenz *et al.*, 2009) – but not with a dicotyledonous straw and rarely at a pilot-scale (Lindedam *et al.*, 2010). Furthermore, these studies have generally aimed to establish statistically significant differences in yields between cultivars and not necessarily on the underlying changes in CW structure governing the differences. Therefore, the chemical basis for these variations is not fully understood.

This chapter aimed not only to observe differences in straw quality between OSR cultivars using pilot-scale processing, but also to relate those differences to straw composition. To do this, straw derived from a selection of OSR cultivars and other crop types of the same species (*B. napus*) was pretreated at near-optimal conditions established in the previous chapter (210 °C, 10 min) using pilot-scale SE. The chemical composition of the original material, pretreated substrates and products released during processing were established. IR-spectra were also taken from these materials also gave an insight into their polymeric structure. Glc and ethanol yields of the pretreated material were quantified after hydrolysis and SSF respectively. This data allowed differences in yield between cultivars to be traced back to genotypic differences in straw composition.

5.3 Materials and methods

5.3.1 Materials

Seventeen *B. napus* cultivars were grown under field conditions at KWS UK Ltd, Cambridge, UK (+52°, 8', 32.40", -1°, 6', 19.66") in a randomised order, in adjacent 3 m x 12 m plots. The cultivars selected were a genetically diverse selection of *B. napus* genotypes, representative of the most common sub-groups – Winter OSR (WR), Spring OSR (SR), Fodder Rapes (FR) and Swede (SW) (Harper *et al.*, 2012). The cultivars analysed in this study were: Canard (FR), Canberra x Courage (WR), Darmor (WR), Erglu (SR), Hansen x Gaspard (WR), Judzae (SW), Lincrown x Express (WR), Madrical x Recital (WR), Major (WR), POH285 Bolko (WR), Quinta (WR), Ramses (WR), Sensation NZ (SW), Shannon x Winner (WR), Slapka Slapy (unspecified), Slovenska Krajova (WR) and York (SW).

All cultivars were harvested at maturity (8 Aug 2012). Approximately 3 Kg of OSR straw was collected upon ejection from a combine harvester, which directly threshed and chipped the straw from a single cultivar into 2-3 cm pieces. The straw sample was taken from the centre of each 3 m strip to prevent contamination from adjacent cultivars. The straw was stored in woven polypropylene bags in a dry, unheated room before analysis.

The cellulase cocktail used in this chapter was Cellic® CTec2 (Novozymes, Denmark) with a stock cellulase activity of 180 FPU/mL determined following (Ghose, 1987). Unless otherwise stated, all chemicals used were analytical grade, purchased from Sigma-Aldrich, UK.

5.3.2 Steam explosion of oilseed rape straw

A sample of OSR straw (1 Kg F.W.) from each cultivar was steam exploded into hot water (6.6 L) at a near-optimum pretreatment severity (210 °C, 10 min) using a Cambi™ Steam Explosion Pilot Plant. After SE, the heating chamber was cleared twice by applying 2-3 bar of pressure to dislodge the majority of residual material. The pretreated biomass was filtered immediately through a 100 µm nylon mesh bag in a low speed centrifuge. The solid and liquid products were measured and a representative sample of each fraction was taken for analysis. The SE unit was extensively rinsed between each pretreatment to prevent cross contamination between cultivars.

5.3.3 Analysis of steam explosion liquors

The concentration of fermentation inhibitors (organic acids and furfural derivatives) retained in each liquor (water soluble fraction created after SE) was quantified by HPLC after filtration (96 well filter plate, 0.2µm).

5.3.4 Chemical composition of the untreated and pretreated solids

The matter content of the pretreated solid produced from each cultivar was established using an infrared drying balance (Mettler LP16, Mettler-Toledo, Belgium) drying duplicate samples (0.5 g) at 105°C, to constant mass. A sample of each SE solid and untreated material was frozen in liquid nitrogen and freeze-milled into a fine powder to gain a homogenous sample for chemical analysis (3 min, SPEX 6700 freezer/mill, Spex Industries, NJ) and dried to constant mass (40 °C, overnight). The neutral sugar and UA compositions of the solid were established following protocols outlined in Chapter 3.

5.3.5 Fourier transform infrared spectroscopy

FT-IR spectra were collected in the 800–4000 cm^{-1} region for each freeze-milled sample using methodology outlined in Chapter 3. Triplicate spectra were taken for each material, truncated (800-1800 cm^{-1}), baseline corrected (to 1800 cm^{-1}), area normalised before analysis. Partial least squares (PLS) regression of spectra and chemical data was conducted in Matlab® with assistance from Klaus Wellner.

5.3.6 Determining saccharification yields for each cultivar

A 1g (DW equivalent) sample of each pretreated solid was suspended in 20 mL sodium acetate / acetic acid buffer (5% substrate, 0.1M, pH 5, 0.01% thiomersal) in 30 mL screw-top vials (Sterlin, UK), held in a shaker plate incubator (50 °C, 150 RPM). Cellic® CTec2 was added to the equilibrated solutions at a cellulase dose of 0.2 mL/g substrate (ca. 36 FPU/g). Digestions were conducted in triplicate and the amount of Glc quantified after 96 h of incubation, using the GOPOD assay outlined in Chapter 3. The amount of cellulase-derived Glc was subtracted from the total.

5.3.7 Simultaneous saccharification and fermentation of pretreated straw

A sample of each pretreated substrate was suspended in 10 mL solution with a final concentration of 5% substrate in nitrogen base (Formedium, Hunstanton, UK) and held in 20 mL screw-topped glass vials. Both cellulase (36 FPU cellulase/g substrate) and a concentrated yeast inoculum were added to each vial and incubated for 96 h, 40°C.

The yeast inoculum used (*Saccharomyces cerevisiae,* strain NCYC 2826, National Collection of Yeast Cultures, Norwich, UK) was grown from a slope culture, inoculating 1L of Yeast Mould (YM) broth (3 d, 25°C) before centrifuging, discarding the supernatant and partially reconstituting the yeast in nitrogen base. The final solutions contained 3.83 x 10^7 viable cells / mL when inoculated – assayed using a NucleoCounter® YC-100™ (ChemoMetec, Denmark). SSFs were conducted as three independent replicates and the ethanol released from a cellulase + yeast control was subtracted from each sample. Ethanol concentrations were quantified using HPLC.

5.4 Results and discussion

5.4.1 Composition of untreated OSR straw derived from different cultivars

Despite being grown, harvested, stored and analysed under the same conditions, significant variations in the abundance of constituent sugars were observed between straw from different cultivars (Table 7).

The abundance of minor sugars (UA, Man and Ara) differed most notably between cultivars. Rha was the only sugar not to differ in abundance significantly between cultivars before pretreatment. The moisture content of the straw was ca. 9.5% and did not differ significantly between cultivars (Table 7). This established the main variations in straw carbohydrate compositions between cultivars.

Cultivar name	Composition (g/kg original air-dry straw)*									
	Glc	Xyl	UA	Man	Gal	Ara	Rha	Fuc	MC	Other
Canard	315	139	44	16	11	11	4	1	89	370
Canberra x Courage	319	146	39	16	12	9	5	1	95	359
Darmor	299	133	40	18	12	10	5	2	100	381
Erglu	339	145	37	20	14	12	5	2	100	325
Hansen x Gaspard	353	130	38	19	14	10	5	2	100	330
Judzae	342	149	40	21	15	11	6	3	99	314
Licrown x Express	377	146	38	24	13	11	6	1	96	288
Madrical x Recital	331	137	44	20	13	13	5	2	101	334
Major	315	130	44	19	12	9	5	1	95	369
POH285, Bolko	334	145	47	18	13	11	5	1	84	342
Quinta	289	125	34	15	14	14	5	1	95	408
Ramses	292	115	40	15	13	13	5	1	86	419
Sensation NZ	349	148	34	19	13	9	5	1	85	337
Shannon x Winner	320	140	45	18	13	12	5	1	107	339
Slapka Slapy S3	310	134	33	17	11	10	5	1	101	379
Slovenska Krajova	321	135	44	16	12	10	5	1	86	370
York	349	150	40	20	13	9	5	1	102	312
Mean	327	138	40	18	13	11	5	1	95	352
Range	88	35	14	9	4	5	2	2	23	131
Range (% mean)	27%	26%	35%	50%	31%	48%	35%	113%	24%	37%
*ANOVA** (p value)*	<0.05	<0.01	<0.001	<0.001	<0.05	<0.001	0.564	<0.01	0.399	-

* Glc = Glucose, Xyl = Xylose, UA = Uronic acids, Man = Mannose, Gal = Galactose, Ara = Arabinose, Rha = Rhamnose, Fuc = Fucose, MC = Moisture content, Other = Other non-carbohydrate matter by difference.

** One way analysis of variance

Table 7 – Sugar composition of untreated *B. napus* straw derived from different cultivars

5.4.2 Variation in the composition of steam exploded solid residues

Straw derived from each cultivar (1kg) was steam exploded into hot water at near optimal conditions (210°C, 10 min). The sugar compositions of the SE solids were established to see if genotypic variation in composition observed between the untreated straw of different cultivars was retained after pretreatment (Table 8).

Although commercially relevant, pilot-scale processing is not as conducive to biomass screening of varying genotypes as other methods. Incomplete recovery of material during pretreatment and washing steps is inevitable when analysing discrete samples using SE (Rocha *et al.*, 2012). Therefore, yield estimates based on the original material presented here are not representative of those that would be obtained under continual use, in an industrial setting.

Nevertheless, biomass from different cultivars were treated identically and processed in a random order in relationship to their CW compositions. Therefore, the relative differences in chemistry of the pretreated material and saccharification yields are likely to reflect the genotypic differences in biomass composition. After SE, Man, Gal and Fuc (Man, Gal and Fuc) were almost completely removed from the water insoluble fraction (< 5% of the original remained) (Table 8). Likewise, other non-cellulosic sugars (Xyl, Ara, Rha and UA) were also removed, but retained a higher proportion of their sugars in the pretreated residue (10-20% of the original). By contrast, > 60%, and in some cases 80% of the original glucan available in the original material was retained in the SE residue (Table 8).

After SE, the largest quantitative difference between substrates produced from different cultivars was in the abundance of glucan retained in the water insoluble material (Table 8). However, larger variations were observed in the retention of non-cellulosic sugars such as Xyl, Ara and Rha. Interestingly, the abundance of Rha retained in the pretreated pellet differed between cultivars after PT ($p<0.05$), despite having similar abundance in the untreated straw. Straw from some cultivars, such as Canard, retained small quantities of Ara after SE (≈5g/kg), as others, such as York, retained almost none. No significant differences in UA, Man, Gal and Fuc abundance were observed between cultivars following pretreatment ($p>0.05$, Table 8).

Cultivar name	Recovery (g/kg original FW)		Composition (g/kg pretreated material DW)								
	Mass DW	MC (%)	Glc	Xyl	UA	Man	Gal	Ara	Rha	Fuc	Other
Canard	340	65	416	30	13	2	1	5	2	Trace	544
Canberra x Courage	514	67	413	30	13	2	2	5	2	Trace	546
Darmor	447	68	426	30	14	2	1	4	2	Trace	536
Erglu	488	66	394	30	19	2	2	5	2	Trace	566
Hansen x Gaspard	520	61	430	33	11	Trace	2	4	2	Trace	528
Judzae	554	65	400	34	17	Trace	2	4	2	Trace	558
Licrown x Express	528	71	407	27	12	2	1	3	2	Trace	558
Madrical x Recital	616	65	427	29	13	Trace	1	3	2	Trace	539
Major	460	71	427	26	13	Trace	1	0	2	Trace	543
POH285, Bolko	447	71	437	32	19	Trace	1	1	2	Trace	528
Quinta	513	69	361	31	13	Trace	1	4	1	Trace	602
Ramses	478	61	439	28	17	Trace	1	2	2	Trace	528
Sensation NZ	542	64	384	31	12	2	2	3	2	Trace	576
Shannon x Winner	523	66	395	26	16	Trace	1	1	2	Trace	576
Slapka Slapy S3	480	70	413	26	12	1	1	3	2	Trace	555
Slovenska Krajova	475	71	377	32	12	3	1	5	2	Trace	581
York	503	65	455	28	17	Trace	0	0	2	Trace	515
Mean	496	67	412	30	14	-	1	3	2	-	552
Range	276	10	94	8	8	-	2	5	1	-	87
Range (% mean)	56%	15%	23%	27%	56%	-	135%	171%	47%	-	16%
ANOVA (p-value)	-	<0.01	<0.001	<0.001	0.511	-	0.287	<0.05	<0.05	-	-

* Glc = Glucose, Xyl = Xylose, Ara = Arabinose, Gal = Galactose, Fuc = Fucose, UA = Uronic acids, Other = other non-carbohydrate matter.
MC = Moisture content

Table 8 – Matter recoveries and monomeric sugar composition of straw derived from different cultivars steam exploded at 210°C, 10 min.

5.4.3 Variations in infra-red spectra between OSR cultivars

Fourier transform infra-red spectroscopy has been used extensively to probe the structure of plant CWs (Kacurakova *et al.*, 2000; Szymanska-Chargot & Zdunek, 2013). Here, spectra for OSR straw from different cultivars before and after SE was used to assess cultivar-specific differences at a polymeric level (Figure 23).

Spectra collected from untreated straw showed greater variation between cultivars than those from the same cultivars after SE. The largest spectral differences were observed at wavenumbers typically associated with structural carbohydrates - cellulose, XG and pectic structures - 875, 1020, 1240, 1315, 1420, 1600 cm^{-1} and 1730 cm^{-1} (Szymanska-Chargot & Zdunek, 2013). Particular cultivars showed above average deviation in absorbance at certain wavelengths. For example, Ramses straw exhibited higher absorbance at 875 cm^{-1} (C1-H bending in XG and cellulose) compared to other cultivars. Similarly, Hansen x Gaspard showed above-average absorbance at 1600 cm^{-1} (COO- asymmetric stretching) – suggesting differences in pectic cross-linking (Szymanska-Chargot & Zdunek, 2013).

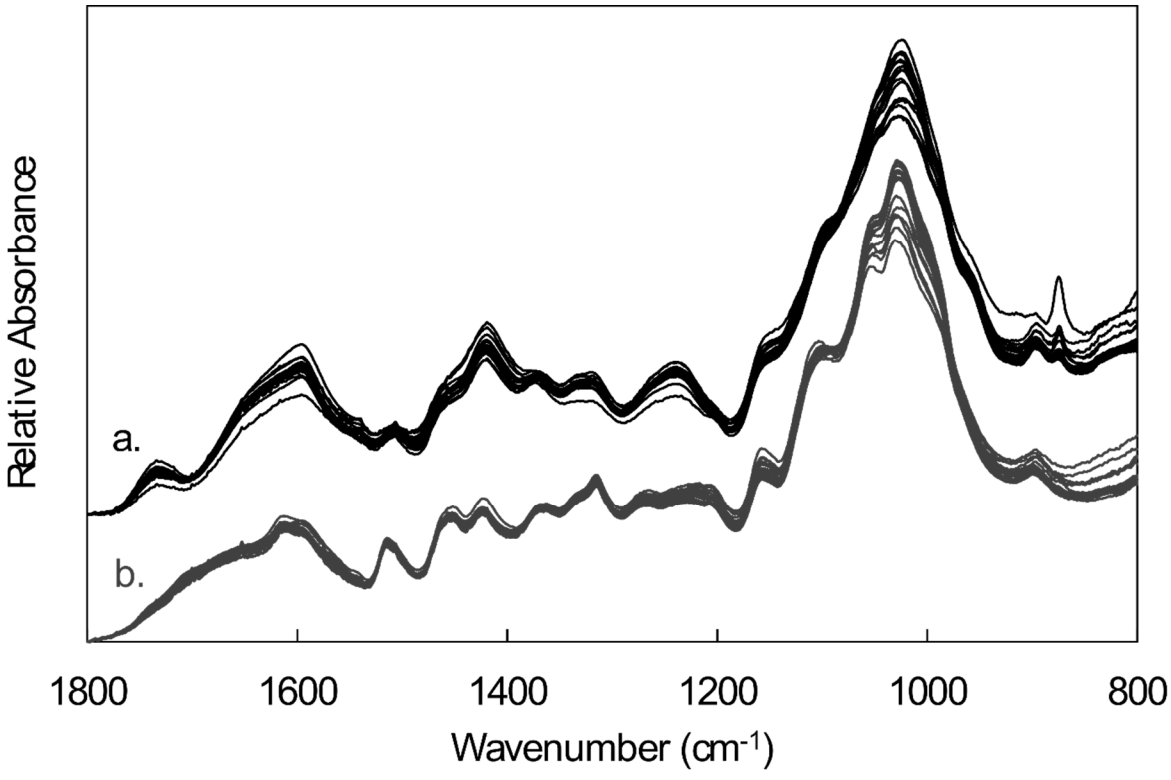


Figure 23 – Average FT-IR spectra collected from straw, derived from different cultivars before (a) and after (b) SE at 210 °C, 10 min.

After pretreatment, spectra taken from the residues of different cultivars were more uniform (Figure 23). Largest variation between cultivars was observed at wavenumbers related to non-cellulosic components: 1020 cm^{-1} (C–O stretching, C–C stretching in XG and pectins) and 1155 cm^{-1} (C-O-C glycosidic linkages in xylan) (Kacurakova *et al.*, 2000). Spectral variations between cultivars identified at other wavenumbers were diminished following SE, reflecting the extent of their removal from the biomass.

5.4.4 Variation in fermentation inhibitor release from different cultivars

SE of *B. napus* straw at severities required to gain reasonable saccharification yields (> 60%) tends to produce a significant amount of organic compounds (see Chapter 4). It is important to quantify these compounds as they can inhibit downstream processes – particularly fermentation. Therefore four common inhibitory compounds (HMF, 2FA, acetic acid and formic acid) released into the pretreatment liquor were quantified (Table 9). The abundance of fermentation inhibitors retained in the pretreatment liquor after SE differed significantly between cultivars ($p < 0.001$) (Table 9).

Cultivar name	Volume (L)	Concentration - g/L SE liquor			
		Acetic	Formic	2FA	HMF
Canard	6.60	2.86	2.02	0.51	0.15
Canberra x Courage	7.08	2.68	2.04	0.54	0.16
Darmor	7.13	2.75	2.11	0.46	0.11
Erglu	6.95	2.72	1.97	0.32	0.15
Hansen x Gaspard	6.96	2.35	1.77	0.49	0.12
Judzae	7.00	2.25	1.81	0.37	0.11
Licrown x Express	6.25	3.20	2.08	0.50	0.18
Madrical x Recital	6.50	2.76	2.02	0.42	0.13
Major	6.84	2.59	1.91	0.34	0.14
POH285, Bolko	7.00	3.13	2.18	0.46	0.18
Quinta	6.76	2.91	1.95	0.67	0.20
Ramses	6.55	2.68	1.60	0.45	0.23
Sensation NZ	7.15	2.43	1.96	0.28	0.11
Shannon x Winner	6.85	3.40	1.92	0.39	0.26
Slapka Slapy S3	6.92	3.08	1.97	0.41	0.20
Slovenska Krajova	6.78	2.39	1.50	0.76	0.17
York	6.18	3.23	2.08	0.43	0.23
Mean	6.79	2.79	1.93	0.46	0.17
Range	0.97	1.15	0.68	0.48	0.15
Range (% mean)	14%	41%	35%	105%	90%

*Acetic = Acetic acid, Formic = Formic acid, 2FA = 2-furfuraldehyde, HMF = hydroxymethylfurfural

Table 9 – Concentration of organic acids and furfural derivatives retained in the pretreatment liquors of straw derived from different cultivars. The abundance of all compounds in the hydrolysis liquors differed significantly between cultivars (ANOVA, $p < 0.001$)

5.4.5 Straw from different cultivars obtained different hydrolysis and fermentation yields

A portion of the steam exploded biomass derived from each cultivar was converted to either Glc or ethanol by enzymatic hydrolysis or SSF respectively using a near-optimum cellulase dose (36 FPU/g substrate) (Table 10). Although all 17 cultivars were grown, processed and analysed in the same manner, significant differences ($p < 0.001$) in product yields were observed between cultivars (Table 10).

A number of cultivars achieved atypical yields, which may suggest genotypic differences in yield applicable to a single cultivar. These included Ramses, which produced much lower saccharification and fermentation yields compared to other cultivars (Table 10). Similarly ethanol yields produced by SSF generally reflected Glc yields saccharified from the material except for Hansen x Gaspard and Canberra x Courage, which obtained good saccharification yields but performed very poorly under SSF conditions (Table 10). Without these outliers, reducing sugar yields correlated strongly with Glc yields ($p < 0.001$, $r = 0.8824$, $n = 14$), and Glc yields correlated with ethanol yields as one might expect ($p < 0.05$, $r = 0.6005$, $n = 14$). Typically, 95% of the Glc hydrolysed from the pretreated material was fermented to ethanol under SSF conditions. It was therefore decided to discuss the general trends applicable to most cultivars separately, and suggest possible reasons why 'Ramses', 'Hansen x Gaspard' and 'Canberra x Courage' resulted in lower product yields.

Cultivar name	Product yield g/kg pretreated material						
	Glucose		Reducing sugar		Ethanol		
	g/kg	% in PT	g/kg	% in PT	g/kg	% PT*	% hydrolysed*
Canard	289	(69%)	445	(95%)	142	(67%)	(96%)
Canberra x Courage	352	(85%)	510	(109%)	107	(51%)	(60%)
Darmor	367	(86%)	483	(101%)	173	(80%)	(92%)
Erglu	286	(73%)	415	(91%)	154	(77%)	(106%)
Hansen x Gaspard	361	(84%)	564	(117%)	107	(49%)	(58%)
Judzae	269	(67%)	395	(86%)	171	(84%)	(125%)
Licrown x Express	289	(71%)	438	(96%)	147	(71%)	(100%)
Madrical x Recital	302	(71%)	452	(95%)	135	(62%)	(88%)
Major	344	(81%)	484	(103%)	177	(81%)	(101%)
POH285, Bolko	331	(76%)	496	(101%)	146	(66%)	(87%)
Quinta	312	(87%)	460	(112%)	137	(74%)	(86%)
Ramses	215	(49%)	325	(67%)	91	(41%)	(83%)
Sensation NZ	277	(72%)	423	(97%)	135	(69%)	(96%)
Shannon x Winner	266	(67%)	388	(88%)	125	(62%)	(92%)
Slapka Slapy S3	374	(90%)	514	(112%)	185	(88%)	(97%)
Slovenska Krajova	332	(88%)	511	(118%)	157	(81%)	(92%)
York	344	(76%)	456	(91%)	141	(61%)	(80%)
Mean	312	(76%)	456	(99%)	143	(68%)	(90%)
Range	159	(41%)	239	(51%)	94	(47%)	(67%)
Range (% mean)	51%	-	52%	-	66%	-	-

*Est 0.51 g ethanol/g Glc

Table 10 – Mass of glucose and ethanol produced from pretreated straw derived from different cultivars (5% substrate, 37 FPU/g, 96 h) incubated at 50°C or 40°C respectively. The percentage yields are included in parentheses. Ethanol yields are quoted as either% of the theoretical in the pretreated substrate (%PT) or of the hydrolysed Glc (% hydrolysed). Significant differences in product yields were observed between cultivars (ANOVA $p < 0.001$).

5.4.6 Relationship between straw composition and product yields

To understand the potential relationship between cultivar straw composition and product yields, the abundance of each component sugar present in the untreated and pretreated residues were compared with Glc and ethanol yields after processing (Table 11).

Cultivars that produced glucan-rich straw did not necessarily produce higher Glc (p = 0.2394, n = 16) or ethanol yields after processing (p = 0.424, n = 14). This was also true for the glucan content in the PT residue to the variation in glucose and ethanol yields (p = 0.068 and 0.5483, n = 16 and 14 respectively). This observation concurs with previous findings in maize, where ethanol yields between cultivars was more closely related to glucan convertibility rather than glucan content (Lorenz *et al.*, 2009).

Component	Glc yield (g/kg PT, n=16)		RS yield (g/kg PT, n = 16)		Ethanol yield (g/kg PT, n = 14)	
	R^2	p	R^2	p	R^2	p
Original straw composition (g/kg FW)						
Glucose	-0.3120	0.2394	-0.1622	0.5485	-0.2326	0.4236
Xylose	-0.4319	0.0948	-0.4900	0.0540	-0.2502	0.3882
Uronic acid	-0.1127	0.6778	-0.0452	0.8680	-0.1833	0.5305
Arabinose	-0.4280	0.0981	-0.3445	0.1913	-0.3574	0.2096
Galactose	**-0.5602**	**0.024***	-0.4523	0.0786	-0.3506	0.2190
Rhamnose	-0.3909	0.1344	-0.3754	0.1519	0.1685	0.5647
Mannose	-0.3375	0.2010	-0.3702	0.1581	0.0114	0.9691
Fucose	-0.1722	0.5236	-0.2262	0.3997	0.3366	0.2393
Other	0.4121	0.1127	0.3577	0.1738	0.2472	0.3943
Pretreated straw composition (g/kg DW)						
Glucose	0.4667	0.0684	0.3169	0.2317	0.1756	0.1206
Xylose	-0.0917	0.7356	0.1489	0.5821	-0.0493	0.3837
Uronic acid	-0.2849	0.2848	-0.4593	0.0735	-0.1021	0.5511
Arabinose	-0.0597	0.8260	0.1330	0.6234	0.1286	0.4473
Galactose	-0.1838	0.4956	0.0121	0.9644	0.2918	0.1343
Rhamnose	0.2092	0.4369	0.2757	0.3014	0.1691	0.9760
Other	-0.4753	0.0628	-0.3694	0.1591	-0.2063	0.1440

Table 11 – Correlation coefficients and p-values for a two-sided test from zero. Bold values suggest some relationship with significant correlations ($p < 0.05$) marked with an asterisk.

Imperfect fractionation of steam and biomass within the SE apparatus made precise yield estimates difficult to determine. Nevertheless, cultivars that contained less Gal in their original straw tended to produce higher Glc yields after SE and hydrolysis (p < 0.05, n = 16). Notable but not significant negative correlations with saccharification yields were also observed for xylan and arabinan abundance in the original straw (p = 0.053-0.098 n.s., n = 16). Interestingly, Ramses contained relatively high Gal and Ara content compared to other cultivars (Table 7), which strengthened the correlation of Ara and Gal with Glc yields after processing when all cultivars were included (R = -0.5587 and 0.5303 respectively, p < 0.05, n = 17). It is therefore possible that particularly low saccharification yields were produced from Ramses straw because of this difference in CW chemistry.

Conversely, variations in the composition of the pretreated materials between cultivars did not correlate with the range in saccharification yields between cultivars (g/kg pretreated material) (Table 11). This suggests that the genotypic variation in recalcitrance of the pretreated material was most closely related to the variation in hydrolysed components, rather than the retention of recalcitrant sugars in the pretreated material. There was some indication that the amount of polymeric UA and Glc retained in the pretreated solid limited RS and increased Glc yields respectively, however these correlations were not significant between cultivars (p = 0.0628 and 0.0735 respectively).

Previous work conducted using *B. napus* straw showed that Ara typically responds to SE in a similar way to pectic sugars, whereas Gal decomposition is more similar to that of hemicellulosic components (Ryden *et al.*, 2014). The results showed that Gal, and possibly Ara showed a negative correlation with Glc yield. This would suggest that polymers containing higher Gal and Ara contents such as XG side-chains, AGPs or pectins (particularly RGI), might be involved in determining biomass recalcitrance between cultivars after pilot-scale SE. Recent evidence has shown that AGPs can covalently link the hemicellulose-pectin network (Tan *et al.*, 2013). AGPs are thermally resistant CW associated polymers (Seifert & Roberts, 2007), therefore it is possible that genotypic differences in the abundance of these, or similarly Ara and Gal-rich polymers, could have a significant impact on substrate recalcitrance after pretreatment. Further work, involving the more detailed characterisation of biomass, would be needed to ascertain the effect that these carbohydrates may have on CW recalcitrance.

5.4.7 Relating genotypic variation in IR-spectra with variation in ethanol yields using partial least squares (PLS) regression.

Partial least squares (PLS) regression is a convenient way of correlating quantitative measurements with spectral data. As mentioned previously, FT-IR spectra give an overview of the constituent bonds present in the material – thereby giving information as to its polymeric structure. This can make spectral interpretations of biological material difficult, as many infra-red absorbance peaks overlap. Splitting the spectral variation into successive, principal components, using multivariate analysis makes the data more accessible: highlighting areas of the spectra that correlate with variances in quantitative measurements.

Previously, this methodology has been used to identify the main polymeric changes that occur in OSR straw following SE at varying pretreatment severities (Ryden *et al.*, 2014). The crucial effects that these changes in severity had on subsequent Glc (via enzymatic hydrolysis) and methane generation (after anaerobic digestion) were also identified (Ryden *et al.*, 2014; Vivekanand *et al.*, 2012).

Here PLS modelling was used to match spectral variations between cultivars to variations in ethanol yields after processing – summarising them into six PLS components (PLS 1-6). Spectra taken from the untreated straw were correlated with ethanol yields in g/kg original straw (Figure 24, A). Likewise, spectra taken from pretreated straw were correlated with ethanol yields expressed as g/kg steam exploded straw (Figure 24, B).

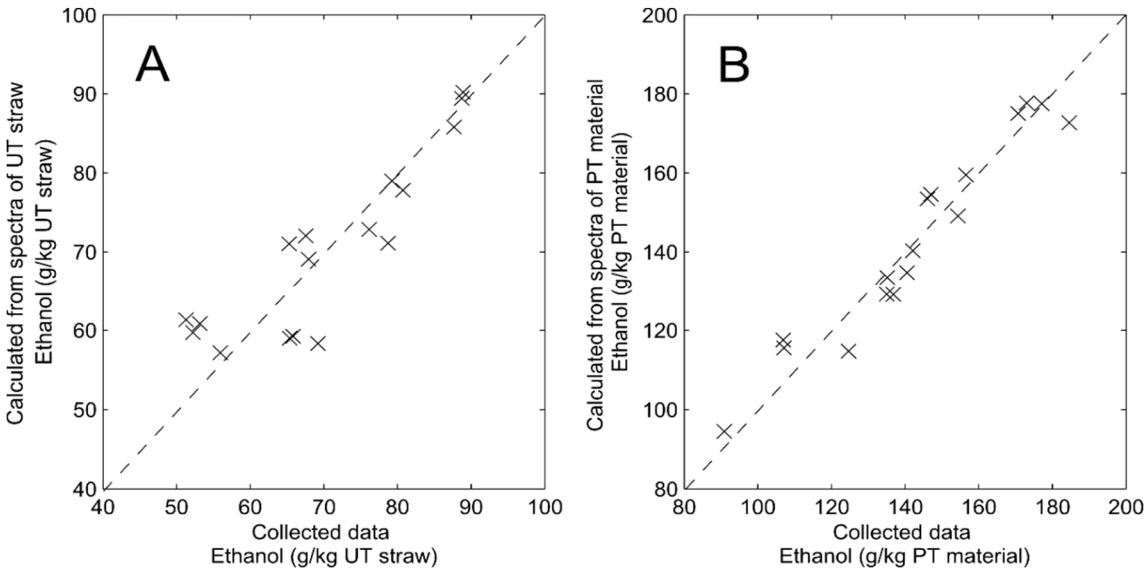


Figure 24 – Partial least squares (PLS) models correlating variance in IR spectra collected from untreated (A) and pretreated (B) materials with the variation in ethanol yields produced between cultivars. Correlations using six PLS components are shown.

The loadings for each component were examined to identify what differences in polymeric associations between cultivars are likely to influence ethanol yields (Figure 25). In total, the first six PLS components could describe 97-98% of intraspecific variation in ethanol yields utilising 78-83% of the spectral variations observed between cultivars.

The majority of variation in ethanol yields observed between cultivars after SE (76.83%) could be explained by a single PLS component (PLS 1) – utilising 29.61% of the variation in spectra collected from the pretreated residues (Figure 25, RHS). This spectral variation was mostly isolated to the 1020-1025 cm^{-1} region (C–O stretching, C–C stretching in xylans and pectins), suggesting that residual non-cellulosic carbohydrates were the main spectral differences between cultivars related to ethanol yields (Figure 25, RHS) (Szymanska-Chargot & Zdunek, 2013).

Other spectral variations between pretreated residues derived from different cultivars (identified in PLS components 2-6) explained much less of the variance in ethanol yield. PLS 2 could explain a further 6.46% of the variation in ethanol yields - primarily using variation in absorbance at cellulose associated wavenumbers (1000, 1030, 1103, 1160 cm^{-1}). PLS 3 described a further 7.67% of the variation in ethanol yields, attributed mostly to residual pectin moieties (1600 cm^{-1}, COO- asymmetric stretching) (Figure 3, RHS). The remaining components (4-6) explained less than 7% of the remaining variation combined, highlighting subtle differences in spectral regions previously identified by higher components (Figure 25, RHS).

More PLS components were needed to explain the variation in ethanol yields (g/kg untreated straw) when correlating them against spectra collected from the original straw (Figure 25, LHS). The majority variation in ethanol yields between cultivars (52.42%) can be explained by 15.83% of the intraspecific spectral variation between the untreated straw (PC1). Loadings for PLS 1 identified spectral variation found at wavenumbers related to non-cellulosic components: 1020 cm^{-1} (C–O stretching, C–C stretching in xylans and pectins), 875 cm^{-1} (C1-H bending in XG and cellulose), 1600 cm^{-1} (COO^- asymmetric stretching in pectins) and 1730/40 cm^{-1} (C=O stretching vibration of alkyl ester in pectins) (Kacurakova *et al.*, 2000; Szymanska-Chargot & Zdunek, 2013).

The variation in spectra at 1600 and 1730/40 cm^{-1} in PC1 are particularly interesting as they suggest that increased absorbance of pectic HG between cultivars increases ethanol yield (Szymanska-Chargot & Zdunek, 2013). Unlike most crop residues, *B. napus* is a dicotyledonous plant - with pectin-rich, type I CWs similar to the model plant *Arabidopsis* (Yokoyama & Nishitani, 2004). Genetic manipulation in other species, including *Arabidopsis* has independently shown that saccharification yields

produced from CW material is related to the degree of pectin methyl-esterification (Francocci *et al.*, 2013; Lionetti *et al.*, 2010). It is therefore interesting to see that these changes may also influence genotypic variation in saccharification quality between *B. napus* cultivars after pilot-scale processing.

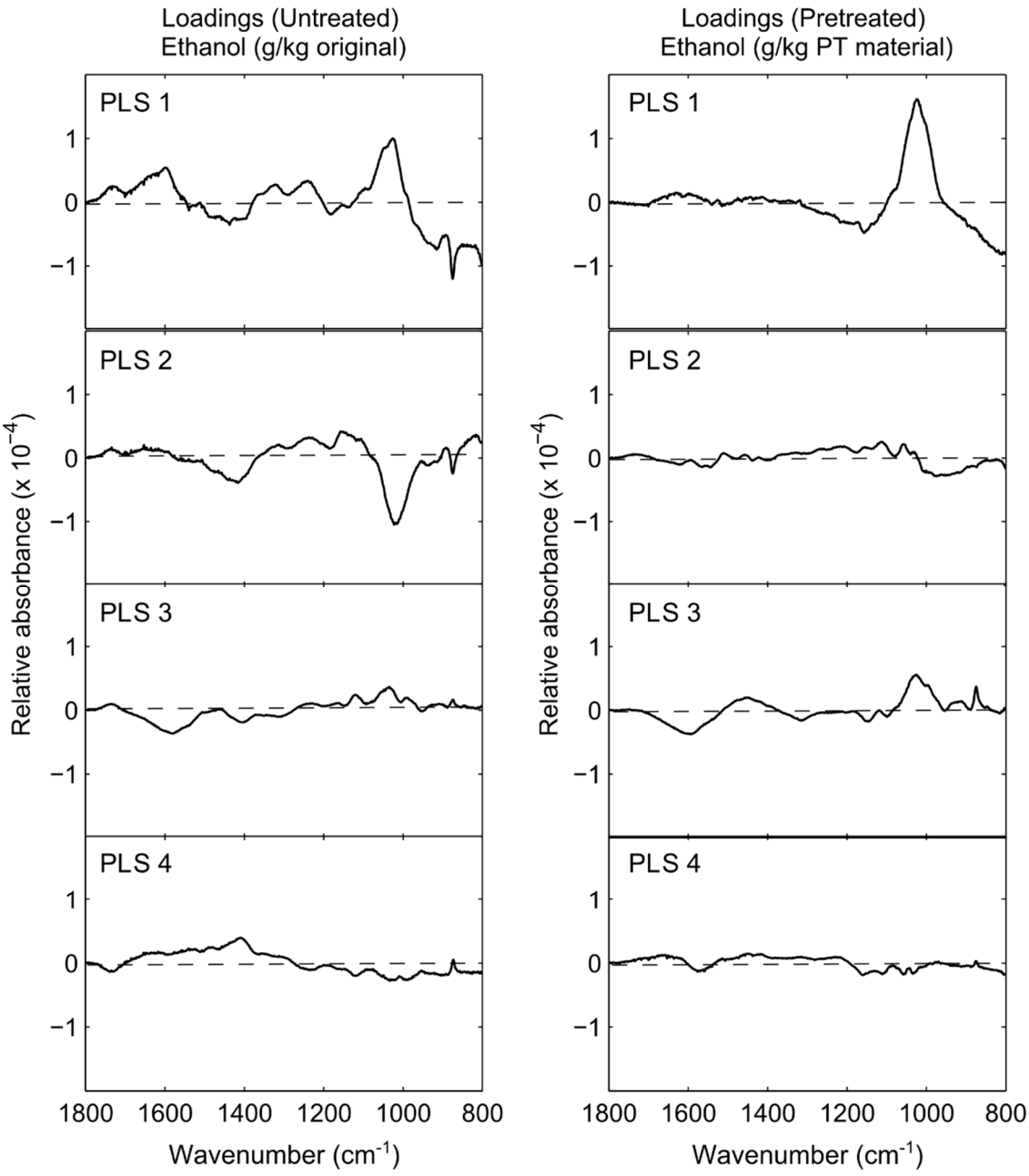


Figure 25 – PLS loadings showing spectral variations correlated with ethanol yields in untreated (LHS) and pretreated (RHS) straw produced from different cultivars. The first four PLS components are displayed (PLS 1-4).

The loadings for the second PLS component (PLS 2), which explains a further 26.37% of the variation in ethanol yields can be explained with spectral variation in pectin associated peaks - the largest being at 1115 cm^{-1} (C-O, C-C stretching in pectin). Minor cellulose-associated peaks also contribute to PLS 2: 1415 cm^{-1} (C-O, C-C stretching in cellulose) and 1160 cm^{-1} (O-C-O asymmetric stretching of the glycosidic bond in cellulose) (Figure 25, LHS).

The main spectral differences in lower PLS components, for example PLS 3, explaining a further 7.05% of variation in ethanol yield, included 1034 cm^{-1} (glucan/glucomannan ring vibrations) and 1580 cm^{-1} (CW proteins) (Kacurakova *et al.*, 2000; Szymanska-Chargot & Zdunek, 2013). PLS 4 explained a further 9.15% of the variation in ethanol yields, primarily from the variation in absorbance at 1408 cm^{-1} (COO- symmetric stretching in pectins) (Figure 25, LHS). Loadings for smaller components (PLS 5-6) showed subtle variations in absorbance in areas previously described in by higher components and together only explained a further 2.28% of the variation in ethanol yields.

5.5 Conclusions

Significant variation in Glc, ethanol and fermentation inhibitor yields were observed between cultivars - despite being grown, harvested and analysed under identical conditions. Interestingly, genotypic differences in straw quality were not simply governed by glucan content of the original material, but by the integrity of the non-cellulosic components. These differences can have an effect on final product yields when using pilot-scale methodology and therefore require further study.

Gal-containing polymers in the original straw correlated negatively with the range in saccharification yields between cultivars. There was also an indication that cultivars with more Xyl and Ara in their original material might yield lower saccharification yields after processing. Interestingly, variations in carbohydrate compositions of the SE materials did not correlate strongly with the variations in sugar or ethanol yields between cultivars. This suggests that differences in Glc yield associated with cultivar variation may be improved by changes to the original material. The pretreated residues were more similar in composition, but the absence of particular components after pretreatment may improve hydrolysis.

Closer inspection of FT-IR spectra using PLS modelling revealed that variations in Glc and ethanol yields between cultivars was probably influenced by the removal of non-cellulosic carbohydrates, with minor contributions made by differences in cellulose structuring, glucomannan and CW proteins implicated by changes in absorbance revealed in lower PLS components.

Practical considerations such as sample size and incomplete sample recovery limit the use of pilot-scale machinery for large-scale biomass screens (Rocha *et al.*, 2012). Improved precision and throughput could be achieved by using smaller-scale techniques. This also opens new experimental possibilities, such as the interaction between genotype and processing conditions, which is an interesting and novel topic.

Therefore, in the following chapter, several existing methods were adapted to make them suitable for biomass screening. Rational screening conditions that result in varying degrees of biomatrix opening were selected based on information gathered from a single cultivar and applied to a larger range of cultivars to explore interactions between processing conditions and genotype selection.

6 Screening biomass for traits relevant to cellulosic ethanol production

6.1 Chapter outline

High throughput (HT) screening methods can provide large datasets, which would be impossible to achieve using conventional techniques. Nevertheless, technical challenges limit the use of HT methods for biomass analysis (Decker *et al.*, 2009). In this chapter, existing methods were tailored specifically for the HT analysis of biomass. Particular care was taken to select accurate but flexible routes for biomass analysis that reflect the different processing permutations available. Screening conditions were selected with the aim of producing complementary and contrasting datasets. The ultimate goal of this chapter was to obtain data relevant to lignocellulosic bioethanol production from many cultivars, of sufficient quality to conduct a pilot-scale GWA study.

Results in this chapter show the optimisation of small-scale methodology to pretreat, dispense, hydrolyse and quantify saccharification products. Precise quantification of reducing sugars in biomass hydrolysates can be achieved by manipulating the ratio of sample to DNS reagent and reading wavelengths. Straw from a single cultivar was initially pretreated at 130 to 220 °C, 10 min, and saccharification yields modelled to show the change in sugar yields following pretreatment at a range of severities. From these results, rational screening conditions were selected. Seventy-seven accessions of *B. napus* straw, derived from 49 cultivars were processed using four pretreatment conditions. This showed that cultivars that produced more Glc under one set of pretreatment and hydrolysis conditions also tended to produce higher yields when using another conditions. However, some process-specific interactions between genotype and processing conditions were observed. Some traits such as the variation in fermentation inhibitor yields between cultivars were clearly process specific. Interestingly, cultivars that released more organic acids under suboptimal pretreatment conditions also achieved higher Glc yields after hydrolysis. This shows that variations in endogeneous organic acid release between cultivars can have a significiant effect on saccharificiation yield.

6.2 Introduction

Although commercially relevant, pilot-scale processing is not conducive to analysis of many different cultivars. Practical limitations of working with large volumes of material, combined with the need to analyse many discrete samples, in a system that works best under continual use are major limitations. Only partial recovery of the sample between runs using large-scale machinery makes precise quantification

suitable for comparison difficult (Rocha *et al.*, 2012). Therefore, establishing smaller scale biomass processing methods is desirable.

Furthermore, HT analysis of biomass opens new lines of investigation that would not be technically, or financially feasible before. Scaling down processes also reduces capital and consumable costs considerably when compared to larger systems (Berlin *et al.*, 2006 and Chundawat *et al.*, 2008) and increases the speed at which data can be collected. Therefore, HT biomass analysis is ideal for optimising multi-component cellulase cocktails and identify germplasm with superior saccharification potential (Gomez *et al.*, 2010). Indeed, collecting suitable phenotypic data, with sufficient replication for GWA studies, can only be achieved using HT methods (Cobb *et al.*, 2013).

Before collecting screening parameters at a small scale, a number of logistical challenges must be addressed (Comprehensively reviewed by Decker *et al.,* (2009). Key difficulties in HT screening techniques include accurate biomass delivery, using appropriate pretreatment and hydrolysis conditions and, rapid and accurate quantification of products (Decker *et al.*, 2009).

Previous biomass screening attempts show how important it is to consider these parameters before collecting screening data. For example, Oakey *et al.,* (2013) demonstrated that when screening ≈1500 barley straw accessions, the environmental variation seen between cultivars over 2 years (22-28% of the total) was dwarfed by the variation introduced during processing (57-58%). This shows the amount of variation that can potentially be introduced if experimental design is not carefully considered.

The authors advocated the use of post-experimental statistical correction to gain meaningful screening results (Oakey *et al.*, 2013). However, others would argue that if consistent protocols are developed that remove all avoidable sources of variability, post-experiment filters should not be needed (Lambert & Black, 2012). Furthermore, in direct opposition to the view presented by Oakey *et al.,* (2013), Lambert & Black (2012) stated that automated post-experimental clean-up is an insufficient remedy for poor experimental design.

In the case of Oakey *et al.,* (2013), key weaknesses in the experimental protocol undoubtedly led to a disproportionately large processing variation. Examples of avoidable sources of experimental variation included: (i) Analysing minute amounts of original material (4 mg) dispensed with an initial accuracy of ±0.1 mg (±2.5%) (Gomez *et al.*, 2010) and using very low substrate concentrations (0.8% w/w) which inevitably magnified errors in subsequent steps. (ii) All technical replicates were analysed consecutively in adjacent wells, in the same plates, over multiple days, with no

randomisation of samples (Oakey *et al.*, 2013). (iii) Selection of the 3-methyl-2-benzothiazolinone hydrazone assay as a method for quantification may also have caused difficulties as the most abundant sugars present in biomass hydrolysates (Glc and Xyl) have very different response factors (Abs/nmol sugar)(Gomez *et al.*, 2010). At the concentrations used in the diversity panel (40-110 nmol/mg DM) (Oakey *et al.*, 2013), every nmol of Xyl produced from the biomass would be quantified as ≈2 nmol of Glc (Gomez *et al.*, 2010). Therefore, even subtle differences in Xyl released between cultivars could cause major discrepancies in quantification, even before questioning whether the sample concentration was in the same, linear region as the calibration standards - a surprisingly common error with HT platforms (Lindedam *et al.*, 2014).

The aforementioned examples provide excellent illustrations of the key difficulties in developing HT screening techniques. It cannot be overstated that the careful consideration of screening techniques is fundamental to the success of any HT platform. Therefore, in this section, the approach taken to pretreat, dispense, hydrolyse and quantify the sugars released from biomass is outlined, attempting to minimise the variance introduced at each stage. This included the optimisation of the dinitrosalicylic acid (DNS) assay for quantifying RS abundance in biomass hydrolysates. In the past, this method has been discounted as too difficult to optimise for HT quantification (Gomez *et al.*, 2010), due to its comparatively low sensitivity to sugars and requiring high temperatures and heating durations to achieve coloration (Decker *et al.*, 2009).

Nevertheless, the DNS method is routinely used to estimate the concentration of RS in enzymatic hydrolysates, offering a non-specific alternative to more in depth and time-consuming methods such as GC or HPLC. In order to facilitate HT screening of biomass, automation of this simple colorimetric assay has been sought (Decker *et al.*, 2003; Goncalves *et al.*, 2010; King *et al.*, 2009; Miyazaki *et al.*, 2006; Navarro *et al.*, 2010; Shankar *et al.*, 2009; Song *et al.*, 2010).

However, recent attempts to scale down the DNS method to a 96-well format suffer from a number of limitations. For example, previous multiplexed methods have improved liquid transfer times, but heating and cooling times remain high. In some cases, incubation time is lengthened considerably (10 min) to accommodate for poor heat transfer when heating reagents in microtitre plates (Goncalves *et al.*, 2010). Others heat the reagents using a thermocycler, but do not take advantage of the improved heat transfer and reproducibility afforded to this method. All heat the sample at 95-100 °C for 10 min to ensure full coloration (Decker *et al.*, 2003; Miyazaki *et al.*, 2006; Navarro *et al.*, 2010; Shankar *et al.*, 2009; Song *et al.*, 2010). Extended heating regimes at high temperatures not only limit the overall speed of the assay, but potentially cause

inaccuracies via evaporation (Decker *et al.*, 2003), particularly when using pierceable sealing mats (Navarro *et al.*, 2010).

In addition, the original DNS method was designed to detect low concentrations of RS in diabetic urine (< 0.2 g/L) as opposed to concentrated hydrolysates required for efficient fermentation (50-100 g/L) (Zacchi & Axelsson, 1989). Therefore, extensive dilutions of hydrolysates prior to or immediately after heating are often necessary to bring the reaction products within range of detection (Decker *et al.*, 2003). The dilution of samples should be avoided wherever possible as this can significantly reduce the precision of this technique through errors in liquid transfer (Bailey, 1988; Decker *et al.*, 2003).

Poor standardization of the DNS method is compounded by the use of a range of non-conventional wavelengths for final quantification. For example, current studies analyse the reaction products using wavelengths from as low as 490 nm (Xu *et al.*, 2010) to as high as 580 nm (Iandolo *et al.*, 2011) without consideration as to the effect this may have on accuracy.

In this chapter, the effect of quantification wavelength, heating regime and reaction volumes on assay performance is explored. The resulting assay provides a HT method specifically tailored to analyse reducing sugar concentrations in biomass hydrolysates.

The final challenge addressed in this selection was the selection of suitable conditions for screening. Logistically, screening parameters must be established and applied to large datasets. However, restricting screening conditions limits the interpretation of the results. This concept is crystallised in the "you get what you screen for" axiom established by (Decker *et al.*, 2009). In an attempt to address this challenge, straw from a single cultivar was processed under a range of pretreatment conditions and selected rational parameters based on this data.

Screening parameters used here were selected based on work conducted in this, and previous chapters (particularly Chapter 4) to yield ≈100% (high PT, high cellulase dose), ≈50% (high PT, low cellulase), ≈50% (low PT, high cellulase) and ≈25% (low cellulase, low PT) Glc yields. Fermentation inhibitors released to the pretreatment liquors and sugar yields obtained from the pretreated residue after hydrolysis, were quantified from 77 straw accessions, derived from 49 cultivars, to demonstrate the intraspecific diversity in traits relevant to bioethanol production. By taking observations at various levels of biomatrix opening, a fuller impression of how cultivar variation interacts with processing parameters resulting in the final yield was obtained.

6.3 Materials and methods

6.3.1 Materials

Method development work was conducted using OSR straw (stems and empty pods) sourced from Hemp Technology Ltd. Suffolk, UK (52°21'15.7"N 1°30'35.7"E). Straw was harvested in 2011, chipped into < 350 mm pieces, dust extracted and baled. The straw was stored as a 20 Kg bale in a dry, unheated room before analysis (< 1 year). Further size reduction was necessary to gain a representative sample, therefore chipped straw was milled to < 2 mm, followed by freeze-milling using methods outlined in Chapter 2.

Screening work used threshed straw collected by Andrea Harper (University of York) and colleagues. Straw from 77 *B. napus* accessions (49 cultivars in total) was sourced from existing diversity panels grown at two sites - KWS UK Ltd., Cambridge (+52°, 5', 51.24", +0°, 6', 19.66") and Rothemstead Research, Harpenden (+51° 48' 31.49", -0° 21' 19.99") and collected over two years (2009-10 and 2010-11). The majority of accessions (n = 63) were collected from KWS, UK (Y_1 = 21, Y_2 = 42) with additional accessions sourced from Rothemstead Research (Y_1 = 14). The materials were grown under field conditions in adjacent 100 x 150 cm plots arranged in randomised blocks. The straw from three to five individual plants were collected at maturity, threshed by hand and stored in a dry, unheated room before analysis.

Straw from each cultivar (stems and empty pods) was chipped into 1-2 cm pieces using an industrial pin-mill, homogenised further to pass a 2 mm mesh and a representative sample (≈3 g) cryogenically ground into a fine powder suitable for screening.

6.3.2 Pretreatment

Freeze-milled straw from each cultivar were made into 15 mL slurries containing 5% substrate (FWB) in ultrapure H_2O (Milli Q) and sealed in 10-20 mL Biotage Microwave Vials. The milled straw was pretreated in a randomised order using a Biotage Initiator$^+$ Microwave Synthesiser for 10 min at the desired temperature. A sample of the liquor (1 mL) was taken so organic acid and furfural derivatives released from the materials could be quantified. The water insoluble pretreated substrates were extensively rinsed by centrifuging (2000 RPM, 2 min) and replacing the supernatant with ultrapure water (MilliQ) until clear, followed by sodium acetate / acetic acid (0.1M, pH 5) to remove soluble material and equilibrate the solutions to pH 5. Microbial activity was prevented using thiomersal (0.01% w/w).

6.3.3 **Enzymatic hydrolysis**

The rinsed slurries from each cultivar (1 mL) were transferred to eight polypropylene cluster tubes (1.2 mL, Corning, NY) using a multipipette equipped with wide orifice tips. The position of the tubes were randomised over four 96 well plates (Figure 26) and enzymatically hydrolysed simultaneously using either 5 or 25 µL of Accellerase 1500 at 50 °C (Microtherm, Camlab) with a batch cellulase activity of 72.9 FPU/mL. After 96 h, the racked tube plates were centrifuged (2000 RPM, 3 min) and the RS and Glc concentration of the liquor was quantified using DNS and GOPOD, respectively.

		1	2	3	4	5	6	7	8	9	10	11	STD
Plate 1	A	125A	23B	2A	50	106	100	125	143	180	168	130	0
	B	86	164	BLANK	73	42	130B	111A	151A	65	78	184A	5
	C	2	70A	133A	24A	37	177	150	15A	155	43	28	10
	D	7	40	91B	141A	184	166	131	100A	53	185	91	15
	E	49	166B	164A	11	60	99A	80A	37B	BLANK	23A	49A	20
	F	24	155A	141	80B	62	99	47	50A	169	141B	16	25
	G	151	29	110	111	45	24B	21	73A	52	143B	52B	30
	H	1	48	57	23	115	15	120	133	52A	70	140	35

		1	2	3	4	5	6	7	8	9	10	11	STD
Plate 2	A	151	73	185	11	49	168	115	133	42	52A	52	0
	B	86	21	23A	1	7	140	70	169	15A	49A	143B	5
	C	164	23B	100	106	99A	120	151A	133A	24A	40	130B	10
	D	52B	141B	166B	91	37	155A	110	111	184A	23	80B	15
	E	78	50A	131	2	15	80A	150	141A	48	BLANK	141	20
	F	28	62	65	16	184	73A	37B	166	180	100A	24	25
	G	60	164A	50	47	177	143	125	BLANK	24B	91B	155	30
	H	2A	53	45	130	29	111A	43	70A	57	99	125A	35

		1	2	3	4	5	6	7	8	9	10	11	STD
Plate 3	A	2	155	131	130B	78	106	37B	166	140	47	99A	0
	B	166B	185	57	15A	45	133	49	125A	100A	23	130	5
	C	24B	29	177	15	111	141A	42	21	120	100	184	10
	D	24A	151	115	62	80A	53	73A	169	180	184A	168	15
	E	7	70A	60	91B	16	155A	65	43	99	50A	91	20
	F	52A	23B	133A	52	24	143	125	BLANK	49A	52B	111A	25
	G	50	141	110	164	48	164A	150	23A	73	37	11	30
	H	151A	BLANK	143B	1	86	80B	28	2A	141B	70	40	35

		1	2	3	4	5	6	7	8	9	10	11	STD
Plate 4	A	52A	131	120	24	143	110	21	166B	180	43	184A	0
	B	47	15A	185	52B	62	37	143B	80A	91B	52	23	5
	C	29	57	24B	184	99	130	50	168	70	133A	151A	10
	D	166	130B	2	100A	65	86	155A	23A	42	141A	150	15
	E	111	78	151	40	BLANK	91	73A	125	23B	1	2A	20
	F	106	140	11	141	115	BLANK	24A	141B	49A	45	50A	25
	G	100	28	15	164A	111A	49	80B	125A	155	16	177	30
	H	48	37B	99A	7	164	60	73	169	70A	133	53	35

Figure 26 – Randomised hydrolysis plate layouts, including room for 0-35 g/L glucose standards in the final column (STD). Cellulase controls (substrate BLANK) were also positioned at random locations across the four plates.

6.3.4 Statistical analysis

All statistical analysis was conducted using Genstat 15.2 (VSN International Ltd.). Restricted maximum likelihood (REML) was used to predict the average product yields derived from each cultivar to control for year/site related differences observed in the dataset. Correlations were produced using R version 3.0.2.

6.4 Results and discussion

6.4.1 Small-scale pretreatment of biomass

To allow screening of straw samples from multiple cultivars, it was necessary to gain consistent pretreatment of small, reprehensive biomass samples. As illustrated in a previous section, fine milling of material acts as a pretreatment in itself, opening the structure of the biomass and therefore increases saccharification yields (Chapter 4). Therefore, milling the substrate before processing is not ideal. Nonetheless, the heterogeneity of lignocellulosic biomass is the main challenges to gaining representative results at a small scale (Gomez *et al.*, 2010). Fine-milling of biomass is therefore an unfortunate but essential prerequisite for all biomass screening platforms, so that a representative sample can be analysed at a small-scale (Chundawat *et al.*, 2008).

All samples were treated in an identical manner and in a randomised order to remove any structure introduced at this stage. Nevertheless, it is important to note that the results produced using screening conditions may underestimate the role of physical structure in determining the saccharification performance of these materials (Vidal *et al.*, 2011).

Freeze-milled samples of *B. napus* straw (0.75 ± 0.0001g, RSD ± 0.013%) were weighed into 10-20 mL Biotage Microwave Vials, adding 14.25 mL of dd$H_2$0 before pretreatment. Pretreating larger quantities of straw minimised error due to weighing compared to smaller-scale alternatives (Gomez *et al.*, 2010). Air-dried *B. napus* straw was used in this study, as opposed to drying the sample before analysis. Air-dried straw was chosen as the variations in moisture content (MC) between a randomly selected group of cultivars was very small (MC = 6.36 ± 0.23% FWB, n = 24).

Pretreating biomass slurries in a Biotage® Initiator+ allowed very precise control of the pretreatment applied to each vessel when pretreating continually stirred biomass slurries at various temperatures (Figure 27). Moreover, holding the samples in airtight pressure vials throughout pretreatment provided a fully closed system, increasing the precision in which product yields could be calculated.

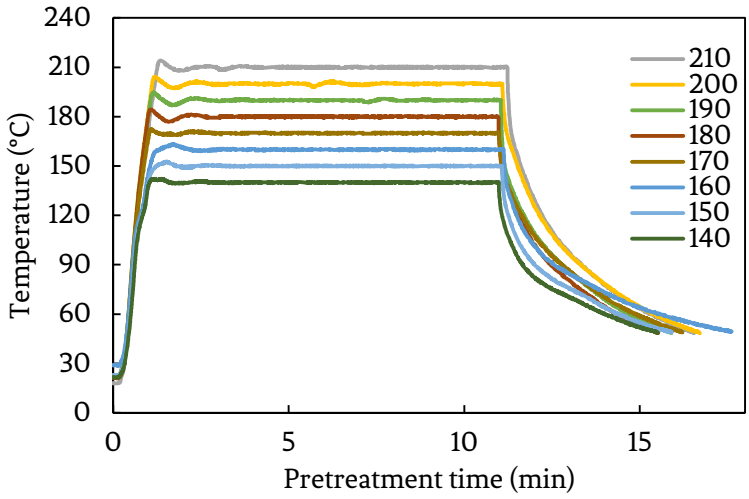


Figure 27 – Small scale pretreatment of freeze-milled OSR straw (15 mL) using a Biotage® Initiator⁺ Microwave Synthesizer (Biotage®, UK) – 5% [substrate], adsorption level "Normal", 20 mL pressure tubes, holding at desired temperature (140-210°C) for 10 min.

Another advantage of using this equipment was that the temperature of the vial was logged approximately every 1 sec, allowing the precise calculation of the pretreatment severity factor that the vial was exposed to over the course of a run. The variation in pretreatment severity between each cultivar was estimated by integrating the area under each curve. The variation in severity observed between samples during each screen was therefore very small (R_o = 3.51 ± 0.0018 and 4.24 ± 0.0008, RSD = 0.02% and 0.05% respectively, n = 77), allowing accurate and reproducible pretreatment.

6.4.2 Solid transfer of pretreated substrate to 96-well tube plates

Precise biomass delivery is perhaps the most challenging part of developing any HT assay using lignocellulose (Chundawat *et al.*, 2008). Biomass suspensions (or slurries) are often used in HT screening platforms. These techniques can be used to conduct multiple saccharification assays simultaneously and in a reproducible manner (Decker *et al.*, 2009). Transferring pretreated slurries has a number of benefits over small scale weighing of biomass, for example: it is less labour intensive and allows the use of material that has not been dried between pretreatment and hydrolysis. This prevents the partial reversal of biomatrix opening by drying pretreated biomass before hydrolysis (Zhao *et al.*, 2012).

To assess the potential of using pretreated OSR slurries for screening, 15 mL suspensions of freeze-milled steam exploded OSR straw (220°C, 10 min) were made at various concentrations (0.5-10% substrate, w/w) in ddH_2O and transferred using wide-bore tips to individual, pre-weighed tubes held in a 96-well tube rack. The slurries were dried to constant mass (105°C) and re-weighed to estimate the transfer of

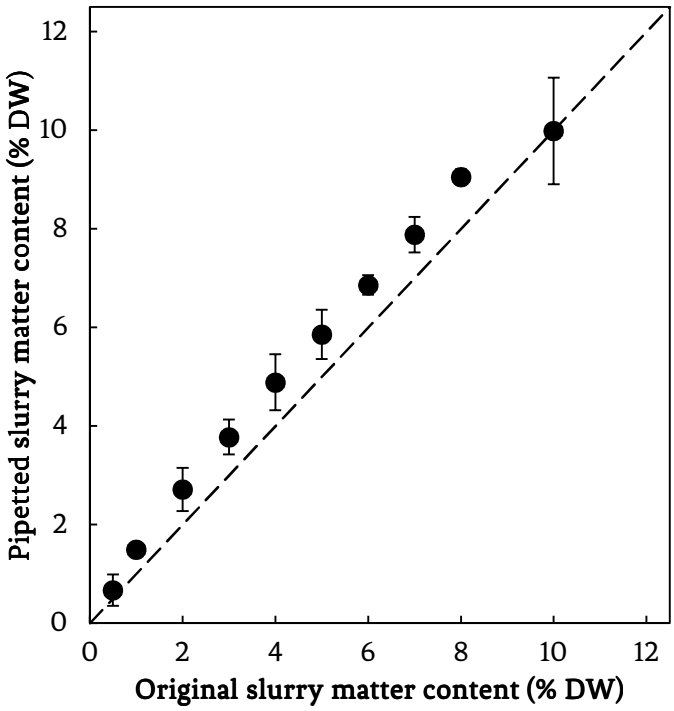


Figure 28 – Accuracy and precision of transferring 1 mL aliquots of freeze milled pretreated OSR straw at varying substrate concentrations.

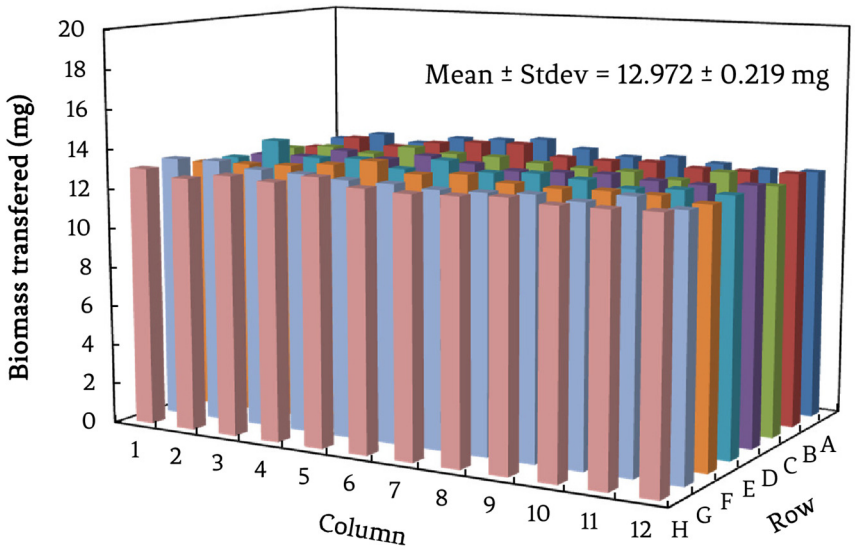


Figure 29 – Variation in 1% slurry transfer across a 96 well plate, RSD ± 1.69%, n = 96.

substrate the plate (Figure 28). Slurries that contained over 10% substrate (w/w) were too viscous to pipette. However, pipetting of slurries below this concentration was possible. Slightly more substrate was transferred to the plate than the original substrate concentration (Figure 28), but the substrate could be transferred to 1 mL tubes in a consistent and reproducible manner (Standard error of the mean = 0.2-0.5% DW, n = 3). In a separate experiment, pipetting a 1% DW biomass solution over an entire 96-well plate, the transfer of slurries was again relatively precise (12.97 ± 0.22 g/L, RSD ± 1.69%, n=96) (Figure 29).

6.4.3 Optimising the DNS assay for high throughput sugars analysis

Biomass hydrolysis can be conducted at a small scale by transferring biomass slurries into individual 1 mL tubes and adding cellulase. Following this, a similarly rapid methods for product detection is required (Decker *et al.*, 2009). Chromatographic methods are time consuming, as samples must be analysed consecutively. Colorimetric methods are more useful for analysing multiple samples, as a whole plate (n = 96) can be analysed simultaneously. Therefore, colorimetric methods are a desirable quantification method for HT screening. As final reaction volumes of multiplexed screening methods are typically small < 300 µL (Berlin *et al.*, 2006), the selected method also needs to operate using small sample volumes. Here, a common method of determining the concentration of RS in biomass hydrolysates was optimised to analyse samples at a microplate scale.

6.4.3.1 Scaling and optimising reaction volumes

The DNS method is a widely-used, non-specific method of quantifying reducing sugars in solution. Typically, this assay is conducted in test tubes. Three mL of sample is added to 3 mL of DNS reagent and boiled (5 min) before quantification at 575 nm (Miller, 1959). When using conventional sample to DNS ratios (1:2, sample: DNS reagent or 50% v/v) (Miller, 1959), extensive dilution of concentrated hydrolysates is sometimes necessary to bring them into a readable range. The dilution of samples prior to analysis reduces precision and extends assay time by adding further liquid transfer steps. Moreover, appropriate dilutions require *a priori* knowledge of the RS concentration in the hydrolysate.

Two calibration curves covering 0-4 g/L (0.1 g/L increments) and 5-29 g/L (1 g/L increments) were analysed using varying ratios of sample to DNS (1, 5, 10, 20, 30, 40 and 50% sample/DNS reagent, v/v) made up to 180 µL to establish if smaller sample volumes could be used to eliminate the need for sample dilutions. Solutions were heated (5 min, 100 °C) and quantified at 540 nm (Figure 30). At relatively low Glc concentrations (0-30 g/L), a sample to DNS ratio of 1:20, sample: DNS reagent (5% v/v)

maximum linearity and resolution was achieved (y = 0.0919x - 0.065, r^2 = 0.9993, Figure 30). Larger sample volumes improved resolution but impaired linearity over this range; the opposite was true for smaller sample volumes. For this reason, 1:20, sample: DNS reagent was selected for further optimisation.

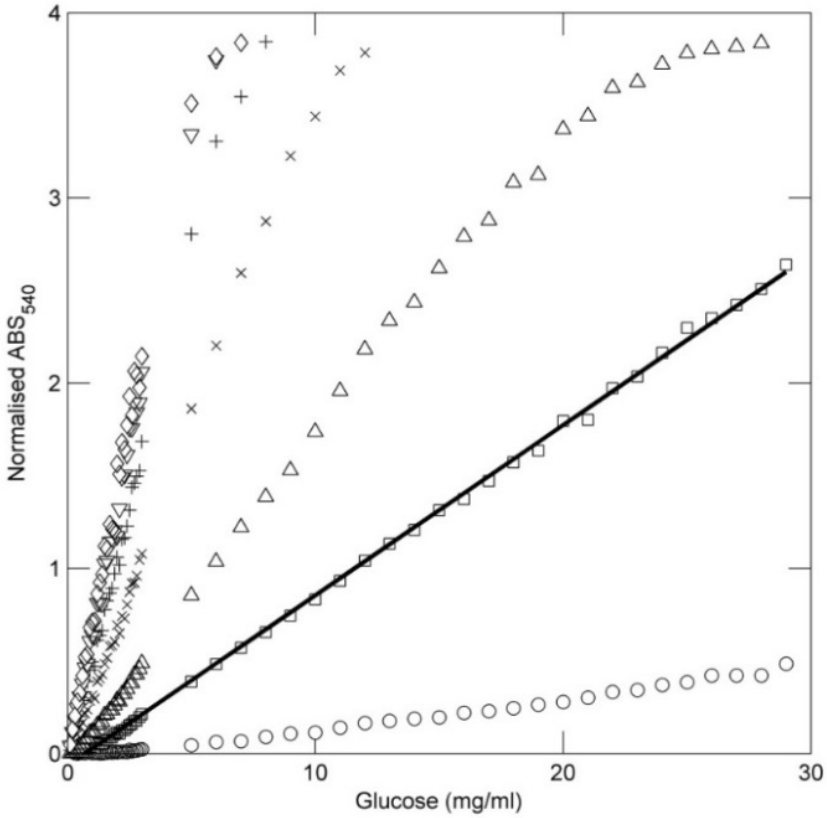


Figure 30 – Differing ratios of sample to DNS on the absorbance of the reaction products after heating at 100 °C, 5 min, read at 540 nm. Samples contained 1% (○), 5% (□), 10% (△), 20% (x), 30% (+), 40% (▽) or 50% (◇) sample/DNS reagent (v/v). Reaction volumes containing 5% sample/DNS reagent (v/v) (or 1:20 sample: DNS reagent) were selected as they displayed a linear calibration curve (black line, y = 0.0919x - 0.065, r^2 = 0.9993) over this range (0-29.5 g/L n = 60).

6.4.3.2 **Optimising temperature and timing regimes**

When using a manual DNS method, the time taken to transfer liquids is much greater than the sum of all the other steps, lengthening the total assay time. However, multiplexing considerably improves the speed of liquid handling (1 min 50 s, in total) such that the time required to heat the reaction mixture becomes a significant time restraint.

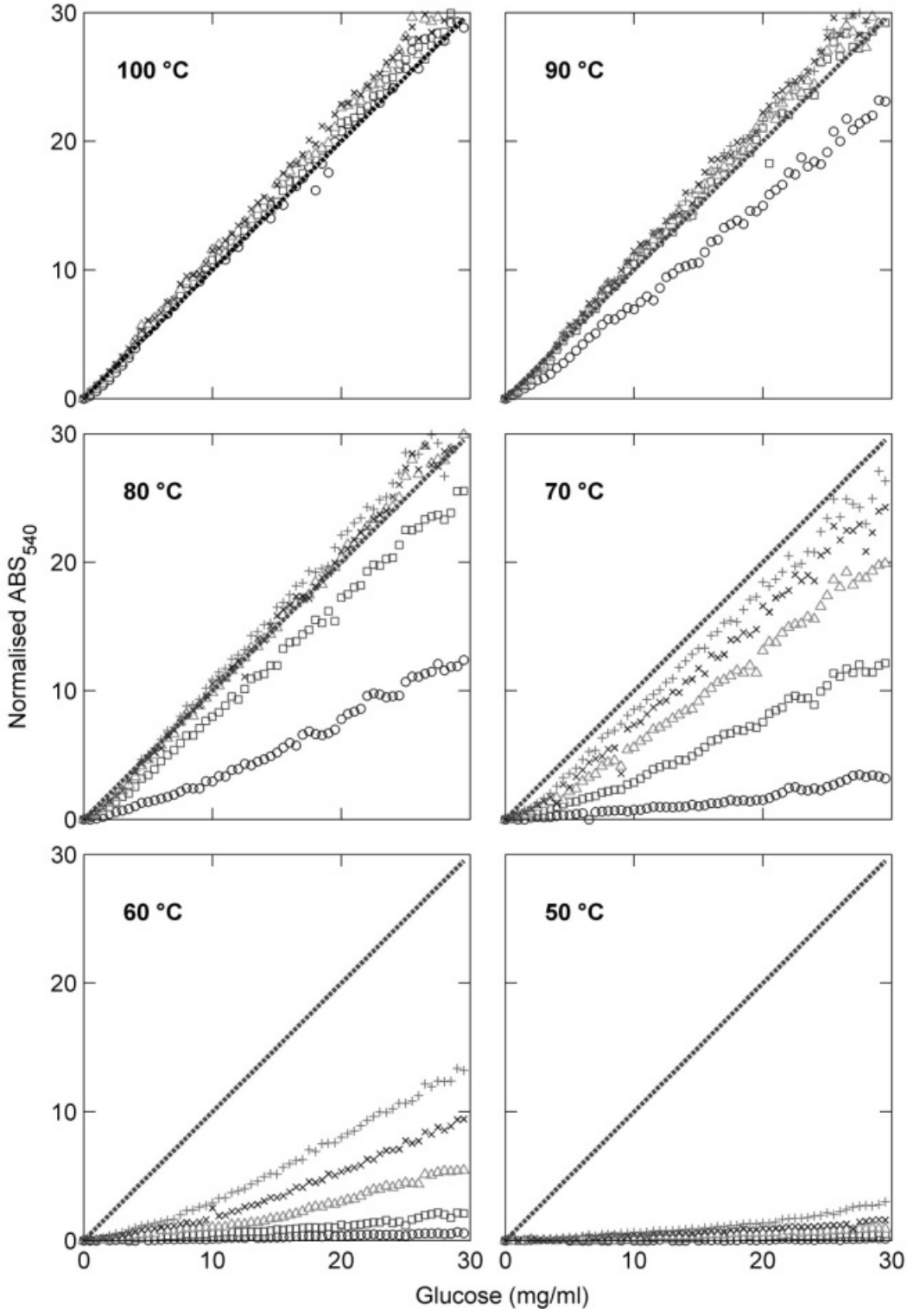


Figure 31 – Optimising reagent heating regimes. Glucose standards (0-29.5 g/L *n* = 60) were added to DNS reagent (1:20 standard: DNS reagent) and heated at a variety of temperatures (50-100 °C) for various lengths of time: 1 (○), 2 (□) 3 (△) 4 (x), 5 (+) min. The dashed line is indicative of full coloration (100 °C, 5 min). Absorption was quantified (540 nm) and values have been expressed relative to full coloration (x = y).

Following the original DNS method (Miller, 1959), reaction mixtures were heated in test tubes for 5 min, 100 °C. However, most recent multiplexed methods heat the reagents for considerably longer lengths of time (10 min) despite being heated in a thermocycler which should improve heat transfer and provide more precise heating control (Miyazaki *et al.*, 2006, Navarro *et al.*, 2010, Shankar *et al.*, 2009, Song *et al.*, 2010).

To establish the most rapid heating regime necessary to ensure full coloration, a sixty point Glc calibration curve (0-29.5 g/L, 1:20 sample: DNS reagent) was generated from samples heated for varying temperatures and durations (Figure 31). In a thermocycler, heating at 100 °C, 1 min was sufficient to gain full coloration. This is significantly faster than previous methods which typically involve heating periods of up to 10 min (Miyazaki *et al.*, 2006, Navarro *et al.*, 2010, Shankar *et al.*, 2009, Song *et al.*, 2010).

In certain circumstances, using lower temperatures may be preferable to limit evaporation e.g., when using pierceable adhesive sealing mats (Navarro *et al.*, 2010). When heating samples at lower temperatures linear calibration curves ($r^2 > 0.99$) can also be obtained, but longer heating durations are needed (1, 2 and 5 min at 90, 80 and 70 °C respectively). Even longer heating times are needed to ensure full coloration of the samples (2, 3 and > 5 min for 90, 80 and 70 °C respectively). These heating regimes, although not the fastest, could be used to minimise evaporative loss and allow greater flexibility in the sealing method used.

6.4.3.3 **The effect of reading wavelength on quantification**

A range of wavelengths are currently used to quantify the end product of the DNS reaction. To observe the effect this has on reading resolution and precision, two overlapping Glc calibration curves – 0-29.5 g/L (n =60) and 25-100 g/L (n = 10) – were generated (1:20 sample: DNS reagent, heated 100 °C, 1 min) and analysed at varying wavelengths (Figure 32).

When quantifying reaction products at narrower wavelengths (approaching 500 nm), greater resolution is attained, but linearity is restricted to intermediate wavelengths (Fig 3). Wavelengths between 540–605 nm all yield linear calibration curves ($r^2 > 0.9999$) and are therefore suitable to use when sample concentration is < 30 g/L. Optimum linearity is reached when reading at 575 nm ($r^2 = 0.999986$), but quantifying the reaction products at this wavelength causes poorer resolution (slope = 0.024480). Conversely, reading at 520 nm gave optimum resolution (slope = 0.133689) but linearity was impaired ($r^2 = 0.993796$). The greatest resolution, while still gaining a linear calibration curve ($r^2 > 0.9999$) was obtained at 540 nm (slope = 0.079916, $r^2 = 0.999927$) therefore this wavelength was selected for further optimisation.

Unfortunately, calibration curves begin to plateau at concentrations > 32 g/L when reading at 540 nm, limiting the use of just a single wavelength for quantification. However, when analysing samples with higher RS concentrations, longer wavelengths can be used to maintain linearity.

When analysing more concentrated substrates (30-100 g/L) linear calibration curves are generated at wavelengths between 580 – 660 nm (r2 > 0.9980). Optimum linearity over this range was obtained at 600 nm (r^2 = 0.99851) but again, resolution is decreased (slope = 0.0135). Reading at 580 nm gives the greatest possible resolution (slope = 0.0228) while maintaining a linear calibration curve up to 100 g/L (r^2 > 0.998).

Therefore, to analyse samples of unknown concentrations two calibration curves covering 0-25 g/L and 25-100 g/L can be included on the same plate. The former calibration curve should be read at 540 nm and the latter at 580 nm – an appropriate wavelength should be selected depending on sample absorbance.

Although intermediate wavelengths could be used, reading at two wavelengths ensures the best resolution and linearity is achieved when analysing a particular sample. Using a longer quantification wavelength is also preferable to diluting samples as this introduces inaccuracies and increases the overall assay time (Bailey, 1988; Decker *et al.*, 2003).

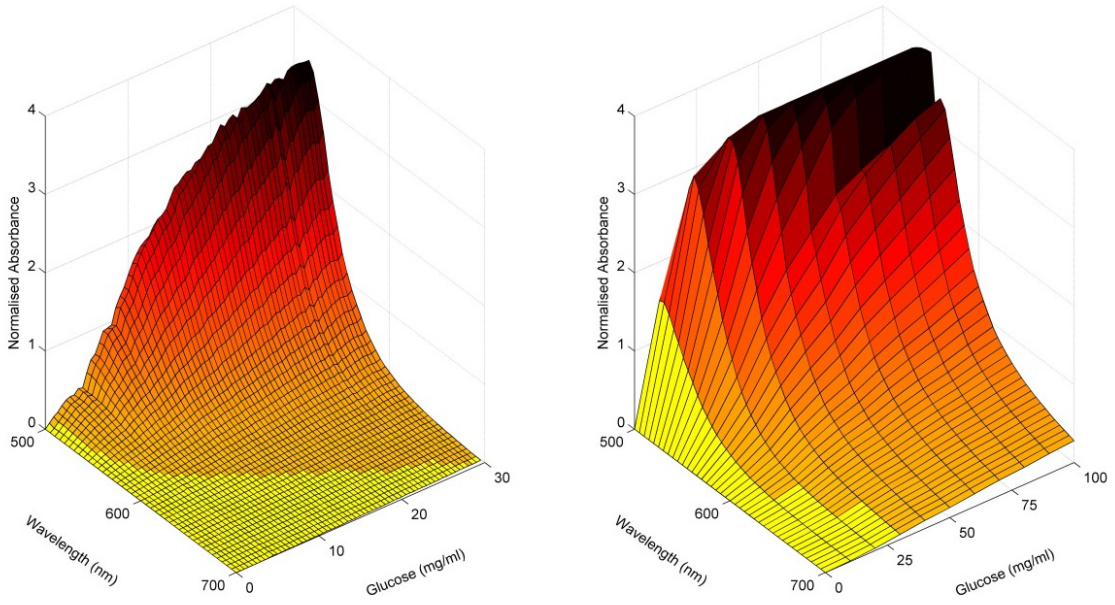


Figure 32 – Changes in resolution and linearity when two glucose calibration curves (0-29.5 g/L, *n* = 60, LHS and 0-100 g/L, *n* = 10, RHS) are generated using 1:20 standard: DNS ratios, heated (100 °C, 1 min) and quantified at differing wavelengths.

6.4.4 **Establishing pretreatment conditions for screening**

Initially, a batch of commercially grown *B. napus* straw derived from a single cultivar was pretreated in liquid hot water at varying severities (130-220°C, 10 min) and hydrolysed with an excess of cellulase (50 FPU/g original material) to establish rational process conditions with which to screen (Figure 33).

By modelling the sigmoidal curves of Glc and RS release under varying severities, sugar yields at all pretreatment temperatures from 130 to 220 °C can be predicted at any severity (Table 12) using the following equations:

$$Reducing\ sugar\ yield\ (mg/gFW) = \frac{268.28}{1 + Exp(-0.182 * (Temp - 175.14))}$$

Equation 5

$$Glucose\ yield\ (mg/gFW) = \frac{252.97}{1 + Exp(-0.105 * (Temp - 185.09))}$$

Equation 6

Liquid hot water pretreatment resulted in a 6-fold increase in Glc yield under optimal conditions (≈220°C, 10 min). The maximum Glc yield that could be achieved using this straw was ca. 252.97 mg/g of original straw (FW) (ca. 88.08% max Glc). The hydrolysis of other (non-Glc) CW carbohydrates increased until an intermediate pretreatment severity was reached (ca. 185°C, 10 min), after this point the breakdown of these less-robust sugars becomes greater than the rate of release (Figure 33). These parameters were therefore used to calculate precise pretreatment conditions which to screen (Table 12 & Figure 33). The temperatures 185 °C and 210 °C were chosen as they yield ≈50% and ≈90% of the maximum Glc yield released by LHW pretreatment respectively.

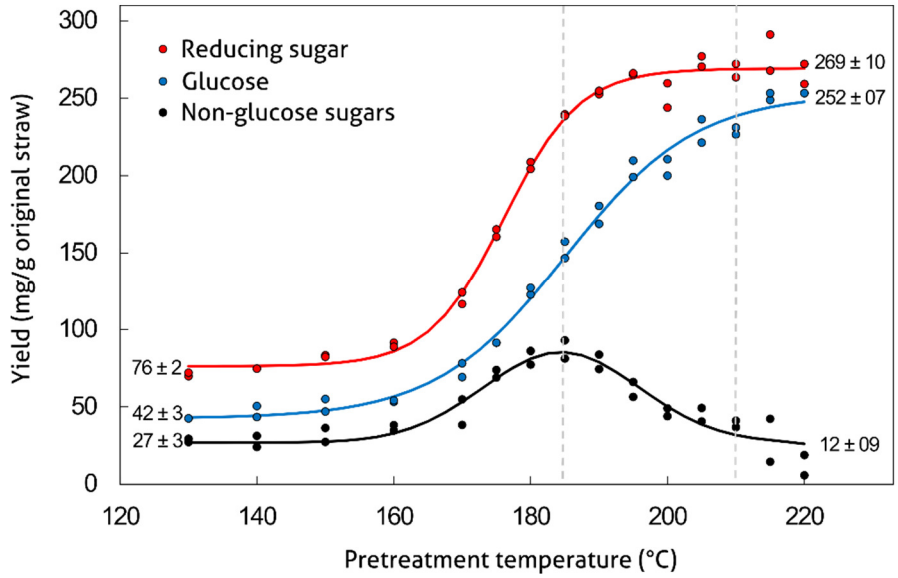


Figure 33 – Changes in reducing sugar (red) and glucose (blue) yields from OSR straw after pretreatment at a range of severities (130-220 °C for 10 min) and hydrolysed with excess cellulase (50 FPU/g). The abundance of non-glucose sugars released from the water insoluble portion was estimated by subtracting the abundance of glucose from the abundance of reducing sugars (black). Minimum and maximum yields predicted by modelling the curves are displayed on the left and right of the graph, respectively. Vertical dashed lines indicate selected screening conditions.

	Base Yield (A, mg/g)	**Max additional yield (C, mg/g)**	**Mid-point (M, °C)**	**Slope (B)**
Reducing sugars	76.22 ± 2.98	193.06 ± 4.03	175.14 ± 0.46	0.182 ± 0.014
Glucose	42.48 ± 3.38	210.49 ± 6.65	185.09 ± 0.76	0.105 ± 0.008

Table 12 – Parameters describing the relationship between pretreatment temperature and sugar yield from OSR straw. Data (n=30) was fitted to the following equation: Sugar yield = A + C/(1 + EXP(-B*(temp - M))). Parameters can therefore be used to estimate the Glc and RS yield expected at any pretreatment temperature between 130 and 220°C.

6.4.5 **Phenotype data collected from various cultivars**

After establishing rational pretreatment conditions, hydrolysis conditions were selected based on work in Chapter 4 (Figure 20, pg. 93). Conditions were selected to yield a range of saccharification yields ranging from ≈25% to 90% of the maximum Glc yield using a combination of pretreatment severity and cellulase dose. These conditions selected in this study were as follows:

(I) 185L – LHW 185°C, 10 min, hydrolysis ≈7 FPU/g original material.

(II) 185H – LHW 185°C, 10 min, hydrolysis ≈36 FPU/g original material.

(III) 210L – LHW 210°C, 10 min, hydrolysis ≈7 FPU/g original material.

(IV) 210H – LHW 210°C, 10 min, hydrolysis ≈36 FPU/g original material.

Combinations of pretreatment severity and cellulase dose were used to gain four levels of biomatrix opening, limited by various parameters. The main limitation for each of these conditions were: cellulase dose and pretreatment severity (185L), pretreatment severity (185H), cellulase dose (210L), and neither cellulase dose nor pretreatment severity (210H).

Product yields were collected from 77 accessions of field-grown *B. napus* straw, gathered from existing diversity panels. The majority of accessions (n = 63) were collected from KWS, UK (Y_1 = 21, Y_2 = 42) with additional accessions sourced from Rothemstead Research (Y_1 = 14). The principle sugar breakdown products released into the pretreatment liquor were quantified for all cultivars at both severities (185 and 210 °C). Water insoluble material produced after pretreatment was then rinsed and hydrolysed at two cellulase concentrations (ca. 7 and 36 FPU/g original material, L and H respectively). In total, data regarding sixteen traits relevant to bioethanol production were collected, under various conditions as follows:

(I) Acetic acid yields after pretreatment (185 and 210)

(II) Formic acid yields after pretreatment (185 and 210)

(III) Furfuraldehyde yields after pretreatment (185 and 210)

(IV) Hydroxymethylfurfural yields after pretreatment (185 and 210)

(V) Glc yields after pretreatment and hydrolysis (185L, 185H, 210L, 210H)

(VI) RS yields after pretreatment and hydrolysis (185L, 185H, 210L, 210H)

Full summaries of the screening data collected for each accession is included in the Appendix Tables 1 and 2 (pg. 226-234). Summaries for each cultivar, controlling for variations between year and site using Residual maximum likelihood (REML) estimation are included in Appendix Tables 3 to 6 (pg. 235-238).

REML estimation is an effective way of controlling for environmental conditions for field trials, which typically produce unbalanced datasets (Virk *et al.,* 2009). As straw was sourced from a number of origins, REML analysis was used as a control for field conditions, predicting means for each cultivar expected in an average year at an average site. This is also a common way to prepare data for GWA studies (Penning *et al.,* 2014). The significance of each term also indicates the effect of environmental variations on a particular product yield (Table 13).

The results illustrate that some traits were more sensitive to agronomic conditions than others. For example, Glc 185L, HMF 185, HMF 210, 185 2FA all showed significant variation across both years and sites (Table 13). These phenotypes are likely to be effected by low abundance, readily hydrolysed sugars (soluble sugars, residual starch and callose). These components are known to be more heavily influenced by environmental conditions, particularly the amount of residual starch retained in the straw (Decker *et al.*, 2012) and varying callose deposition in response to stress (Delmer *et al.*, 1993).

Other traits showed significant differences ($p < 0.05$) between years, but not sites (Glc210L, RS210L and Formic 185), as others showed differences between sites and not years (RS185L, Glc185H, 2FA 210). The likely reason for these differences is that growth conditions between years and sites resulted in quantifiable changes in CW composition that needed to be controlled for.

Trait	Year			Site		
	Wald	d.d.f.	*P*	Wald	d.d.f.	*P*
Dropping individual terms from full fixed model						
185 Glc Low	**6.62**	**297.5**	**p<0.05**	**17.79**	**303.4**	**<0.001**
185 RS Low	2.10	290.3	0.149	**4.12**	**303.3**	**<0.05**
210 Glc Low	**92.90**	**278.3**	**<0.001**	0.53	292.4	0.469
210 RS Low	**67.09**	**289.6**	**<0.001**	1.02	302.8	0.314
185 Glc High	0.21	269.9	0.644	**12.61**	**280.2**	**<0.001**
185 RS High	0.00	276.2	0.995	0.38	288.9	0.541
210 Glc High	0.23	293.8	0.632	3.52	274.3	0.062
210 RS High	2.33	299.7	0.128	0.62	303.6	0.432
185 Acetic	1.43	123.4	0.234	2.75	136.4	0.099
185 Formic	**82.83**	**123.7**	**<0.001**	0.37	137.2	0.546
185 2FA	**22.36**	**113.8**	**<0.001**	**13.99**	**122.2**	**<0.001**
185 HMF	**45.45**	**133.3**	**<0.001**	**23.95**	**146.5**	**<0.001**
210 Acetic	0.03	130.9	0.867	0.03	146.2	0.862
210 Formic	0.80	130.1	0.372	0.01	145.2	0.917
210 2FA	0.05	111.4	0.824	**8.36**	**120.2**	**<0.01**
210 HMF	**19.68**	**120.7**	**<0.001**	**22.02**	**134.6**	**<0.001**

Table 13 – Controlling for variation in traits caused by differences in years and sites using REML modelling (Yield = Constant + Year + Site + Cultivar).

Traits that were less dependent on agronomic conditions showed no significant effect of site or year ($p > 0.05$). These included Acetic 185, Acetic 210, Formic 210 and various saccharification yields produced using a high cellulase dose (RS 185H, RS 210H, Glc 210H). It is likely that these traits were less sensitive to environmental conditions, probably because the yields reflected variations in the majority of CW components available for hydrolysis, as opposed to more variable low abundance components.

Crucially, when cultivar number was added to the model as a fixed term, the effect of cultivar was always a highly significant term ($p < 0.001$). This demonstrates that genotype always had a significant effect on product yields for every trait, irrespective of conditions used.

6.4.5.1 Saccharification yields after processing

Screening conditions were carefully selected to give a range of bio-matrix openings – producing incrementally higher Glc yields by manipulating both pretreatment and hydrolysis conditions. The average Glc yields produced from the 49 cultivars under the four processing conditions were 71, 115, 174 and 234 mg/g original straw, for 185L, 210L, 185H and 210L respectively (Figure 34).

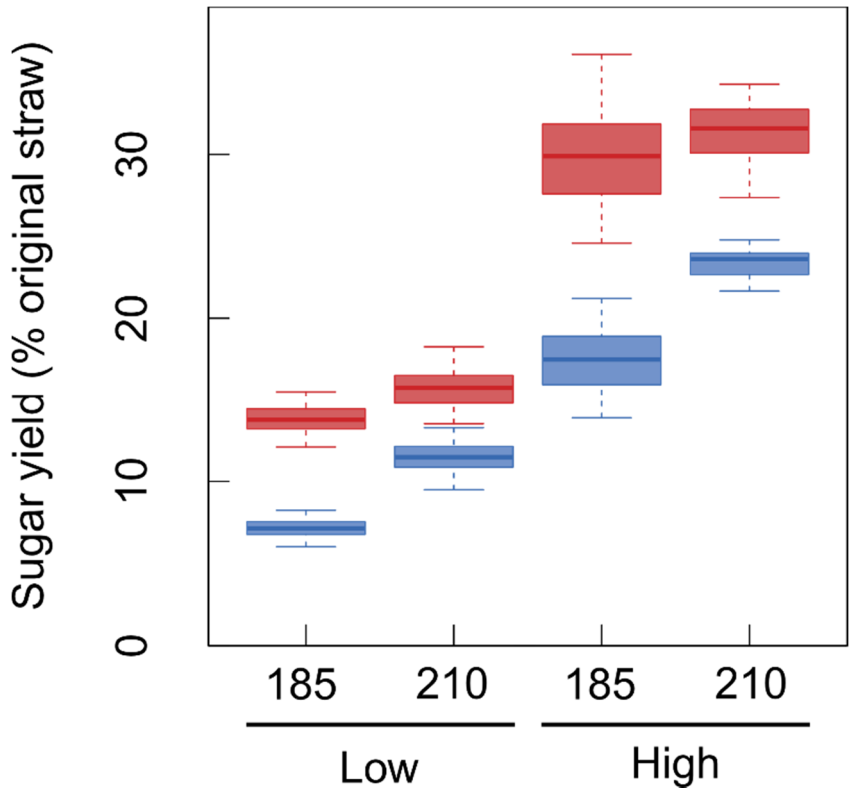


Figure 34 – Variation in reducing sugar (red) and Glucose (blue) yields from OSR straw derived from different cultivars (n = 49) after pretreatment at either 185 °C, or 210 °C for 10 min and hydrolysis at either a lower or higher cellulase dose (ca. 7 and 36 FPU/g original straw respectively).

In some cases, similar saccharification yields could be achieved when pretreating the straw from one cultivar at 185 °C, than another pretreated at 210 °C (Figure 34). This emphasises the potential benefit of selecting cultivars that produce more sugar at lower pretreatment severities, particularly if both hexose and pentose sugars are used for fermentation.

Pretreatment severity had the largest effect on variability in Glc yields between cultivars when sufficient cellulase was added. When *B. napus* straw was pretreated at 185 °C and hydrolysed with a high cellulase dose, the range of Glc yields between cultivars was higher (mean ± 42%) than when straw from the same cultivars were pretreated at 210 °C (mean ± 13.34%) (Figure 34). Therefore, in instances where feedstock composition can be assured, biorefineries could make significant savings in processing costs by using a cultivar that produces high sugar yields at low severities. On the other hand, if the processing plant is situated in an area where diverse feedstock genotypes are used, using a higher pretreatment severity will give greater uniformity in saccharification yields between cultivars.

6.4.5.2 **Glucose yield after processing, do you get what you screen for?**

Screening platforms typically limit processing regimes to those that yield near maximum Glc yields (Lindedam *et al.*, 2014). These conditions minimise the effect of screening conditions on the final yield thereby giving more uniform yields between platforms. However this potentially limits subsequent interpretation of the results as they may only reflect the variations in straw composition relevant to specific conditions. This concept is crystallised in the axiom, "you get what you screen for" (Decker *et al.*, 2009).

Recently, Lindedam *et al.*, (2014) compared three independently optimised screening platforms. Despite the differences in pretreatment and hydrolysis conditions used between the different laboratories, the different screening platforms were broadly comparable in terms of assessing the differences in saccharification potential of twenty wheat cultivars (Lindedam *et al.*, 2014). This suggests that variations in biomass recalcitrance between genotypes may be independent of screening conditions (Lindedam *et al.*, 2014). However, the screening platforms compared by Lindedam *et al.*, (2014) all pretreated finely milled straw at similar severities ($R_0 \approx 3.6$) and hydrolysed the remaining substrate at an above optimum cellulase dose (> 40 FPU) and duration (> 70 h). Restricting screening to 'optimal' conditions (severe pretreatment and high cellulase dose) may mask interesting differences in CW chemistry that could give higher sugar yields at lower energy inputs.

To understand the role of processing conditions on Glc yields produced from various *B. napus* cultivars, trait data was correlated to reveal consistent genotypic trends at various levels of biomatrix opening (Figure 35). Cultivars that released more Glc under one processing condition also released significantly more Glc under different processing conditions ($p < 0.001$, $n = 49$). This suggests that at least some of the properties that make a cultivar produce more Glc are common to all severities. Alternatively, Glc released under less severe processing conditions may largely determine the variation in Glc yields between cultivars.

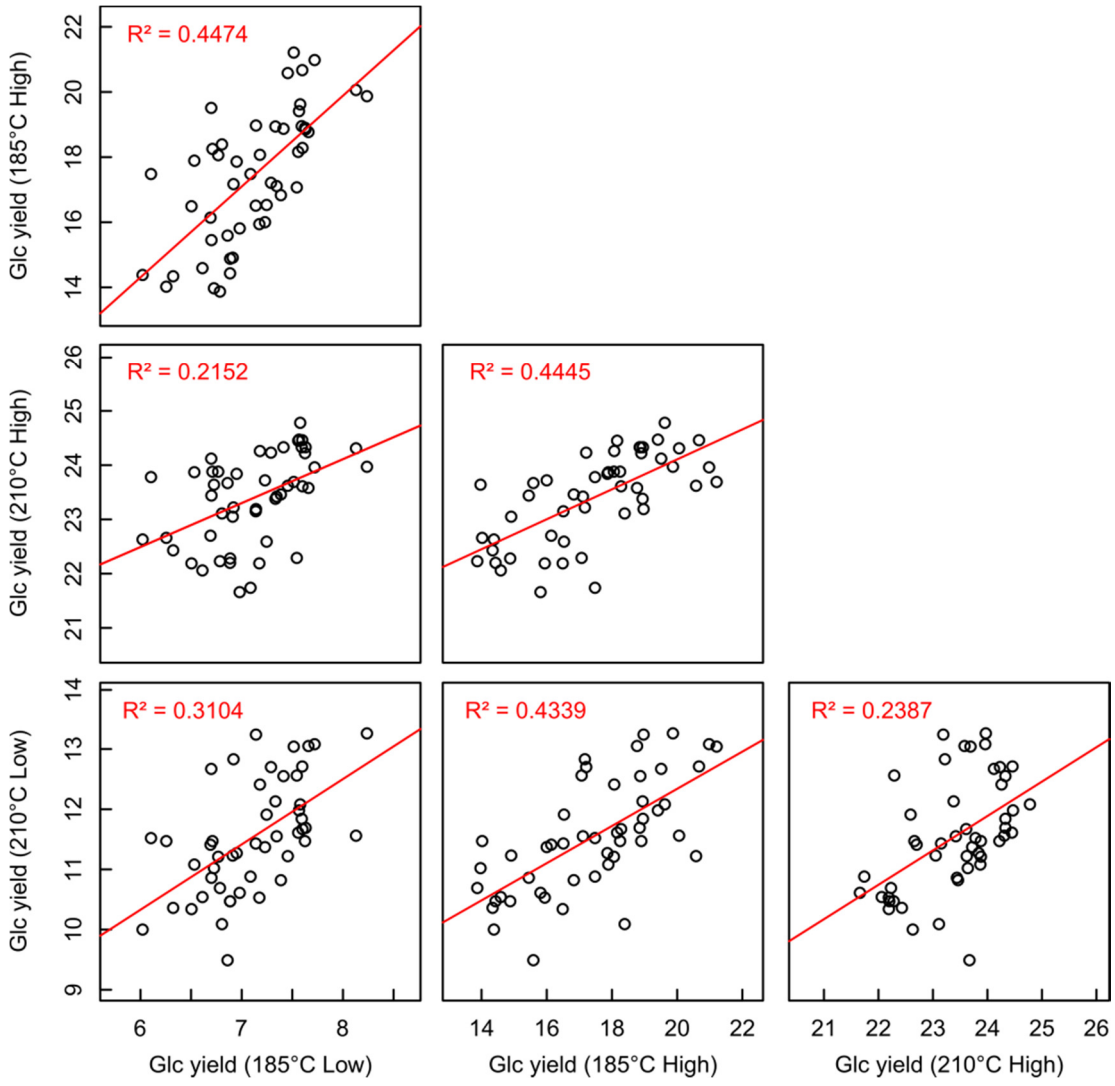


Figure 35 – Comparisons between glucose yields produced from *B. napus* straw derived from 49 cultivars after various processing regimes. Straw was pretreated in liquid hot water at either 185 °C or 210 °C, 10 min followed by hydrolysis using a higher (High) or lower (Low) cellulase dose (ca. 36 or 7 FPU/g original straw respectively).

Screening conditions that were limited by the same factor (either cellulase availability or pretreatment severity) correlated most strongly with each other ($r^2 > 0.44$) (Figure 35). Similarly, those limited by either cellulase availability or pretreatment severity (185H and 210L) also correlated well ($r^2 \approx 0.43$) (Figure 35). This suggests that the variation in Glc yields between cultivars is not dependent on either pretreatment or cellulase dose *per se,* but rather variations in the material itself.

By contrast, when Glc yields produced from most disparate screening conditions were compared (185L with 210H), variations in yields between cultivars did not correlate as strongly ($r^2 \approx 0.2$) (Figure 35). This suggests that some cultivar-specific differences in Glc yield are caused by particular processing conditions.

It would be interesting to see if variations in screening conditions utilising more different processing methodologies (for example, alkali or solvent-based pretreatments) also show commonalities in substrate-specific trends between processing regimes.

6.4.5.3 **Fermentation inhibitor yields after pretreatment**

When biomass is pretreated, partial hydrolysis of the material occurs. This causes some polysaccharides to depolymerise and degrade to organic acids and furfural derivatives (Palmqvist & Hahn-Hagerdal, 2000). The variations in concentration of these compounds reveal interesting information about how each cultivar responds to processing. Cultivars that produce greater concentrations of these compounds could have negative effects on downstream processes. However, these cultivars may contain less robust components that make processing more effective.

Consequently, four of the most common breakdown products formic acid, acetic acid, 2FA and HMF contained within the pretreatment liquor were quantified for each cultivar, pretreated at either 185 or 210 °C (Figure 36). The abundance of these products were correlated to reveal common variations in fermentation inhibitor production between cultivars (Figure 37).

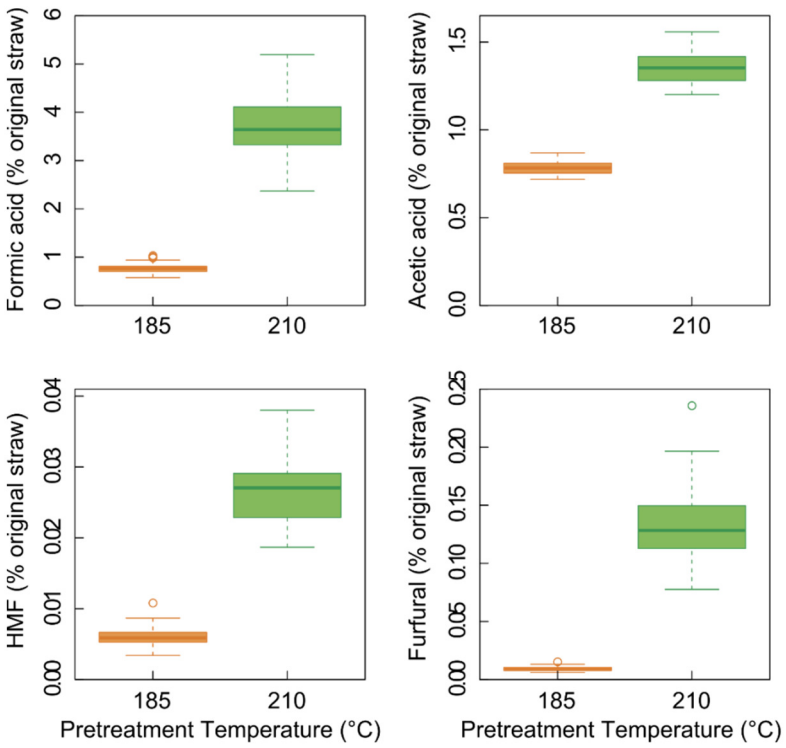


Figure 36 – Fermentation inhibitors released from *B. napus* straw derived from different cultivars (n = 49) after pretreatment at 185 °C (orange) or 210 °C (green) for 10 min.

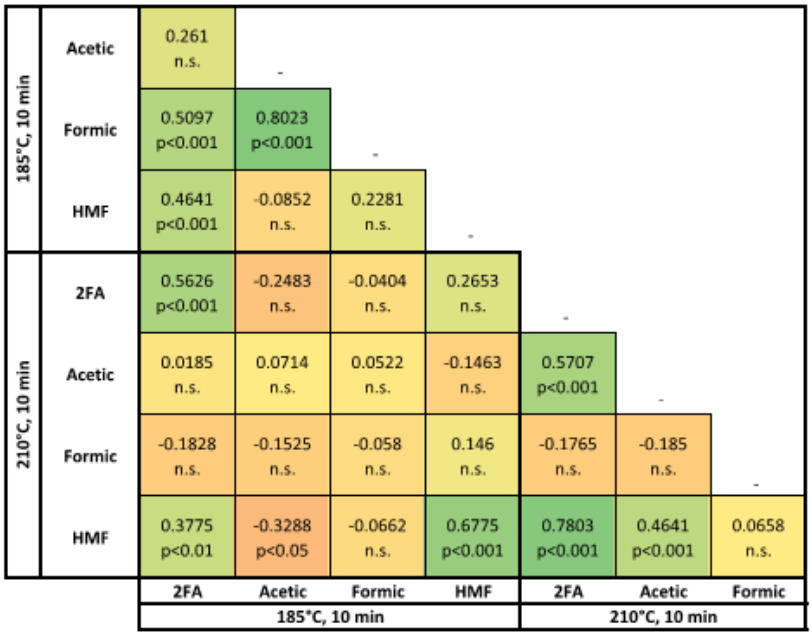


Figure 37 – Correlations between fermentation inhibitors contained within the pretreatment liquors after pretreatment (185 and 210 °C, 10 min), from straw derived from different *B. napus* cultivars (n = 49). Cells are coloured according to the direction of the regression line (+1 to -1). Results from a two-sided significance test from zero are displayed as either: not significant (n.s.) or significant to varying degrees ($p < 0.05$, $p < 0.01$, $p < 0.001$).

As expected, straw pretreated at 210 °C produced more fermentation inhibitors than when the same cultivars were pretreated at 185 °C. Acetic and formic acid were produced in similar proportions (≈ 1-2% original straw), but at the higher severity (210 °C), formic acid was the predominant break-down product released (2.5-5% original FW) (Figure 36). Pretreatment at 185 °C released ≈58% of the acetic acid produced at 210 °C from an average cultivar, whereas ≈21% of the formic acid and HMF released at 185 °C compared to 210 °C. By contrast, only 7% of the 2FA released at 210 °C was produced at 185 °C for an average cultivar. Collectively, these data suggest that at 185 °C less robust acetylated CW components rapidly hydrolyse to form formic acid (from pentose sugars), HMF (from hexose sugars) and acetic acid (acetyl groups). Only some 2FA was contained within the liquor, potentially reflecing the small amount of xylan hydrolysed at 185 °C compared to 210 °C.

The likely origin of HMF, 2FA and formic acid released at 185 °C could be from PCW components such as acetylated xyloglucans, which contain a relatively even proportion of pentose and hexose sugars. Only some of glucuronoxylan, the main repository of pentose sugars in SCW material (20-30% w/w) (Scheller & Ulvskov, 2010), may be hydrolysed to 2FA at 185 °C (Figure 37).

Under more severe pretreatment conditions it is likely that the majority of degradation products come from SCW components, namely glucuronoxylan. Unlike pretreatment at 185 °C, the variation in formic acid yields between cultivars did not correlate with other breakdown products when pretreated at 210 °C. Instead, cultivars that yielded more acetic acid also produced more 2FA and HMF (Figure 37). This is consistent with the idea that glucuronoxylan dominates the variation in organic acid abundance between cultivars when pretreated at 210 °C, whereas XG dominates those released at 185 °C.

The “you get what you screen for” axiom appears to be particularly important in regard to fermentation inhibitor yields, both severities rarely correlated (Figure 37). Nevertheless, cultivars that produced more HMF at 185 °C also produced more HMF and 2FA at 210 °C. Despite only 7% of the 2FA released at 210 °C was released at 185 °C, variations between cultivars also correlated (Figure 37). This suggests that the variation in 2FA released between cultivars may be determined at intermediate pretreatment severities. Strong associations between the ranges of HMF and 2FA abundances between cultivars pretreated at 185 °C and 210 °C suggest that much of the varietal variations in these components are determined at lower severities.

6.4.6 Biomass-derived organic acids catalyse pretreatment and increase glucose yields after hydrolysis at sub-optimal severities

Generally, organic acids and furfural derivatives are viewed as an inconvenient by-product of pretreatment (Jonsson *et al.*, 2013; Rasmussen *et al.*, 2013). However, the release of these products is not only an indication of the destruction of potentially fermentable sugars, but also of increased depolymerisation, biomatrix opening and cellulose accessibility. The release of biomass-derived organic acids during pretreatment is also assumed to catalyse depolymerisation further by lowering the pH of the pretreatment liquor (Hendriks & Zeeman, 2009) – but this has not been definitively shown. With this in mind, the abundance of organic acids released during pretreatment were correlated with sugar yields produced after hydrolysis to reveal any relationships between the abundance of these breakdown products and sugar yields (Figure 38).

When pretreating *B. napus* straw at 185 °C, cultivars that produced higher organic acids concentrations (acetic and formic acid) also produced more Glc after saccharification (Figure 38). This result can be interpreted in two ways, either the early hydrolysis of acetylated polysaccharides to acetic acid and formic acid improve the accessibility of cellulases to cellulose or, organic acids released under low pretreatment severities have a significant effect on the pH of the pretreatment liquors, thereby catalysing the hydrolysis of CW components. As the abundance of initial carbohydrate break-down products HMF and 2FA (pKa < 12) (Klasson *et al.*, 2011) do not correlate with hydrolysis yields, but the abundance of formic and acetic acid does (pKa = 3.75 & 4.76 respectively) (Perrin *et al.*, 1981), this provides strong evidence that organic acid released from the biomass catalyse biomatrix opening in a pH-dependent manner.

Interestingly, when straw is pretreated at 210 °C, the amount of organic acids released from a cultivar has no relationship to the sugar yields obtained from the straw (Figure 38). Therefore, when severe pretreatments are used, intraspecific variations in organic acid release and subsequent catalysis of biomass hydrolysis, is less important. This also demonstrates interactions between cultivar selection and process parameters that would not have been detected if a single severity was used.

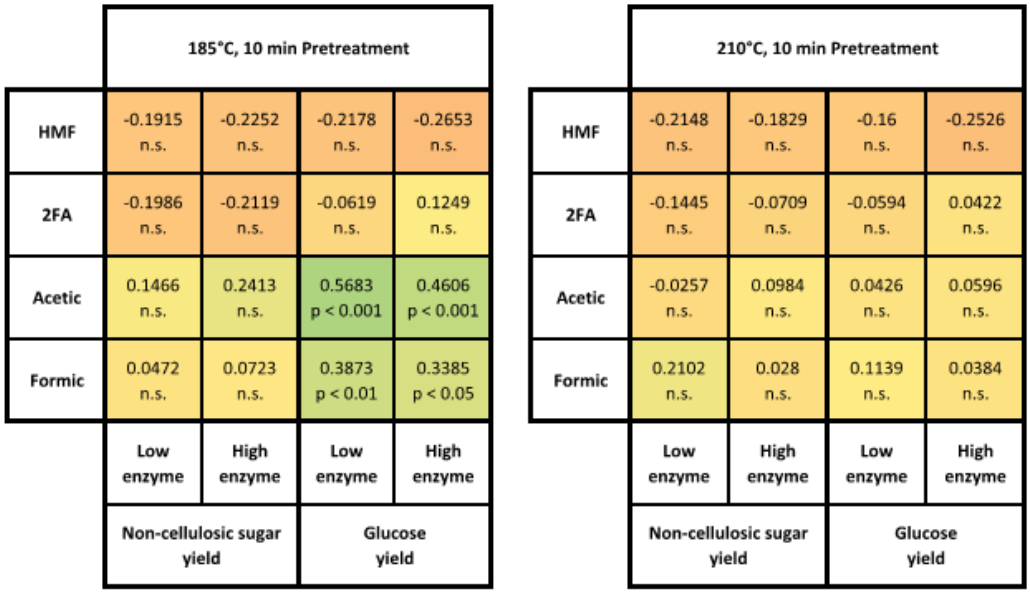


Figure 38 – Correlations between organic acid and sugar yields from *B. napus* straw derived from different cultivars (n = 49) after pretreatment at either 185 °C, or 210 °C for 10 min and hydrolysis at either a low or high cellulase dose (ca. 7 and 36 FPU/g original straw respectively). Cells are coloured according to the direction of the regression line (R = +1 to -1). *P*-values for a two-sided significance test from zero are also displayed.

6.5 Conclusions

The aims of this chapter were to: (i) Tailor existing methods specifically for the HT analysis of biomass. (ii) Establish rational, complementary and contrasting screening conditions that will maximise the interpretation of later screening work. (iii) Evaluate to what extent genotypic variations in traits relevant to bioethanol production are related to process conditions. (iv) Obtain data relevant to 2G bioethanol production suitable for a pilot-scale GWA study.

To analyse biomass in a higher-throughput manner, OSR straw was first finely milled to gain a representative sample, portioned out at a scale large enough to minimise the effect of weighing errors and pretreated in a precisely-controlled heating system. To add flexibility in hydrolysis conditions, the biomass was rinsed, reconstituted in buffer and transferred to multiple 1 mL wells as slurries. This allowed multiple hydrolyses to be conducted using never-dried substrates in a randomised layout. For quantification, a commonly used assay was adapted to a 96-well plate format and tailored specifically for the use on biomass hydrolysates. The resulting DNS assay is significantly faster and more precise than previous methods and eliminates the need for sample dilution. This provided a suitable platform with which to screen multiple straw samples.

Rational screening parameters were established by pretreating OSR straw from a single cultivar at a range of conditions and selecting the most informative conditions. When

processing straw from various cultivars under four different conditions, intraspecific variation was found for all traits (fermentation inhibitor abundances after pretreatment and sugar yields after hydrolysis).

The strength of correlations in Glc yields between processing regimes suggests that a significant proportion of the variation in yields is common to all conditions. Nevertheless, comparing data collected using the disparate screening conditions suggests that some subtleties in Glc yields between cultivars may depend on processing conditions.

Cultivars that produced more acetic and formic acid from their straw also yielded more Glc after pretreatment at an intermediate severity (185 °C), but not under more severe conditions (210 °C). As other breakdown products did not correlate with Glc yields, this suggests that intraspecific differences in endogenous organic acid abundances catalyse biomatrix opening to various degrees. This could be exploited to increase pretreatment efficiency using less severe pretreatment conditions.

By analysing product yields produced from *B. napus* straw, varying both process parameters and straw genotypes, the previous chapters (4-6) provide likely physiological mechanisms for biomass improvement. However, to implement these changes *in planta,* the underlying genetic variations that determine the variation in saccharification yields must be elucidated. Therefore, in the next chapter, trait data collected in this chapter was used to conduct a pilot-scale GWA study, to highlight areas of the *B. napus* transcriptome that correlate most strongly with the variation in these traits relevant to bioethanol production. This should allow greater insight into the genetic determinants of biomass recalcitrance between *B. napus* cultivars.

7 Identifying areas of the *Brassica napus* transcriptome associated with cellulosic ethanol production traits

7.1 Chapter outline

Elucidating the genetic determinants of biomass recalcitrance could improve both our exploitation of biomass and expand our understanding of the natural world. Association mapping has been identified as one strategy suitable to identify these genes (Slavov *et al.*, 2013). However, few studies have collected data suitable for a GWA study. In this chapter, GWA mapping was used to locate likely areas of the *B. napus* transcriptome that correlated most strongly with traits relevant to bioethanol production. The close synteny between *Brassica* and *Arabidopsis* genomes allowed candidate genes in these regions to be suggested, drawing on the detailed functional annotation available in *Arabidopsis*. Although candidates could only be tentatively assigned, based on their proximity to associated SNP markers, the results allowed a series of falsifiable hypotheses regarding the nature of biomass recalcitrance to be established. Observations of this nature could potentially accelerate the identification of genes associated with these complex traits, paving the way for further research in this field.

The positions of SNP markers associated with biomass-related traits suggested that genes involved in sugar nucleotide provisioning, particularly those governing UDP-Xyl synthesis and recycling, were particularly important in determining saccharification yields between cultivars. The molecular machinery involved in the synthesis and endogenous hydrolysis of cellulose, production of glucuronoxylan, pectin modification (particularly pectin methylesterification), xyloglucan synthesis and glycoprotein abundance were all implicated in altering saccharification yields. Few markers were found in close proximity to lignin genes, but those that were are likely to modify syringyl/guaiacyl (S/G) ratios between crop cultivars. Probing the genetic basis of plant CWs also revealed other plausible routes for biomass improvement and may help clarify how CW synthesis is coordinated.

7.2 Introduction

Significant interest has been developed in recent years in the identification of key 'bioethanol genes' controlling plant CW saccharification suitable for biomass improvement (Abramson *et al.*, 2010; Akin, 2008; Bosch & Hazen, 2013; Burton & Fincher, 2014; Carpita, 2012; Farrokhi *et al.*, 2006; Himmel *et al.*, 2007; Pauly & Keegstra, 2008; Phitsuwan *et al.*, 2013; Torres *et al.*, 2013; Yang *et al.*, 2013).

Reductionist techniques are necessary to determine the precise role of specific genes but this is difficult to achieve without guidance from forward-genetic screens. However, progress using reverse genetics is hampered by the complex and dynamic nature of plant CWs, which are very abundant and can adapt to genetic manipulation (Seifert *et al.*, 2010). With nearly 2700 genes involved in CW formation, of which only about 8% are annotated (Wang *et al.*, 2012), complementary hypothesis generation is required to highlight those worthy of closer inspection.

Genome-wide searches for 'bioethanol genes' are likely play a central role in expanding our understanding of CW genetics (Yokoyama & Nishitani, 2004) and could be an excellent way of focussing resources on key genes suitable for biomass improvement (Kintisch, 2008; Slavov *et al.*, 2013; Torres *et al.*, 2013). Moreover, GWA mapping can potentially be used to identify genetic variants suitable for crop breeding (Slavov *et al.*, 2013).

Until recently, GWA studies had only been conducted in a plant species with fully sequenced genomes (Zhu *et al.*, 2008), which restricted association mapping to mainly model species. However, Harper *et al.,* (2012) recently demonstrated that GWA studies can be conducted in *B. napus*, a crop species, with a large, un-sequenced genome using transcriptome sequencing. In a proof-of concept study, Harper *et al.,* (2012) showed that markers derived from transcriptome sequencing associated with seed glucosinolate abundances, surrounding unigenes known to control this trait.

The success of the GWA study conducted by Harper *et al.,* (2012) illustrates that *B. napus* is an excellent species for association mapping for many reasons. Firstly, the transcriptomic diversity found within *Brassica* species is greater than that found in other crop species, as sub-groups have been bred for different purposes (Tang & Lyons, 2012). Secondly, the *B. napus* genome has a relatively high rate of decay in LD (low degree of LD), with LD typically extending ±300–1000 Kb depending on the collections studied (Delourme *et al.*, 2013; Harper *et al.*, 2012). Therefore, associated markers are likely to cluster in close proximity to a potential candidate gene. This perhaps explains why Harper *et al.,* (2012) could detect associations related to a trait which is controlled by few loci, using a comparatively small dataset (n = 53)(Harper *et al.,* 2012).

GWA studies typically require hundreds of samples to be analysed (Korte and Farlow, 2013). This is to increase the confidence in any observed associations to an acceptable level when selected from many thousands of markers. As the material available in this study was limited to 77 accessions collected from 49 cultivars (Chapter 6), the likelihood of detecting individually statistically significant associations using a pilot-scale GWA study was low. Nevertheless, the patterns of association seen might be

biologically informative and aid in the identification of falsifiable hypotheses related to biomass recalcitrance.

Unlike other studies, the close synteny between the *Arabidopsis* and *Brassica* genomes allows 'candidate genes' found in close proximity to associated SNP markers to be identified with greater accuracy; Potentially pinpointing rational genes involved in the variation in that trait (Zhu *et al.*, 2008). *Arabidopsis* has the most comprehensively annotated plant genome available (TAIR, 2014) and is therefore an excellent plant to assign potential functions to *B. napus* unigenes. Therefore, *B. napus* offers an excellent opportunity to probe our fundamental understanding of plant CW genetics, particularly in dicotyledonous plants.

Recently, two similar GWA studies applied a single set of screening conditions to collect saccharification data from two monocotyledonous plant species, Sorghum (Wang *et al.*, 2013c) and Maize (Penning *et al.*, 2014). Penning *et al.*, (2014) used liquid hot water pretreatment (180 °C, 17.5 min - R_o = 3.6), while Wang *et al.*, (2014) used acid pretreatment (2% H_2SO_4, 121 °C, 40 min) before hydrolysis at a single cellulase dose. These considerably larger studies (n = 242 and 282 respectively) highlighted 7 and 115 QTL associated with saccharification yields respectively (Penning *et al.*, 2014, Wang *et al.*, 2013c).

In this study four saccharification conditions were used, pretreating substrate in LHW at two severities (185 °C and 210 °C, 10 min - R_o = 3.50 and 4.24), followed by hydrolysis at two cellulase doses. This provides a novel opportunity to explore interactions between screening conditions and associated regions after GWA mapping. Screening parameters were designed to maximise the chance of capturing genotype-specific variants in product yields with biological and industrial relevance.

Specific aims of this chapter were to:

- Locate likely areas of the *B. napus* transcriptome that correlate most strongly with traits relevant to 2G bioethanol production – thereby highlighting potential genetic determinants of CW recalcitrance between cultivars.
- Suggest likely candidate genes using functional annotations in *Arabidopsis.*
- Use the identity of likely candidates to develop testable hypotheses that could reveal genetic determinants of biomass recalcitrance.

To achieve these aims, quantitative traits relevant to cellulosic ethanol production collected in the previous chapter were combined with a transcriptomic SNP dataset collected from the same cultivars. Outputs from a mixed-linear model combining SNP variants with phenotype data, controlling for population structure, were used to select

areas of the *B. napus* transcriptome associated with the intraspecific variation in each trait.

7.3 Materials and methods

7.3.1 Datafiles provided from mixed-linear-models

Quantitative traits relevant to cellulosic ethanol production collected in the previous chapter were used in a GWA study, utilising a transcriptomic dataset developed by Harper *et al.,* (2012). In total, 16 traits associated with 2G bioethanol production, collected from 49 cultivars under varying processing regimes were sent to Dr Andrea Harper (University of York). Dr Harper combined these datasets with complementary transcriptomic SNP data, collected using Illumina sequencing of RNA extracted from the leaves of 21 day old *B. napus* plants of the same cultivars (Harper *et al.*, 2012).

Dr Harper produced a *regression file* containing outputs from all mixed-linear-models, including *p*-values showing the degree of correlation between each SNP variant with each trait, controlling for population structure (e.g. Table 14). This output was produced using TASSEL (Trait Analysis by aSSociation, Evolution and Linkage, V5.0, available at www.maizegenetics.net). A column containing the $-\log_{10}P$ value was added and "#NUM!" values changed to "0" before loading into R.

	Trait	Marker ID	F	p-value	markerR2	log10P
1	185_2FA	JCVI_690:199	17.82	0.000146	0.4504834	3.8356
2	185_2FA	JCVI_12941:853	15.87	0.000297	0.4011598	3.5272
3	185_2FA	JCVI_5795:576	15.37	0.000381	0.3895108	3.4191
4	185_2FA	JCVI_17322:235	15.38	0.00042	0.4231168	3.3768
5	185_2FA	JCVI_17322:312	15.11	0.000447	0.4204259	3.3497
6	185_2FA	JCVI_17322:267	14.58	0.00051	0.3734621	3.2924

Table 14 – Example header rows of the *regression file* provided

The full genome sequence of *B. napus* is not available, however molecular marker maps exploiting the close synteny between *Brassica* and *Arabidopsis* genomes allows the relative position of each marker to be estimated (Parkin *et al.*, 2005; Parkin *et al.*, 2003). A second, *directions file*, containing the likely genome assigned positions for each SNP marker on each of the *B. napus* chromosomes was provided (Table 15). As the SNP markers identified are based on transcribed sequences, the relative positions of each SNP (Pseudo) are the relative positions of each transcriptomic unigene (likely genes based on transcriptome sequences) on each Chr. An additional column "PseudoSpaced" was added to give the relative positions of each SNP across the entire transcriptome, so that the automated peak searches treated those on each Chr separately.

	Marker ID	Chr	Graph	Sort	Pseudo	PseudoSpaced
1	JCVI_971:763	A1	1	1	8303	8303
2	JCVI_971:773	A1	1	2	8303	8303
3	JCVI_21108:829	A1	1	3	9482	9482
4	JCVI_31415:188	A1	1	4	83741	83741
5	JCVI_31415:208	A1	1	5	83741	83741
6	JCVI_41170:189	A1	1	6	114724	114724

Table 15 – Example header rows for the *directions file* provided

A third "*look-up"* file, containing the orthologous position of each SNP marker (Marker ID), in relation to the position on the *Arabidopsis* genome (AGI) was also provided (Table 16).

	Marker ID	AGI
1	JCVI_10398:46	AT1G06290.1
2	JCVI_10398:58	AT1G06290.1
3	JCVI_10398:86	AT1G06290.1
4	JCVI_10398:100	AT1G06290.1
5	JCVI_10398:106	AT1G06290.1
6	JCVI_10398:175	AT1G06290.1

Table 16 – Example header rows for the *look-up file* provided

Finally, an *effects file*, was provided showing the identity (Allele), estimated effect size (Effect) in % untreated material, for each SNP allele at each locus, relative to a randomly assigned allele (Table 17). The number of cultivars that had the increasing or decreasing allele at a particular locus was also included (Obs).

	Trait	**Marker**	**Allele**	**Effect**	**Obs**
1	210_2FA	AM057018:293	R	0.004510013	37
2	210_2FA	AM057018:293	A	0	1
3	210_2FA	AM059296:66	C	0.00208085	12
4	210_2FA	AM059296:66	M	0	26
5	210_2FA	AM060041:149	Y	0.004151017	17
6	210_2FA	AM060041:149	C	0	21

Table 17 – Example header rows for the *effects file* showing the quantitative effect of each allele on product yield

7.3.2 **Automated peak detection and identifying candidate genes**

The *regression file* was split so each trait could be analysed separately using Kutools for excel (Detong Technology LTD). All further data handling was conducted R version 3.0.2. The *regression file* and *directions file* were merged based on "Marker ID". The resulting dataset was sorted by "*p-value*" in ascending order. All rows containing SNPs with a *p*-value < 0.01 were extracted.

In some cases, multiple SNP markers found above this threshold were located in the same unigene (for example: JCVI_17322:267, 235 and 312). To allow peaks to be detected based on their relative distances from adjacent high probability SNPs the dataset was first sorted by order on pseudo-molecule (PseudoSpaced) and lower *p*-value duplicates found in the same unigene were removed.

The relative distance between SNPs on the pseudo-molecule (in base pairs) is given in column "*PseudoSpaced*". Therefore, each high-probability-SNP was sorted by their relative positions on the pseudomolecule and distances calculated between each SNP and its neighbouring high-*p* SNP. A search distance was set at 1700 Kb. Each SNP was assigned a label describing its position in a peak – either:

(I) "LONE", > 1700 Kb from another marker,
(II) "START", upstream SNP > 1700 Kb, downstream SNP < 1700 Kb
(III) "END", upstream SNP < 1700 Kb, downstream SNP > 1700 Kb
(IV) "IN PEAK" if flanked by SNPs < 1700 Kb apart

The resulting SNP markers were annotated with their corresponding positions on the *Arabidopsis* pseudomolecule, merging with the *look-up file* based on "Marker ID". The results were sorted by position on the pseudomolecule and required columns extracted. Areas of the *B. napus* pseudomolecule surrounding these highly associated SNPs were searched, and likely candidate genes assigned, using functional annotation of *Arabidopsis* gene orthologues using "The *Arabidopsis* Information Resource" (TAIR 10) genome browser (TAIR, 2014).

7.4 Results and discussion

7.4.1 Association mapping of traits relevant to bioethanol production

7.4.1.1 Extracting most highly associated SNPs for further inspection

GWA studies seek areas of the genome that associate most strongly with the variation in particular trait (Lambert & Black, 2012), thereby revealing potential genetic determinants worthy of further investigation. However, the large number of SNP associations observed across the transcriptome make it necessary to focus on a more manageable number of associated areas.

To achieve this, previous studies have taken one of two approaches: either selecting very strict selection criteria that focus on only a small number (< 10) of highly associated regions (Wang *et al.*, 2013c), or using more liberal selection criteria and drawing key conclusions from coincidental associations in the same regions (Chan *et al.*, 2010, Pasam *et al.*, 2012).

Stringent selection criteria are favoured by some in an attempt to limit potential false discovery rate. However, selecting only very highly associated regions does not necessarily safeguard the results from potential false positives associated with confounding variables (Lambert & Black, 2012), particularly if the associations have low *p*-values. Instead, strict criteria highlight few variants that explain a very small proportion of the overall phenotypic variation (Pasam *et al.*, 2012). This is particularly true for CW-related traits where the phenotype is unlikely to be determined by single loci. Instead, it is likely that final yield is determined by the concerted efforts of multiple genes that make a small but significant contribution to the overall phenotype.

In this pilot-scale GWA study, *p*-values associated with saccharification traits were relatively low considering the number of statistical tests involved. This was perhaps inevitable as relatively few accessions were analysed in this study. Although some studies conducted in *Arabidopsis* and *B. napus* have located SNP markers related to simple traits, controlled by few loci, using small datasets (fewer than 100 cultivars)(Korte and Farlow, 2013, Harper *et al.*, 2013), associations studies typically require much higher replication to isolate statistically significant associations. To gain sufficient confidence that associated regions are related to the trait in question, higher replication may be needed (200-1000 cultivars)(Korte and Farlow, 2013). Results from this study should therefore be viewed as an initial look at what areas might be associated with traits relevant to bioethanol production, which would be greatly improved by greater replication.

Considering the central purpose of GWA studies is to generate testable hypotheses (Lambert & Black, 2012), and the effort involved in collecting genotypic and phenotypic datasets suitable for GWA studies, it makes sense to extract as much useful information from them as possible. Therefore in this study, a liberal approach was initially adopted to probe the relatively weak associations observed, to see if worthwhile information might be extracted.

Liberal selection criteria accept a higher chance of type I errors caused by confounding factors, however, quality control checks conducted before selection should limit the these errors (Pasam *et al.,* 2012). In this study, transcriptomic data (circa. 100,000 SNPs) underwent multiple quality control checks before interpretation. Particularly rare alleles (bottom 5%) were removed to limit the potential errors in genotyping, and a mixed linear model used to correct for kinship (Harper *et al.,* 2012). After applying these quality control checks, a low false-discover-rate is expected in the resulting dataset (Harper *et al.,* 2012, Pasam *et al.,* 2012). The patterns of association of the ≈13,400 remaining SNPs markers should therefore contain few false positives.

Typically, visual inspection of SNP associations across the genome is used to identify associated areas worthy of further consideration (Figure 39, A). In some cases, clear peaks containing multiple associated markers were observed. For example, one of the clearest association peaks was found for SNP markers associated with 185 Formic on Chr A8 (Figure 39, A). However, selecting association peaks in this way is somewhat subjective. When the decay in linkage disequlibrum (LD) is low, marker density is particularly high, or few regions associate with the trait, association peaks are easier to determine. Moreover, visually resolving multiple association peaks located on the same Chr is difficult. Therefore, automated filtering was used to assign association peaks more rapidly and with less subjectivity (Figure 39, B).

Similar studies set an arbitrary threshold line to extract variants likely to show an association with the trait. Typically, SNPs in the top 0.1 percentile are extracted (in these studies, $-\log_{10}(P) > 1.3$) before exploring these peaks in greater depth (Chan *et al.*, 2010, Pasam *et al.,* 2012). Here, a similar approach was taken extracting SNPs associated with each trait ($-\log10(P) > 2$). This threshold captures many visible association peaks above the perceived background level of random associations (Figure 39).

Ideally, SNPs in LD with a single variant could be grouped and the most highly associated marker in that regions reported. However, in this study, transcriptome sequencing was used, ordering SNP variants by location on a pseudomolecule. The pseudomolecule is our best a representation of transcriptome order using current

knowledge but does not perfectly reflect the physical distances between markers. Therefore, precise estimates of LD cannot be calculated from transcriptome data.

Nevertheless, pseudomolecule positions give an approximation of the distances between SNPs across the genome. The decay in LD in *B. napus* genome varies based on the study but is typically very low, typically extending < 300-1000 Kb, or as high 1700 Kb in some cases (Delourme *et al.*, 2013). It is therefore unlikely that two SNPs located > 1700 Kb apart on the pseudomolecule are unlikely to be in LD, irrespective of the dataset used (Delourme *et al.*, 2013). Consequently, more highly associated SNPs (p < 0.01) located < 1700 Kb away from each other were grouped and the most highly associated SNPs in that region were selected for closer inspection (Figure 39, B).

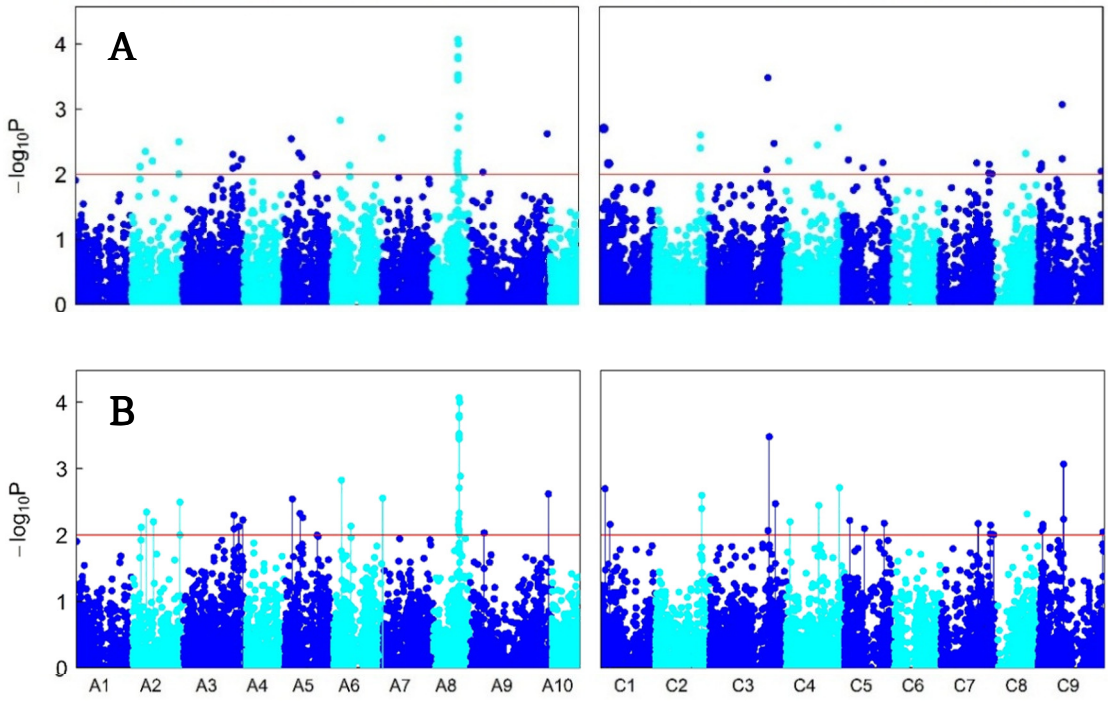


Figure 39 – Example Manhattan plots showing the distribution of genome-assigned SNP markers (● and ●) associated with the variation in formic acid release after pretreatment at 185 °C (y-axis). Markers are ordered along the x-axis according to their position on the *B. napus* pseudomolecule (Chr A1-10 & C1-9 respectively). Most highly associated SNP markers (p < 0.01) were initially selected (A) and those located within ±1700 Kb were grouped and the top marker in that region selected (vertical lines, B).

Information regarding these associated regions, for each of the sixteen traits were tabulated (Appendix Tables 7-22, pg. 239-255) as follows:

(I) **Glc and RS 185L** –Pretreatment at 185°C, 10 min, followed by hydrolysis using ca. 7 FPU/g original material (Appendix Tables 7, 8).

(II) **Glc and RS 185H -** Pretreatment at 185°C, 10 min, followed by hydrolysis using ca. 36 FPU/g original material (Appendix Tables 9, 10).

(III) **Glc and RS 210L** - Pretreatment at 210°C, 10 min, followed by hydrolysis using ca. 7 FPU/g original material (Appendix Tables 11, 12).

(IV) **Glc and RS 210H -** Pretreatment at 210°C, 10 min, followed by hydrolysis using ca. 36 FPU/g original material (Appendix Tables 13, 14).

(V) **Formic 185 and 210 –** Formic acid retained in pretreatment liquors after pretreatment at 185 or 210 °C, 10 min (Appendix Tables 15, 16).

(VI) **Acetic 185 and 210 –** Acetic acid retained in pretreatment liquors after pretreatment at 185 or 210 °C, 10 min (Appendix Tables 17, 18).

(VII) **2FA 185 and 210 –** Furfuraldehyde retained in pretreatment liquors after pretreatment at 185 or 210 °C, 10 min (Appendix Tables 19, 20).

(VIII) **HMF 185 and 210 –** Hydroxymethylfurfural retained in pretreatment liquors after pretreatment at 185 or 210 °C, 10 min (Appendix Tables 21, 22).

In each of these tables, the position of each marker on the *B. napus* pseudomolecule is given (Chr and Pseudo) followed by the degree of association between genotype and phenotype, controlling for population structure (-log10(*p*-value)). The allelic variation at each loci is quoted compared to the modal allele (SNP alleles), with the number of cultivars in the study that contained that allele included in parentheses. The quantitative difference in yield between cultivars with the increasing and decreasing alleles is displayed as the "Effect size". This gives an indication of the degree of change in phenotype expected between those containing either allele.

7.4.1.2 Identifying likely candidate genes associated with traits relevant to bioethanol production

A major advantage of using *Brassica* species for association genetic studies is that they are the most closely related crop species to the model plant *Arabidopsis* (Tang and Lyons, 2012, Ziolkowski *et al.*, 2006). The annotation of *B. napus* unigenes (likely genes based on transcriptome sequencing) surrounding each marker on the *B. napus* pseudomolecule can be achieved by drawing on functional annotations of *Arabidopsis* genes with similar sequence (termed orthologues) (TAIR, 2014). These are tabulated as "Candidate AGI" and "Candidate Name" (Appendix tables 7-22).

Strongest candidates were found when an associated SNP markers were located within a unigene orthologous to a potential candidate gene. For example, JCVI_1973:910 was found within a unigene orthologous to an *Arabidopsis* Xylose isomerase family protein (AT5G57655). Of course, it is possible that these markers are in LD with a neighbouring candidate gene, or coincidentally associated with the gene. For this reason potential candidate genes, however likely, require further validation and in-depth study to ascertain their true role in determining saccharification yields (Wang *et al.*, 2013c).

More commonly, high probability markers were located within a neighbouring region, presumably in LD with a neighbouring candidate gene. A good example of this is the marker JCVI_8788:1017 located on Chr A1, within a unigene (JCVI_8788) orthologous to an Arabidopsis "With No lysine kinase 6" gene (AT3G18750) potentially involved in circadian control of pollen germination (Wang *et al.*, 2008). However, an adjacent unigene (EE490483) ≈31 Kb away from this marker, has a similar sequence to Glucuronic acid substitution of xylan 1 (GUX1) in Arabidopsis (AT3G18660). Reverse genetic studies have recently shown the influence of GUX1 on glucuronoxylan decorations (Bromley *et al.*, 2013, Busse-Wicher *et al.*, 2014) which are very likely to alter the efficiency of both hydrothermal pretreatment and saccharification. One might consider this as a more likely candidate gene.

7.4.2 Areas of the *B. napus* transcriptome associated with saccharification yields across multiple processing conditions

In this study, screening conditions were specifically designed to achieve various degrees of biomatrix opening. Using different process permutations allowed the "you get what you screen for" axiom (Decker *et al.*, 2009), to be investigated in further depth, to include genetic variants resulting from a GWA study.

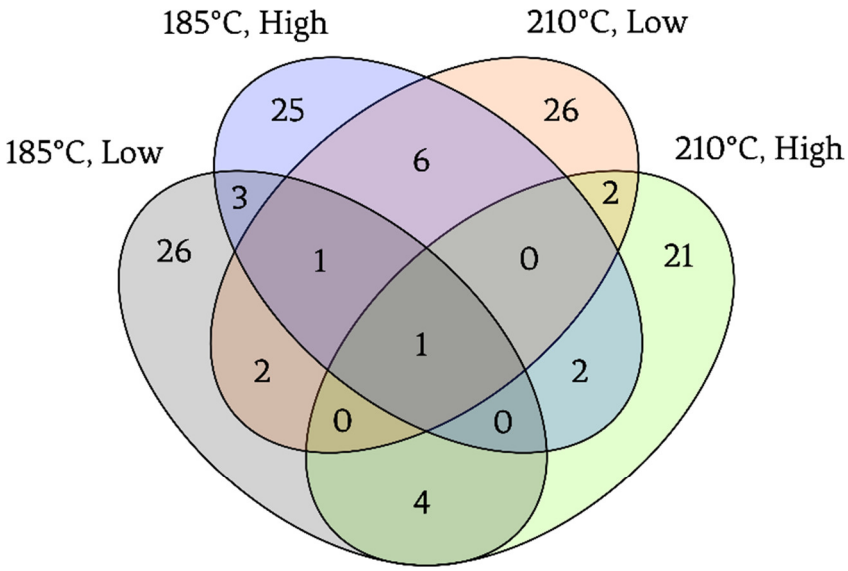
In total, 34 and 21 of the selected markers ($p < 0.01$, highest in a 1700 Kb region) associated with Glc and RS yields respectively, were detected in the same unigene

using more than one screening condition (Figure 40). These markers suggest that at least some areas of the transcriptome associated with the variation in biomass recalcitrance were common to multiple process permutations (Table 18).

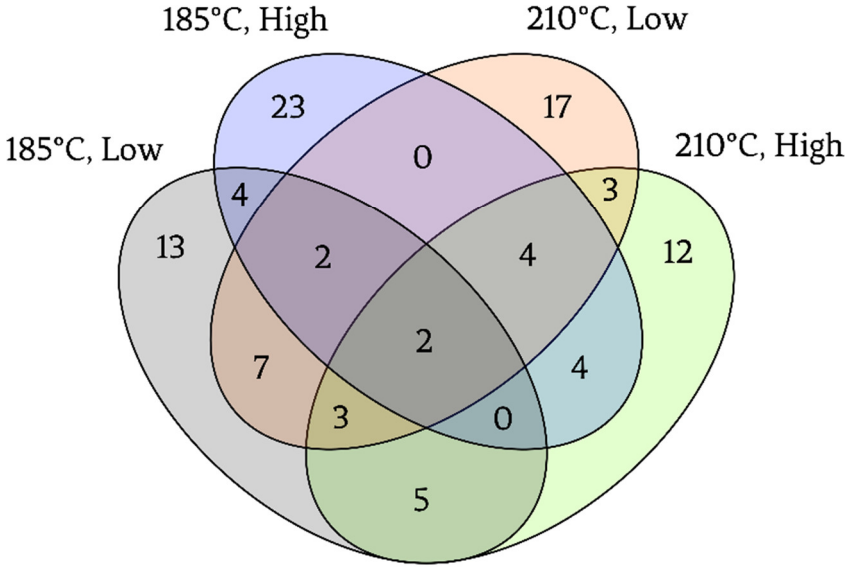
Conversely, some potentially interesting associations appeared to be unique to particular conditions. For example, an area of the *B. napus* transcriptome associated with the range in Glc yields between cultivars within the unigene JCVI_3121 (Chr C7). This marker is interesting as cultivars containing the G allele (n = 25) at this loci tended to produce more Glc than those with the S allele (n = 17) when hydrolysed using a low cellulase dose (Appendix Tables 7, 8, 11 and 12). This trend was seen irrespective of PT condition used. If only a high cellulase dose was used this potentially interesting variant may not have been identified (Table 18). Therefore, in this respect, the "you get what you screen for" axiom appears to apply.

In addition to Glc yields collected from various cultivars, the same liquors were tested for RS concentrations using the DNS assay optimised in Chapter 6. The majority of RS in these hydrolysates will be Glc, particularly after pretreatment at 210°C (See Figure 33, page 134). Therefore, this data gives a second opportunity to detect associations related to Glc yields and isolate variants associated with the release of other, potentially fermentable, sugars (Table 18).

Candidates likely to be involved in determining variations in saccharification yields between cultivars, common to multiple process conditions included: those governing UDP-Xyl/Ara provisioning and recycling (UDP-GlcA decarboxylase 2, XIFP), cellulose properties (CesLA15, RSW4, KOR2), xylan/ pectin modifications (GUX1, GUX5, IRX14 and pectin lyases) and CW acetylation (TBL family members) (Table 18). All these candidates are likely to influence saccharification yields in some way.



Glucose yields
Markers in the same unigene between treatments



Reducing sugar yields
Markers in the same unigene between treatments

Figure 40 – Ven diagrams showing common association peaks between screening conditions. Selected SNP markers associated with each trait ($p < 0.01$, lowest p-value in a 1700 Kb region) that were found in the same unigene were identified.

Unigene	Position		Reducing sugar yields	Glucose yields	Likely Candidate ID	
	Chr	Pseudo			AGI	Name
1 JCVI_2389	C4	25113320	185L+185H+210L+210H	185L+185H+210L+210H	AT3G62830	UDP-Glucuronic acid decarboxylase 2
2 JCVI_430	A8	17643293	185L+185H+210L+210H	185H+210H	AT1G08990	Glucuronic acid substitution of xylan 5
3 JCVI_27767	A7	10349313	185L+185H+210L	185L+185H+210L	AT1G22380	UDP-glucosyl transferase
4 JCVI_8724	A3	30077326	185L+185H+210L	185H+210L	Multiple	Pectin lyase-L superfamily protein (AT4G35670) or F5H 1 (AT4G36220)
5 JCVI_197/ JCVI_1973	A10/ C9	6561639/ 48568519	185L+210L+210H	185H+210L	AT5G57655	Xylose isomerase family protein
6 JCVI_17664	A2/ C5	18739683/ 11041316	185L+210L+210H	185L+185H	AT2G01320	ABC-2 type transporter family protein
7 JCVI_16252	C8	10965209	185H+210L+210H	185H+210L	AT4G14570	Acylamino acid-releasing enzyme
8 JCVI_1056	C3	58549048	185L+210H	185H+210H	AT4G36890	Irregular Xylem 14
9 JCVI_3121	C7	44734046	185L+210L	185L+210L	AT4G22970	Radially swollen 4
10 JCVI_6088	A8	10661406	185L+210L	185L+210L	AT4G30170	Peroxidase family protein
11 JCVI_8788	A1	22635931	185L+210L	185H+210L	AT3G18660	Glucuronic acid substitution of xylan 1
12 JCVI_929	C6	36810175	185H+210H	185H+210H	AT1G65610	Korrigan 2
13 JCVI_18630	C9	54639474	-	185L+185H	AT5G19160	Trichome birefringence-like 11
14 JCVI_32045	A6	22525163	-	185L+185H	AT3G26040	HXXXD-type acyl-transferase family protein
15 JCVI_14508	C8	26011248	-	185L+210H	Multiple	Trichrome Birefringence-like 36 or GH Family 2
16 JCVI_26272	A7	22372439	-	185L+210H	AT1G71220	UDP-glucose:glycoprotein glucosyltransferase.
17 JCVI_26470	C4	33041573	-	185L+210H	AT4G13410	Cellulose-synthase like A15
18 JCVI_7088	A7	12376898	-	185L+210H	AT2G27380	Extensin proline-rich 1
19 JCVI_32230	A7	24389437	-	210L+210H	AT1G76670	Nucleotide-sugar transporter family protein
20 JCVI_16555	A6/ C7	19949020/ 38397626	-	210L+210H	AT5G26810	Pectin lyase-like superfamily protein
21 JCVI_11821	C3	4803363	185L+210L+210H	-	AT5G19160	Trichome birefringence-like 11
22 JCVI_8464	C3	16534486	185H+210L+210H	-	AT2G28760	UDP-Xylose synthase 6
23 JCVI_21919	C3	41744982	185H+210L+210H	-	AT5G61840	Irregular Xylem 10-Like - Glucuronosyltransferase
24 JCVI_14588	A4	6317040	185L+185H	-	AT5G35660	Glycine-rich protein family
25 JCVI_7836	C1	18148122	185L+185H	-	AT4G14360	SAM-dependent methyltransferases superfamily protein
26 JCVI_12842	A1	9738463	185L+185H	-	AT5G58090	O-Glycosyl hydrolases family 17
27 JCVI_15647	A6	9069124	185L+185H	-	Multiple	Cell-wall localised gene / hydroxyproline-rich glycoprotein
28 JCVI_1134	A10	1708228	185L+210H	-	AT1G04910	O-fucosyltransferase family protein
29 JCVI_1336	C4	34725981	185L+210H	-	AT5G37660	Plasmodesmata-located protein 7
30 JCVI_30870	C4	54006138	185L+210H	-	AT2G44110	Mildew resistance locus O 15
31 JCVI_4441	A8	12629468	185L+210H	-	AT4G38080	Hydroxyproline-rich glycoprotein family protein
32 JCVI_27253	A1	12906362	185L+210L	-	Multiple	Carbohydrate-binding protein / Glycosyl hydrolase family 43
33 JCVI_1535	C9	24392590	185L+210L	-	Multiple	Low-level beta-amylase 1 (AT5G47010) or Peroxidase (AT5G47000)
34 JCVI_17961	A9	25881394	185L+210L	-	AT1G28290	Arabinogalactan protein 31
35 JCVI_7695	A3	14117383	185L+210L	-	AT4G01210	Glycosyltransferase family 1 protein
36 JCVI_27578	C4	6392947	185H+210H	-	Multiple	Annexins 3 & 4 / microtubule-associated protein 65-5
37 JCVI_37942	C1	12385276	185H+210H	-	AT4G24530	O-fucosyltransferase family protein
38 JCVI_8475	C7	50748254	185H+210H	-	AT2G20870	Cell wall protein precursor, putative
39 JCVI_1101	C5	24761456	210L+210H	-	Multiple	Granule bound starch synthase 1 / HXXXD-type acyl-transferase
40 JCVI_1124	A7	20231486	210L+210H	-	AT1G65890	Acyl activating enzyme 12
41 JCVI_11801	A7	850820	210L+210H	-	AT2G18800	Xyloglucan endotransglucosylase/hydrolase 21

Table 18 – Areas of the *B. napus* transcriptome that were consistently associated with the intraspecific variance in saccharification yields between 49 *B. napus* cultivars, using one or more processing conditions. Areas were selected by extracting SNPs above a threshold ($p <$ 0.01), locating the highest marker in a 1700 Kb region and observing the coincidence of these markers in the same unigene from multiple saccharification traits.

7.4.3 Mechanisms of biomass recalcitrance: developing falsifiable hypotheses and potential biomass improvement strategies

GWA studies are perhaps the fastest way to generate falsifiable hypotheses, which can focus efforts on key areas of the genome, thereby accelerating gene discovery (Lambert & Black, 2012). The previous section located areas of the genome common to multiple screening conditions, which are likely to harbour candidates associated with saccharification yields under multiple screening conditions. These may be the most interesting genetic candidates to explore in further depth.

However, without knowing the potential mechanism of these genes it will be difficult to identify which candidates are legitimate breeding targets or elucidate the mechanisms of biomass recalcitrance which could expand our fundamental knowledge of CW synthesis. Therefore, in the following sections, likely candidates found in regions associated with the variance in saccharification yields between cultivars are grouped by likely mechanisms. A plethora of potential biomass improvement strategies have been suggested from current knowledge of reverse genetic screens (Abramson *et al.*, 2010, Bosch & Hazen, 2013, Burton & Fincher, 2014, Carpita, 2012, Farrokhi *et al.*, 2006, Himmel *et al.*, 2007, Pauly & Keegstra, 2008, Phitsuwan *et al.*, 2013, Yang *et al.*, 2013). The patterns of association observed here add to these potential strategies, but are developed based on the patterns of association across the transcriptome. Each strategy could be interrogated directly using the potential candidate genes highlighted in this study.

7.4.3.1 Intraspecific differences in UDP-Xyl/UDP-Ara abundance are associated with biomass saccharification

Sugar nucleotides are the building blocks of structural carbohydrates and therefore the relative abundance of these components is likely to alter CW compositions. Although many studies have shown that plants deficient in particular NDP-sugar genes have altered CW compositions, it is not known to what extent CW glycan abundance is regulated at this level (Bar-Peled & O'Neill, 2011).

In this study, multiple SNP markers associated with saccharification yields were located in regions harbouring genes likely to be involved in NDP-sugar synthesis, recycling and movement. This suggests that, at least in mature field-grown *B. napus* plants, the abundance of CW polymers are likely to be influenced by genetically determined differences in NDP-sugar abundances.

For example, one region of the transcriptome within the unigene JCVI_2389 was associated with every saccharification trait collected in this study (Table 18). A unigene

of similar sequence to an Arabidopsis UDP-GlcA decarboxylase 2 (AT3G62830) was located 166-118 Kb away from these markers (Table 18). UDP-GlcA decarboxylase is the final unidirectional control point, providing UDP-Xyl to the UDP-Xyl/Ara pool (Discussed in section 2.6.2, pg. 40) (Bar-Peled & O'Neill, 2011). Therefore, it is likely that this region is associated with saccharification yields by altering the availability of UDP-Xyl or UDP-Ara between cultivars.

Candidate genes of similar functions were also found in close proximity to other associated markers. For example, markers near to a UDP-Xyl synthase 6 (AT2G28760) ortholog were common to all RS yields except for RS 185L (Table 18). Although no study has explored the role of UDP-Xyl synthase 6 directly, other members of the same family are known to control the synthesis of UDP-Xyl for CW polymers (Harper & Bar-Peled, 2002). Therefore, it is likely that this gene encodes a cytosolic protein with similar function to UDP-GlcA decarboxylase (Bar-Peled & O'Neill, 2011).

The patterns of SNP association also concur with those of other association mapping studies (Wang et al., 2011). For example, Wang *et al.,* (2011) identified areas of the Sorghum genome associated with the variation in Glc yields after acid pretreatment and enzymatic hydrolysis, using simple sequence repeat (SSR) markers. One of the most highly associated markers was located 43 Kb from a UDP-GlcA decarboxylase 2 ortholog (Wang et al., 2011). It is therefore likely that this study located the same association region as Wang *et al.,* (2011), albeit in different species and by different methods. Together, these candidates suggest that provisioning of UDP-Xyl to the UDP-Xyl/UDP-Ara pool is likely to play an important role in determining biomass recalcitrance *in planta.*

7.4.3.2 GWA mapping revealed potentially novel routes for the recycling of D-Xyl within the plant, which could alter biomass recalcitrance in the field.

If the provisioning of NDP-sugars is important in controlling biomass recalcitrance, associated regions may help reveal key pathways and novel routes for NDP-sugar synthesis and recycling. For example, the majority of NDP-sugar synthesis and recycling pathways have already been elucidated (Bar-Peled & O'Neill, 2011). However, the pathway responsible for recycling D-Xyl is not currently known (Geserick & Tenhaken, 2013).

It is likely that some D-Xyl is phosphorylated to Xyl-1-P by a yet-to-be discovered sugar kinase, before conversion back to UDP-Xyl by a non-specific UDP-sugar pyrophosphorylase (Geserick & Tenhaken, 2013). This route would be similar to those discovered for other CW sugars. For example, sugar-kinases phosphorylating L-Ara, L-

Fuc, GalA have all been found (Dolezal & Cobbett, 1991, Kotake *et al.*, 2008, Yang *et al.*, 2009), but a potential Xyl-kinase remains elusive (Geserick & Tenhaken, 2013).

Here, two sugar kinases, orthologous to AT3G54690 and AT2G21370, were associated with Glc 185L and RS 185H respectively (Figure 41). The latter gene, has a sequence similar to Xylulose kinase (XK), is expressed in every *Arabidopsis* plant tissue, but does not have XK activity (Hemmerlin, *et al.,* 2006). Therefore, it is possible that AT2G21370 is the yet-to-be discovered gene involved in the recycling of D-Xyl to UDP-Xyl-1-P *in planta*. It would be interesting to see if this genes is involved in recycling D-Xyl in plant CWs, by the phosphorylation of D-Xyl to Xyl-1-P (Figure 41). Mutants deficient in these genes would be expected to accumulate extracellular D-Xyl, thereby testing this hypothesis.

Similarly, many SNP markers were found within (JCVI_1973:910 and 946) and surrounding (JCVI_197:367, JCVI_16118:186, JCVI_11776:469) a unigene orthologous to a Putative xylose isomerase family protein (XIFP - AT5G57655). These markers are associated with almost all saccharification traits apart from 185L (Table 18). Xylose isomerases catalyse the interconversion of D-Xyl and D-Xylulose in a range of species (Jackson & Nicolson, 2002). Bacteria and yeast use Xylose Isomerase to convert D-Xyl into F6P, which can then be used for fermentation (Jackson & Nicolson, 2002). Similarly, plants might use a similar route to recycle free D-Xyl reclaimed from CW polymers and return it to F6P, suitable for the production of new NDP-sugars (Figure 41).

If D-Xyl is recycled via the described routes (Figure 41), the fate of D-Xyl liberated from the CW might be determined by these key genes – either converting it directly back to UDP-Xyl, or into the synthesis of other CW polymers. Considering the potential importance of XIFP (AT5G57655) and UDP-GlcA decarboxylase 2 (AT3G62830) these were selected for further study in *Arabidopsis* (Chapter 8).

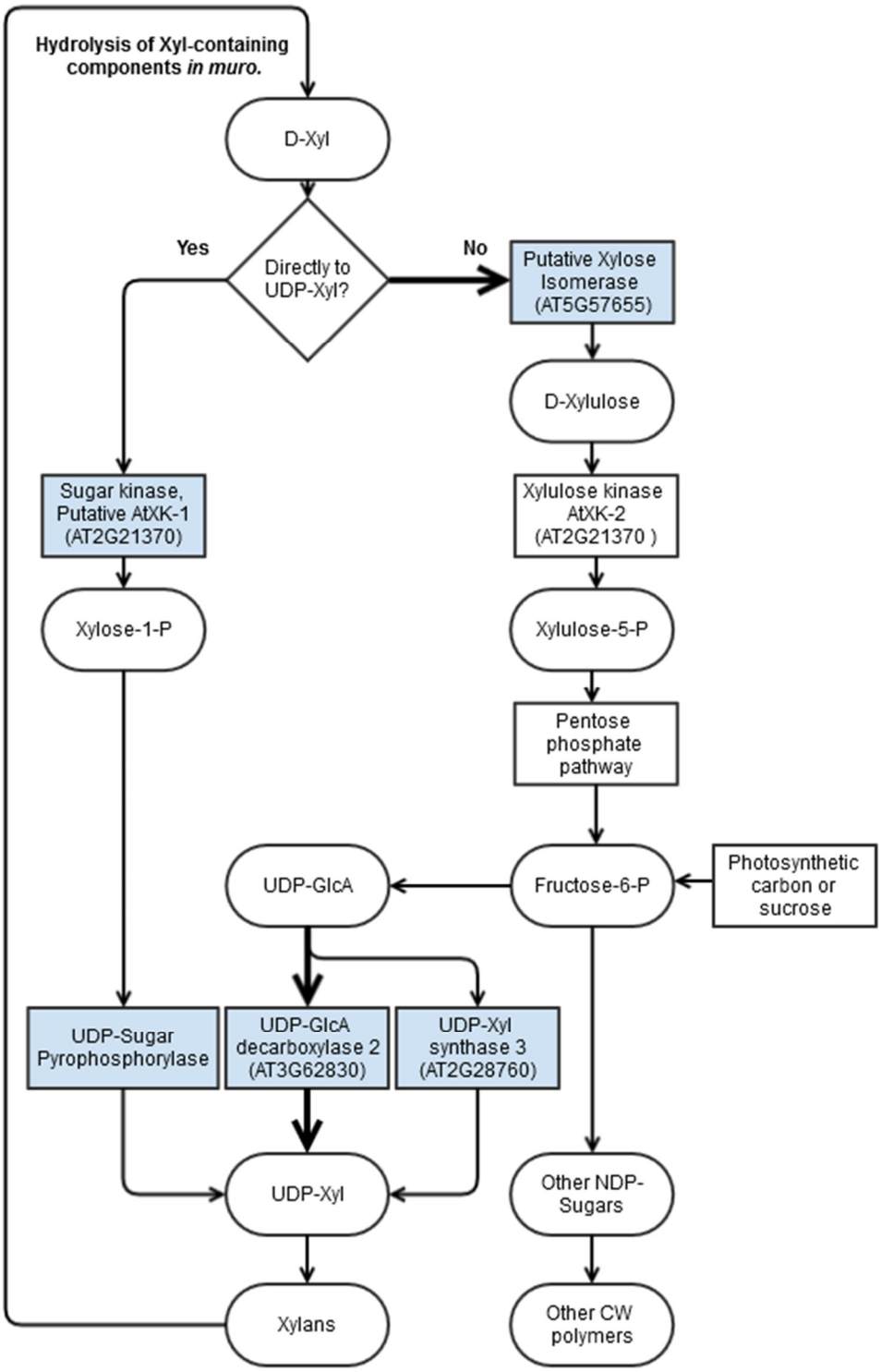


Figure 41 – Schematic describing the predicted routes for D-Xyl synthesis and recycling related to biomass recalcitrance, based on candidate genes revealed in this GWA study. Potential candidates were selected based on areas of the genome associated with saccharification yields after various processing regimes (blue), which are likely to alter biomass recalcitrance. Bold arrows indicate likely dominant pathways. Pathways are consistent with current knowledge (Jackson and Nicolson, 2002, Bar-Peled and O'Neil, 2011, Geserick and Tenhanken, 2013 and Hemmerlin *et al.,* 2006), and provides a framework for further investigation.

7.4.3.3 Plant glycan synthesis and biomass recalcitrance can be influenced by the partitioning of sugar nucleotides between membranes

Multiple candidates genes associated with saccharification yields were identified that may control the movement of sugar nucleotides within the cell. These included: nucleotide-sugar/diphospho-sugar transporter family proteins (AT1G76670, AT4G35335, AT1G57980, AT4G32272 and AT2G02061), sugar porter family proteins 1 & 2 (AT5G27350-60), Polyol/monosaccharide transporter 5 (AT3G18830), and tonoplast monosaccharide transporter 1 (AT1G20840).

Few nucleotide sugar transporters have been categorised to-date (Rautengarten *et al.*, 2014), therefore only one of these candidates (AT1G76670) has been categorised in detail. AT1G76670 (UDP-Gal transporter 2) specifically transports UDP-Gal across plant cell membranes (Bakker *et al.*, 2005).

This observation is interesting for two reasons. Firstly, this observation concurs with pilot-scale observations suggested that the variance in Gal containing polymers in the original material correlated most strongly with the variations in Glc yield between cultivars released after SE and enzymatic hydrolysis (Chapter 5). The appearance of Gal-specific candidate genes using different materials and using different processing regimes, adds greater confidence that these observations are representative of autocatalytic pretreatment of *B. napus* straw.

Secondly, UDP-Gal is produced by the bi-directional epimerisation of UDP-Glc by UDP-Gal 4-epimerase (GALE) (Seifert *et al.*, 2002). Therefore, if the abundance of Gal-containing polymers is controlled at the level of sugar nucleotides, the unidirectional transport of UDP-Gal across membranes is the most likely bottleneck for the availability of Gal for carbohydrate synthesis. It would be interesting to see the functions of the other sugar transporters highlighted in this study and the effect they have on plant glycan synthesis.

7.4.3.4 Endogenous glycosyl hydrolases alter cell wall recalcitrance by hydrolysing cellulose *in planta*.

Multiple candidate genes, located in regions associated with saccharification yields, were members of the CaZy GH family 17 (glucan endo-1,4-β-glucosidases: AT1G11820, AT2G01630, AT4G34480, AT5G56590, AT5G58090, AT5G64790 and KOR2 - AT1G65610) which specifically hydrolyse cellulose chains (Lombard *et al.*, 2014). These enzymes are likely to work in unison with other candidate genes such as βGs (AT1G72990, AT1G75940, AT1G66270-80, AT3G24180, AT3G09260, AT4G38590 and AT5G25980) to hydrolyse cellulose *in planta.* The presence of these fourteen regions in

areas of the transcriptome harbouring genes encoding endogenous cellulases suggest that they are important determinants of CW recalcitrance.

Identification of these genes by reverse-genetics would probably be impossible, as their families are so large. For example, *Arabidopsis* has at least 48 and 51 members of the GH families 1 and 15 respectively (Lombard *et al.*, 2014; Xu *et al.*, 2004). Creating and analysing mutant plants deficient in all 40-50 family members would be exceptionally difficult. However, genes highlighted here could be excellent targets for further categorisation of this family. This illustrates a strength of GWA mapping in developing hypotheses related to CW functioning.

The number of Endo β-1,4-glucanases found in associated areas of the transcriptome may reflect compositional differences specifically relevant to *B. napus* straw. A large proportion of the above ground biomass collected was empty pod tissue (Malagoli *et al.*, 2005, Rossato *et al.*, 2001). As the plant matures, compounds held in other tissues (particularly the pod walls) are mobilised for seed filling. As *B. napus* pods mature, they also release Endo β-1,4-glucanases to hydrolyse cells in the dehiscence zone to initiate seed dispersal (Meakin and Roberts, 1990). The actions of these genes are likely to alter the suitability of the pod material for hydrolysis.

Specific endo-1,4-β-D-glucanases such as KOR2 are likely to alter cellulose deposition particularly in vascular tissues (Mølhøj *et al.*, 2001). As other KOR family members are known to alter cellulose crystallinity (Nicol *et al*, 1998), KOR2 could fulfil a similar function which has implications for biomass hydrolysis. It is therefore likely that the saccharification of stem and pod tissues between cultivars was caused by changes in cellulose crystallinity afforded by various endogenous Endo β-1,4-glucanases.

7.4.3.5 Genes involved in cellulose synthesis are likely to alter recalcitrance: CesA, CESL, KOR, COBL, RSW, and TBL family members

Many areas associated with saccharification yields between cultivars were located in orthologous regions to genes involved in cellulose synthesis. For example: CesA8 (AT4G1878), CesA-L G1, 2 and 3 (AT4G23990-4010), COBRA-like (COBL) protein 5 and 11 precursors (AT5G60950 and AT4G27110), Radially Swollen 3 and 4 (AT5G63840 and AT4G22970) and Radial swelling 9 (AT5G42080). All of these candidates are exceptionally interesting in terms of their role in cellulose deposition and are likely to have a direct impact on cellulose properties, including resistance to hydrolysis.

In *Arabidopsis,* CesA8 is involved specifically in the synthesis of cellulose in secondary CWs (McFarlane *et al.*, 2014). COBL 5 and 11 protein precursor have not been studied in

depth but are likely to be involved in cellulose microfibril modification (Roudier *et al.,* 2002). Radially Swollen 3 is an integral determinant of cellulose synthesis (Burn *et al.,* 2002b) while Radially Swollen 4 is more likely to influence non-cellulosic polysaccharide compositions (Wiedemeier *et al.,* 2002). Radial swelling 9 is also directly involved in cellulose deposition in some way, as mutant plants produce considerably less cellulose compared to WT plants (Collings *et al.,* 2008).

Other particularly abundant candidates that may alter cellulose properties through changes in non-cellulosic polysaccharide compositions included TBL proteins. Candidates found in areas of the transcriptome associated with saccharification yields included: TBL3 (AT1G73140), TBL11 (AT5G19160), TBL32 (AT3G11030), TBL39 (AT2G42570), TBL44 (AT5G58600), TBL36 (AT3G54260), TBL/Powdery-mildew resistant (AT5G58600), Mildew resistance locus O13 (AT4G24250), and Mildew resistance locus O15 (AT5G44110) orthologues. Interestingly, Penning *et al.,* (2014) recently suggested a powdery mildew-resistant protein as a strong candidate related to variation in saccharification yields between maize cultivars. Together these suggest a role of TBL genes in determining biomass recalcitrance.

All of the cultivars used in this study, and that of Penning *et al.,* (2014), were grown under field conditions and therefore might have been exposed to pathogens. Consequently, TBL/mildew resistant genes could directly influence Glc yields through altered CW chemistry (Vogel *et al.,* 2004), or by the fungus directly appropriating Glc from infected tissues (Sutton *et al.,* 1999). If this case, breeding crops for improved yields for bioethanol production and disease resistance may have common targets. Moreover, the coincidence of similar candidates between maize and *B. napus* suggests that despite their different CW structures, some candidates that are important for saccharification of dicotyledonous plants, may also be important for monocots.

Other GTs involved in Glc-related modifications included: Glucosyltransferase F-1 (AT4G01210), UDP-Glucosyltransferase 85A1, 2, 3 (AT1G22360-400) and UDP-Glycosyltransferase superfamily proteins (AT4G09500). It is not known what roles these genes play in CW dynamics, however it is likely that these candidates are important in CW functioning. A particularly interesting candidate are UDP-Glucosyltransferase 85A1-3 as this association is seen in all saccharification traits apart from the most severe conditions (210H) (Table 18). Further work would be required to ascertain their potential roles.

7.4.3.6 Glucuronoxylan production influences biomass recalcitrance

Glucuronoxylan is the main non-cellulosic polysaccharide in dicot SCWs (Scheller & Ulvskov, 2010). Therefore, it was not surprising that more highly associated SNPs are located in orthologous areas of the *Arabidopsis* genome responsible for the synthesis of these polymers. Examples included Glucuronosyltransferase orthologues: Glucuronic acid substitution of xylan (GUX) 1 (AT3G18660) and 5 (AT1G08990), IRX10 (AT1G27440), IRX14 (AT4G36890), IRX10-Like (AT5G61840). Orthologues of IRX10-L and IRX9 (AT2G37090) were also located in regions associated with RS yields using particular conditions. All these genes are already known to be critically important in glucuronoxylan assembly (Doering *et al.*, 2012; Rennie & Scheller, 2014).

Reverse genetic studies have shown that IRX14 creates xylan backbones (Scheller & Ulvskov, 2010). GUX1 adds evenly spaced [Me]GlcA residues to Glucuronoxylan to every 8-10 Xyl residue (Bromley *et al.*, 2013). The backbone of Glucuronoxylan forms a left-handed helix, making a full rotation every three Xyl-residues (Gorshkova *et al.*, 2013). Therefore, xylans decorated by GUX1 are likely to contain GlcA residues facing in opposite directions on the helix, preventing alignment with cellulose on a single side. Changes in GlcA decorations of xylans are also likely to impact other side-groups such as acetyl-groups the hydration properties of key these CW polymers (Busse-Wicher *et al.*, 2014) and so the actions of this gene are likely to alter the efficiency of both hydrothermal pretreatment and saccharification properties.

7.4.3.7 Endogenous degradation and cross-linking of pectins influence saccharification yields.

Two further galacturonosyltransferase orthologues were found in close proximity to highly associated SNP markers related to Glc yields after processing (210L) - Quasimodo 1 (QUA1 - AT3G25140) and Quasimodo 2-Like2 (QUL2 - AT2G03480). QUA genes produce pectic HG. Alterations to these genes alter the chemistry and the porosity of the CW (Rondeau-Mouro *et al.*, 2008) and therefore likely to alter the accessibility of cellulases to cellulose (Adani *et al.*, 2011).

Interestingly, QUL2 functions redundantly with QUA1 (Fuentes *et al.*, 2010), which perhaps explains the coincidence of marker associations in these separate areas of the transcriptome for the same trait. The presence of these candidates complements earlier observations after pilot-scale processing, where UA abundance and properties of the pectic matrix were shown to be related to saccharification efficiency. The most likely mechanism for this is limiting the binding of cellulases to CW components (Chapter 4). It is therefore interesting that these associations were only observed under conditions limited by cellulase dose, but not pretreatment (210L).

The abundance of potentially pectin-degrading candidate genes in areas of the transcriptome to saccharification traits, suggests that differences in pectin restructuring between cultivars may alter biomass recalcitrance. Examples included: polygalacturonase 3 (AT1G23760), pectin-lyase-like superfamily proteins (AT4G35670, AT4G33440, AT3G59850, AT5G26810, AT2G23900, AT3G17100). Other pectin-related genes orthologues found in regions associated with saccharification yields included those altering pectin methylesterification: Pectin methylesterase inhibitors (AT3G17130-52, AT5G46940-90) and Methyl esterase 18 (AT5G58310) putative methyltransferase (AT1G63855) and SAM-dependent methyltransferases superfamily proteins (AT1G55450, AT1G78140, AT2G45750, AT3G15530, AT4G18030, AT4G33110-20, AT5G63100 and potentially AT5G63100). These genes could alter the degree of methyl-esterification of pectins, thereby altering saccharification properties (Francocci *et al.*, 2013; Lionetti *et al.*, 2010). This hypothesis is consistent with previous findings at a pilot-scale (Chapters 3 and 4), which showed that spectral differences between cultivars associated with HG abundance were related to saccharification performance.

Further to these candidates, interesting but more tentative candidates included the presence of Annexins 3 and 4 (AT2G38750-60) near to an orthologous region associated with saccharification yields. Interestingly, a recent GWA study conducted in maize also suggested an gene with a similar sequence to Annexin 8 was associated with Glc yields between cultivars pretreated at a high severity (R_o≈3.60) and hydrolysed using a high dose of Ctec2 (ca. > 50 FPU/g) (Penning *et al.*, 2014). Annexins sense high concentrations of Ca^{2+} in the cytoplasm and migrate to the cell membrane (Mortimer *et al.*, 2008). Therefore, they are likely to be essential for Ca^{2+} homeostasis. This could be important as HG cross-link via Ca^{2+}-mediated ionic bonds between demethyl-esterified groups (Xiao & Anderson, 2013). Annexin 3 and 4 are the only *Arabidopsis* proteins thought to be related to the actin cytoskeleton (Clark *et al.*, 2001), therefore they are more likely than most Annexins to be involved in CW deposition (Wightman & Turner, 2008). Together, the results presented in this study and Penning *et al.* (2014) suggest that Annexins, and Ca^{2+} provisioning in general, could be important determinants of biomass recalcitrance and warrant further study.

7.4.3.8 Role of cell wall structural proteins - GRPs, HPRPs, AGPs and lectins in altering biomass recalcitrance.

A number of markers showing some association with Glc yields under most conditions (except for 185L) were located in regions near to structural protein coding regions. Example included GRP (AT5G35660, AT1G04800, and potentially AT5G46730), PRP (AT2G27380, AT5G43770, AT5G19810 and potentially AT3G18810), HPRP (AT1G21680), an α-Expansin Gene family protein (AT3G15370), a CW protein

precursor (AT2G20870) and FASCICLIN-like AGP family 16 precursor (AT2G35860). Further candidates related specifically to RS yields included an additional HPRP (AT4G38080), a Fasciclin-like AGP (AT5G26730) and AGP31 (AT1G28290).

Interestingly, Wang *et al.,* (2013) also detected an associated SNP variant close to a similar gene - FASCICLIN-like AGP7 - when observing the patterns of association related to saccharification yields of sorghum stems (Wang *et al.*, 2013c). AGPs are particularly interesting as they are a class of chemically and thermally resistant, highly branched, CW associated glycoproteins (Seifert & Roberts, 2007), some of which are known to covalently link CW polymers (Tan *et al.*, 2013).

A particularly remarkable candidate is AGP31. Initially this candidate gene was identified in a regions associated with the variation in RS yields between cultivars pretreated at either severity, but hydrolysed at a low cellulase dose (Table 18, pg. 159). This candidate was initially selected based on its proximity to an associated region alone. But very recent work has independently demonstrated that the orthologous *Arabidopsis* AGP31, forms non-covalent cross-links between galactans, methlyesterified polygalacturonan, RGI and itself, thereby strengthening the CW (Hijazi *et al.*, 2014).

We also identified areas of the transcriptome adjacent to Mannose-binding lectin superfamily proteins (AT5G35940-950 and AT2G43740). These proteins are likely to be involved in the synthesis of complex N-glycans (Strasser *et al.*, 2006) which play important roles in plant defence or coordinating CW synthesis (Peumans & Vandamme, 1995; Vandenborre *et al.*, 2011). It would be interesting to see what influence these candidates, and complex glycans in general, have on determining CW synthesis in plants.

7.4.3.9 Xyloglucan branching likely to alter saccharification yields

Although agricultural residues contain predominantly SCW material, *B. napus* stems contain varying amounts of parenchyma tissue, which has not undergone secondary thickening. These tissues contain more easily hydrolysable PCWs (Hansen *et al.*, 2013) rich in XG (Scheller & Ulvskov, 2010).

A number of genes likely to be involved in XG modification (Rose *et al.*, 2002) were found in orthologous regions associated with saccharification yields. For example: XG endotransglucosylase/hydrolase 26 (AT4G28850) is associated with Glc 210L. Other genes involved in the decoration of side-groups of non-cellulosic polysaccharides (either XG or Glucuronoxylan): O-fucosyltransferase family proteins (AT2G01480, AT1G04910, AT4G24530) and Galactosyltransferase family protein (AT1G53290), were

also present. These are likely to change the structural characteristics of XG or mannan polymers, therefore altering their resistance to degradation. Interestingly Wang *et al.,* (2011), also suggested a XG fucosyltransferase as a potential candidate gene, located 149 Kb from an SSR marker associated with saccharification yields from sorghum stems. This suggests that genetically determined variations in the abundance of terminal fucosyl-residues may play an important role in determining saccharification efficiency across species.

Additional XG endotransglucosylase/hydrolase orthologues 21 (AT2G18800), 27 (AT2G01850) and 29 (AT4G18990) orthologs were also associated with the variation in RS yields for most conditions (185H, 210L, 210H). The increased number of candidate XG endotransglucosylase/hydrolases genes in relation to RS yields, compared to the one candidate associated with Glc yield, suggests that the variation in Xyl released from XGs may be influenced by these genes.

7.4.3.10 Biomass acetylation alters sugar yields and candidates may expose novel genes involved in this process.

We do not fully understand how CW acetylation is achieved *in planta* (Gille & Pauly, 2012b). It is likely that RWA and TBL work in unison to acetylate plant CW carbohydrates (Gille & Pauly, 2012b). As mentioned previously, six TBL orthologues were found in regions associated with saccharification traits (TBL3, 11, 32, 36, 39 and 44). Moreover, markers found in close proximity to TBL11 and TBL36 were associated with saccharification yields after processing under multiple conditions. This, along with the observation that acetic acid concentration in the pretreatment liquors correlated with Glc yields (Figure 38, pg. 144), shows that CW acetylation may be a promising target for biomass improvement.

Although few of the TBL proteins have been categorised to-date, one candidate that associated with saccharification yields, TBL 29, has already been targeted as a gene likely to alter biomass recalcitrance (Xiong *et al.*, 2013). This gene is likely to alter acetylation of xylans and mannans (Xiong *et al.*, 2013). Genes altering CW acetylation could also alter plant growth and development (Ling *et al.*, 2006). Recent work has shown a gene altering CW acetylation "Gene Y" can alter the lignification and recalcitrance of poplar wood (Chen *et al.*, 2014). The candidate genes identified here could fulfil similar roles to "Gene Y", altering biomass recalcitrance. It would therefore be interesting to see the effect that other members of the TBL family have on biomass recalcitrance by reverse genetics.

The identity of acetyl-donor(s) used to decorate CW polymers is also currently unknown, but suspected to be Acetyl-CoA (Busse-Wicher *et al.*, 2014; Gille & Pauly,

2012b). Here many markers associated with the variation in saccharification yields between cultivars were observed in orthologous regions to potential acetyl-donors/transferases. These included: GDSL-like Lipase/Acylhydrolase superfamily proteins (AT1G71691, AT3G14220, AT3G09930, AT4G28780 and AT5G08460), an Acylamino acid-releasing enzyme (AT4G14570), Acyl activating enzyme 12 (AT1G65890), an Acetylated interacting protein 1 (AT3G09980), and Phosphatidylinositol N-acetylglucosaminyl-transferase (AT2G39445). Although these associations might be individually weak, together they suggest that a number of acetyl carriers might be used for biomass acetylation. Moreover, 'Acylamino acid-releasing enzyme' and 'Acyl activating enzyme 12' could be involved in the suspected transfer of Acetyl-groups from carriers hypothesised by Gille and Pauly (2012b). Further study of these genes may provide a more detailed picture of how CW acetylation is coordinated.

7.4.3.11 Lignification plays a minor role in biomass recalcitrance between cultivars, but S/G lignin ratio is likely to be important.

Despite the strong emphasis in the literature that lignin genes are vital determinants of biomass recalcitrance, this study only highlighted seven markers in positions near to those involved in lignin deposition. The relative paucity of lignin genes concurs with the results from a recent GWA study conducted in Maize which showed that genomic regions associated with lignin abundance did not coincide with those related to saccharification yields (Penning *et al,* 2014). Together, these studies suggest that in both monocot and dicot species, lignin-related genes may play a minor part in determining the variation in saccharification yields between crop cultivars compared to carbohydrate-related genes.

Here orthologues of Ferulate-5-hydrolylase 1 (F5H1, AT4G36220), Cinnamoyl CoA reductase 1 (CCR1, AT1G15950) were identified near to markers associated with saccharification yields. CCR1 synthesises coniferylaldehyde (the nearest precursor to G lignin) and F5H1 converts coniferylaldehyde to a precursor of S-lignin. Plants deficient in CCR1 and F5H1 have significant alterations to the amount of S and G lignin produced (Van Acker *et al.*, 2013). Therefore these associations suggest that S/G-lignin ratios are particularly important in altering biomass recalcitrance between *B. napus* cultivars. This is consistent with reverse genetic studies that have shown that S/G ratio of lignin alters the efficiency of liquid hot water pretreatment (Li *et al.,* 2010) and subtle variations in S/G ratio between *Populus* genotypes can have a significant effect on saccharification yields after dilute acid pretreatment (Davison *et al.,* 2006).

Other interesting candidates included a CW-related ABC-2 type transporter (AT2G01320) (Zhou *et al.*, 2010) and the ubiquitously expressed PRX45 (AT4G30170)

(Valerio *et al.*, 2004) which associated with both Glc and RS yields using various processing regimes (Table 18, pg. 159). Two further markers were present near to peroxidase family proteins PRX 29, 58-59 (AT3G17070, AT5G19880-90). These genes have not been studied in depth, however they could be important determinants of lignin deposition which could alter biomass recalcitrance (Wang *et al.*, 2013a). Although their involvement in biomass recalcitrance is currently only tentative, these candidates warrant further inspection.

7.4.3.12 Altering sucrose synthesis and phloem loading in source tissues could benefit both crop and straw quality.

As sucrose is the main source of carbon for non-photosynthetically active cells, it is likely to play an important role in determining the quantity and quality of CW carbohydrates, particularly in sink tissues. Therefore, the provisioning of sucrose to CW polymers has received much attention (Baroja-Fernández *et al.*, 2012a and b; Barratt *et al.*, 2009, Gerber *et al.*, 2014).

Of particular interest is the partitioning of plant carbon into cellulose – the main repository of CW Glc. Debate centres around how sucrose-derived carbon is channelled into the synthesis of cellulose using either sucrose synthase (SuSy) or invertase (Barratt *et al.*, 2009, Baroja-Fernández *et al.*, 2012a and b). The co-localisation of SuSy with cellulose synthase suggests that sucrose-derived UDP-Glc is channelled directly into producing cellulose (Salnikov *et al.*, 2001). It is debatable whether SuSy is essential for cellulose synthesis (Barratt *et al.*, 2009, Baroja-Fernández *et al.*, 2012a and b, Gerber *et al.*, 2014, Smith *et al.*, 2012), but it is likely that SuSy alters the carbon partitioning of the plant in favour of producing cellulose and other CW carbohydrates (Gerber *et al.*, 2014, Coleman *et al.*, 2009).

Here areas of the transcriptome associated with saccharification yields were located near to a UDP-Glc pyrophoshorylase 3 (UGPase3, AT3G56040), Sucrose-6^{F}-phosphate phosphohydrolase (AT2G35840) and Sucrose transporter 4 (AT1G09960), but not SuSy nor sucrose invertase orthologues. Interestingly, Kleczkowski *et al.*, (2004) suggested that UGPase functions in a similar role to SuSy and argued that UGPase would be particularly in important in source tissues of mature plants. The presence of this associated regions suggests that UGPase could be more crucial to sucrose metabolism and CW deposition than previously thought (Kleczkowski *et al.*, 2004). Sucrose-6^{F}-phosphate phosphohydrolase catalyses the final step of sucrose biosynthesis (Lunn *et al.*, 2000) and Sucrose transporter 4 is involved primarily in phloem loading in source tissues (Weise *et al.*, 2000). Together, these three candidates open up the intriguing possibility that increasing the synthesis and transport of sucrose in *source* tissues could improve CW compositions for bioethanol production.

Reallocating sucrose destined for seed production to produce CW compounds would probably have a negative impact on seed filling, which may not be favourable in a crop species. But targeting sucrose synthesis and transport from source tissues could benefit both crop and co-product. It would be interesting to see how direct modification of these three genes alters CW recalcitrance using reverse genetics.

7.4.3.13 Variation in starch and callose synthesis alter Glc yields after processing

Lignocellulosic biomass generally contains very little starch (≈2-3% w/w). However, concerns over small amounts of starch adding to the variance in Glc yields produced from different cultivars after processing has led some researchers to de-starch their material before screening (Decker *et al.*, 2012). In an industrial setting, straw would not be de-starched before use. Therefore, straw was not de-starched before analysis in this study. Instead, it was decided to hydrolyse whole plant tissues, arguing that genotypic variations in residual starch abundance would only have only a slight influence on Glc yields derived from mature tissues from different cultivars.

Indeed, some regions associated with Glc yields between cultivars were adjacent to unigenes with similar sequences to ADP-Glc pyrophosphorylase (AT1G74910), which catalyses the final stage of starch biosynthesis (Mullerrober *et al.*, 1992), a low-level β-amylase (AT5G47010) and chloroplast localized protein with starch binding activity (AT5G39790). These associations were only detected in datasets collected using less intensive pretreatment conditions, where residual starch is likely to be a greater proportion of the total Glc yield. It is therefore likely that these genes relate to the variance in small quantities of residual starch abundance between cultivars.

Likewise, the release of other, potentially variable sources of non-cellulosic Glc such as callose were not controlled for in this study. Callose is a β-1,3-linked glycan found primarily in the sieve elements of vascular tissues, but is also important in plant wound-response (Delmer *et al.*, 1993). The top four SNPs associated with Glc yield after pretreatment at 185L were located in a unigene orthologous to a Plasmodesmata Callose-Binding gene (AT5G08000). Similarly, orthologues to a Plasmodesmata-located protein 7 (AT5G37660) and two Endo-1,3-β-glucosidases (GH family 17 - AT5G58480, AT5G5890) associated with saccharification yields at 185L and 185H also suggest that a significant portion of the variation in sugar release was derived from callose. Under these conditions, one might expect readily hydrolysable sources of Glc, such as callose, to be a greater portion of the overall Glc yield. Therefore, it is likely that this association is related to the abundance of callose between cultivars and not necessarily cellulose-derived Glc.

7.4.3.14 Possible regulatory genes involved in altering biomass recalcitrance

Regulatory genes have the capacity to not only change the composition of CW structures, but also influence straw composition through altering tissue proportions. Certain tissues are likely to have varying degrees of resistance to saccharification (Hansen *et al.*, 2013; Zhang *et al.*, 2014). Therefore, regulatory genes could be excellent targets for biomass improvement (Demura & Ye, 2010; Hussey *et al.*, 2013). Alternatively, these regulators could have global changes to plant development that hinder total crop yield.

Here, two markers were located in orthologous positions to a MYB 4 (AT4G38620) and MYC4 (AT4G17880) could be interesting candidates. Both the candidates were found in orthologous regions to SNPs associated with hydrolysis yields at intermediate degrees of biomatrix opening (185H or 210L respectively). MYB4 functions as a jasmonate-responsive transcription repressor, downregulating C4H expression (Jin *et al.*, 2000), an important protein involved in phenylpropanoid biosynthesis and a gene known to alter saccharification yields (Van Acker *et al.*, 2013). Jasmonate and ethylene signalling pathways have been implicated as being responsible for altered CW properties important for bioethanol production (Cano-Delgado *et al.*, 2003). However, little is known about the other probable candidate gene orthologous to MYC4. As this was a particularly intriguing candidate, MYC4 was selected for further analysis in *Arabidopsis* (Chapter 8).

Similarly, multiple SNPs (JCVI_690:199, 17322:163 and 7836:240) associated with traits related to non-cellulosic polysaccharide abundances when straw was pretreated at 185 °C (RS185L, RS185H, 185 2FA, 185 Formic) were located in regions surrounding orthologues to CW located Transducin/WD40 repeat-like superfamily protein (AT4G14310). Although the function of this gene is currently unknown, and could be coincidentally associated with these traits, it is possible that this gene is important in CW synthesis and warrants further inspection.

7.4.3.15 High probability markers before genome assignment: rare but interesting genetic variants?

Initial candidate selection, used to identify the most interesting candidates for validation in the following chapter were based on early association models where SNP markers were not allocated to a genome (Appendix Table 23). In a few cases, highly associated markers were not present after genome-assignment. Presumably, these markers were removed as the polymorphisms found at these loci were too rare to pass stringent quality control checks after genome-assignment.

Interesting additional markers observed before genome assignment included those near to unigenes likely to be involved in the synthesis of secondary CW xylans such as IRX8 (AT5G54690). IRX8 is thought to be the main GalA transferase involved in constructing the short reducing end oligosaccharide of xylans. IRX8 *Arabidopsis* mutants are deficient in HG and glucuronoxylan and have a strong influence on CW integrity (Scheller & Ulvskov, 2010).

Genome-assigned saccharification data revealed many potential acetyl-donors and many potential acetyl-transferases (TBL family proteins) associated with Glc yields after processing. The distribution of non-genome assigned SNPs highlighted many more areas of the transcriptome potentially harbouring genes involved in the *O*-acetylation of CW polymers (Gille & Pauly, 2012b). Candidates included RWA4, TBL8 and TBL38 (AT1G29880, AT3G11570 and AT1G29050). Moreover, another candidate gene KLUNKER, PIROGI 121 (AT5G18410), which is known to effect cell shape (Saedler *et al.*, 2004) was found in an associated regions to Glc yields before assignment. In this study, data collected from only 49 *B. napus* genotype were compared. It is likely that collecting information from a larger dataset would considerably improve the confidence of the observed markers and could reveal additional layers of information from rarer genetic variants.

7.4.4 Likely candidate genes associated with fermentation inhibitor yields.

Many of the polymers found in plant CWs are hydrolysed during pretreatment, allowing greater access to cellulose. However, the identity these compounds are only known in general terms – *i.e.* cellulose, hemicellulose and lignin (Jonsson *et al.*, 2013; Palmqvist & Hahn-Hagerdal, 2000). Hexose and pentose sugars thermally decompose to produce HMF, 2FA respectively, followed by further degradation to formic acid. Acetic acid is derived from acetyl groups liberated from mainly non-cellulosic polysaccharides (Jonsson *et al.*, 2013; Palmqvist & Hahn-Hagerdal, 2000). The patterns of genetic association observed here implicate more specific polymer groups responsible for the variation in these compounds between cultivars.

To investigate the likely genetic basis for fermentation inhibitor production, associated markers common to multiple traits were compared (Figure 42) and tabulated as before (Table 19). No high probability markers associated with the variance of 2FA, formic and acetic acid retained in the pretreatment liquors were common to both pretreatment conditions. This suggests that the variation in fermentation inhibitor abundances between cultivars is highly dependent on pretreatment conditions. The likely reason for this is that the majority of compounds that break down to form these inhibitors are derived from fragile or robust components at 185 and 210 °C respectively.

Noteworthy marker associations common to multiple traits included those in the unigene JCVI_33410 (Table 19). This marker was found 6.8 to 7.4 Kb from a unigene with similar sequence to UDP-Glc/UDP-Gal-4-epimerase 1 (AT1G12780) which has weak UDP-Gal → UDP-Glc activity (Barber *et al.*, 2002), and strong UDP-Xyl → UDP-Ara activity (Kotake *et al.*, 2009). This gene is therefore likely to work closely with UDP-GlcA decarboxylase to produce UDP-Ara (Kotake *et al.*, 2009). This adds further evidence that changes to the availability sugar nucleotides, particularly those producing UDP-Ara, could be fundamental determinants of non-cellulosic hydrolysis during pretreatment.

An associated SNP marker common to both 185 and 210 HMF (JCVI_25085:69) was found within a gene with a similar sequence to Cellulose synthase 3 (AT5G05170), involved specifically in the synthesis of PCWs (McFarlane *et al.*, 2014). This suggests that some variation in HMF release between cultivars is produced from cellulose in primary cell wall components. This perhaps explains why, unlike other fermentation inhibitors, more markers were associated with HMF yield after pretreatment at both severities (Table 19).

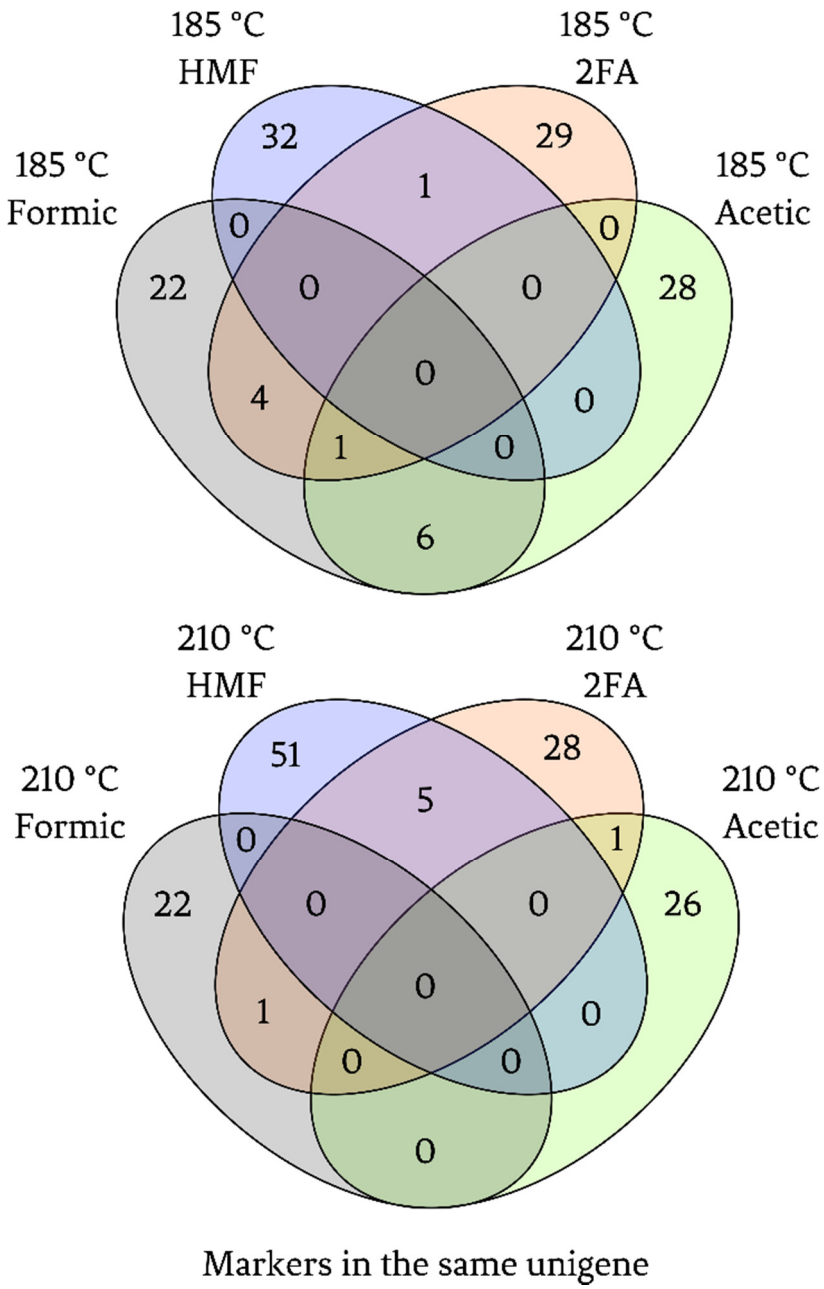


Figure 42 – Ven diagrams showing common peaks associated with the variance in fermentation inhibitors contained within the pretreatment liquors of *B. napus* straw derived from *49* cultivars, pretreated at either 185°C or 210°C. Selected SNP markers associated with each trait (p < 0.01, lowest p-value in a 1700 Kb region) found in the same unigene were highlighted.

	Unigene	Position		Common traits	Likely Candidate ID	
		Chr	Pseudo		AGI	Name
1	JCVI_33410	A6	4824152	185 2FA + 185 Formic + 185 Acetic	AT1G12780	UDP-D-Glc/UDP-D-Gal 4-Epimerase 1
2	JCVI_28444	C3	24863641	185 2FA + 185 HMF + 210 HMF	AT3G09230	MYB 1
3	JCVI_33410	C5	5199859	185 2FA + 185 Formic	AT1G12780	UDP-D-Glc/UDP-D-Gal 4-Epimerase 1
4	JCVI_690:1	A8	13196172	186 2FA + 185 Formic	nr. AT4G14300	-
5	JCVI_17322	C3	56294303	187 2FA + 185 Formic	nr. AT4G14230	-
6	JCVI_23579	C4	50409451	188 2FA + 185 Formic	AT1G56010	NAC domain containing protein 21
7	JCVI_3790	A2	24353905	185 Formic + 185 Acetic	Multiple	ACYL-activating enzyme 3 (AT3G48990), MYB45 (AT3G48920), Pectin lyase-like (AT3G48950)
8	JCVI_7947	A5	16512750	185 Formic + 185 Acetic	AT3G20220	SAUR-like auxin-responsive protein family
9	JCVI_36229	C1	3956907	185 Formic + 185 Acetic	AT4G31370	Fasciclin-like arabinogalactan protein 5 precursor
10	JCVI_15537	C2	43992872	185 Formic + 185 Acetic	AT3G27785	MYB 118
11	JCVI_14456	C9	4637409	185 Formic + 185 Acetic	AT5G60950	COBRA-LIKE protein 5 precursor
12	JCVI_1535	C9	24392590	185 Formic + 185 Acetic	Multiple	Peroxidase (AT5G47000) or Low-level β-amylase 1 (AT5G47010)
13	JCVI_23924	A1	11013742	185 HMF + 210 HMF	AT4G15750	Plant invertase/pectin methylesterase inhibitor superfamily protein
14	JCVI_2668	A4	14942908	185 HMF + 210 HMF	AT2G30933	Carbohydrate-binding X8 domain
15	JCVI_3716	A7	12305247	185 HMF + 210 HMF	AT2G27100	Serrate
16	JCVI_33922	A8	11510806	185 HMF + 210 HMF	AT4G25620	Hydroxyproline-rich glycoprotein family protein
17	JCVI_41807	A8	17341483	185 HMF + 210 HMF	Multiple	UDP-Glycosyltransferase (AT1G10400) or Xyloglucan:Xyloglucosyl transferase 33 (AT1G10550)
18	JCVI_1123	A10	1552565	185 HMF + 210 HMF	AT1G04680	Pectin lyase-like superfamily protein
19	JCVI_25085	C2	1218598	185 HMF + 210 HMF	AT5G05170	Cellulose synthase 3
20	JCVI_20507	C3	24930758	185 HMF + 210 HMF	AT3G08900	Reversibly glycosylated polypeptide 3
21	JCVI_29426	C5	36033682	185 HMF + 210 HMF	AT2G18800	Xyloglucan endotransglucosylase/hydrolase 21
22	JCVI_308	C8	25089768	185 HMF + 210 HMF	AT3G52840	β-galactosidase 2
23	JCVI_11084	C8	30402343	185 HMF + 210 HMF	Multiple	Pectinacetylesterase (AT3G62060), Putative GH (AT3G62050) or SAM-transferase (AT3G62000)
24	JCVI_38459	A3	10748085	210 2FA + 210 Acetic	AT3G60320	-
25	JCVI_9783	C9	62422544	210 2FA + 210 HMF	AT5G04360	Pullunase 1 (Starch mobilisation)
26	JCVI_26090	A8	11707460	210 Formic + 210 2FA	AT4G24530	O-fucosyltransferase family protein
27	JCVI_6656	C1	36465929	210 HMF + 210 2FA	AT3G23590	REF4-Related 1 (associates with phenylpropanoid regulator MEDIATOR)
28	JCVI_2675	C3	3169994	210 HMF + 210 2FA	AT5G13980	Glycosyl hydrolase family 38 protein
29	JCVI_16847	C7	33723111	210 HMF + 210 2FA	Multiple	Inflorescence Deficient in Abscission (AT3G25655) or Devil 6 (AT3G25717)
30	JCVI_99	C8	20843738	210 HMF + 210 2FA	AT5G42080	Radial swelling 9
31	JCVI_1200	C9	56825835	210 HMF + 210 2FA	Multiple	Glc transporter 1 (AT5G16150) or TBL 19, 21 (AT5G15890-900)

Table 19 – Areas of the *B. napus* transcriptome that were associated with the intraspecific variance in fermentation inhibitor yields between 49 *B. napus* cultivars, for one or more trait or processing conditions. Areas were selected by extracting SNPs above a threshold ($p < 0.01$), locating the highest marker in a 1700 Kb region and observing the coincidence of these markers in the same unigene from multiple traits.

7.4.4.1 Candidates likely to alter CW acetylation between cultivars: Acetic acid released during pretreatment

CWs of *B. napus* and other non-grass angiosperms contain considerably more acetyl-groups compared to other biomass sources (Pronyk & Mazza, 2012). In dicot SCWs, un-substituted glucuronoxylan backbones are acetylated to varying degrees (Busse-Wicher *et al.,* 2014, Chong *et al.,* 2014) whereas PCWs are rich in XGs, which are acetylated exclusively on Gal-side chains, typically accompanied by a fucosyl residue. HG and RGI GlcA residues can also be acetylated to varying degrees (Gille & Pauly, 2012b).

As GWA studies are based on association between the trait and phenotype, candidates are likely to include (i) genes directly involved in acetyl addition, (ii) those competing for acetylation sites and (iii) genes adding sugar moieties in similar proportions to acetyl-groups. This perhaps explains why many on the candidates observed in this study were related to XG, Gal, Fuc, GalA and Rha synthesis.

Overall, the candidates suggest that the majority of the variation in acetic acid released following pretreatment was derived from XGs, pectic (mainly galacturonan and RGI) side-groups, Fasciclin-like AGPs and potentially lignin moieties. Markers were found in orthologous positions to XG synthesis included XG endotransglucosylase/hydrolase 18, 19 and 24 (AT4G30270, 80, 90) suggest that XG acetyl-groups may be a dominant source of acetic acid present in the pretreatment liquors.

Dicot XGs typically contain acetylated α-L-Fuc-(1,2)-β-D-Gal-(1,2)-α-D-Xyl branches (Carpita, 2011). Associated regions near to XG fucosyltransferases 0, 7, 6, 8, 9 (AT1G14020, 070, 080, 100, 110), Fucosyltransferase 12 (AT1G49710), β-Galactosidases (ATG35010) and Galactosyltransferases (AT3G14960) are likely to be involved in the synthesis and modification of acetylated glycan side chains, many of which are currently unknown (Pauly *et al.*, 2013). Assessment of these candidates by reverse genetic may help elucidate important genes involved in the synthesis of acetylated glycans.

It is interesting to observe association in regions close to a unigene with a similar sequence to GUX5 (AT1G08990) after pretreatment at 210°C. Two recent studies have shown that other members of the GUX family (GUX1 and 2) add GlcA to xylans before acetyl groups are added (Busse-Wicher *et al.,* 2014, Chong *et al.,* 2014). Molecular competition between carbohydrate side-chains and acetyl-groups may be reflected in the candidates observed in this study.

Here some evidence was found that acetyl-groups released during pretreatment at 185°C were hydrolysed from pectic rhamnogalacturonans. Two candidates associated with acetic acid release, Rha biosynthesis (RHM) 1 and 2 (AT1G78570, AT1G53500) were found in associated regions, in different parts of the transcriptome (Chr A6 and A5, respectively). These genes are required from UDP-Rha synthesis (Oka *et al.,* 2007). As Rha is found exclusively on rhamnogalacturonan side-chains, this strongly implicates these polymers as a source of acetyl-groups. Markers near to RHM2 associated more strongly with acetic acid yields after pretreatment ($p < 0.003$ compared to $p < 0.008$). This is consistent with reverse genetic studies that showed RHM2 as the dominant RHM isoform (Oka *et al.,* 2007).

A number of candidates related to methylation of CW components (presumably GlcA residues of HG and RGI) were located in regions associated with acetic acid release between cultivars. Likely candidates included: methyltransferases (AT2G26810, AT1G50000), plant inverase/methylesterase inhibitors (AT1G50325-340, AT3G10720) and S-adenosyl-L-methionine (SAM)-dependent methyltransferases superfamily proteins (AT1G26850, AT1G69523, AT1G55450 and AT2G45750). These observations are consistent with those seen by other researchers which suggested interplay between CW acetylation and methylation (Manabe *et al.,* 2013).

It is not currently known what substrates are used for CW acetylation, but it is likely that Acetyl-CoA and other acetyl-carriers are used (Gille & Pauly, 2012b). Gille and Pauly (2012b), also suggested that an esterase might be needed to hydrolyse the acetyl-group from its donor molecule. Here a number of potential acetyl-donors and esterases were found near to associated markers, including: Acetyl-CoA oxidase (AT3G06690), Long-chain acyl-CoA synthase 4 (AT4G23850), Acyl-activating enzyme 3 (AT3G48990) and 13 (AT3G16170), Acylamino acid-releasing enzyme (AT4G14570), a Core-2/I-branching β-1,6-N-acetylglucosaminyltransferase (AT1G51770, AT1G03520, AT2G45750) and GDSL-like lipase/acylhydrolases (AT5G03590,610). These genes could be involved in the provisioning of acetyl-groups for CW acetylation and warrant further study.

An associated marker adjacent to a CAD homolog 3 (AT2G21890), likely to be involved in the final stages of lignin deposition (Boerjan *et al.,* 2003), suggests that some of the acetic acid is liberated from naturally occurring acetylated lignin moieties (Del Rio *et al.,* 2007) at 185°C. Potential candidates such as peroxidases (AT1G49570, AT5G64120 and AT5G4700) also suggest that this may be the case.

Markers found in an orthologous region to Cellulose synthase-interacting protein 1 (Pom2/Csi1) coding region suggests that natural variations in cellulose microfibril

alignment genes (Bringmann *et al.*, 2012; Li *et al.*, 2012), could alter CW acetylation between cultivars. Similarly, presence of a COBRA-like protein 5 and 6 precursors (AT5G60950, AT1G09790) and Mildew resistance loci (ATG44110, AT4G02570), suggests a similar role for these proteins in altering CW acetylation.

Together, the candidate genes suggested here are consistent with current theories of CW acetylation which might alter biomass hydrolysis (Gille & Pauly, 2012b; Pawar *et al.*, 2013). The candidates identified also provide potentially interesting and testable, routes for expanding our knowledge of CW acetylation. Particularly interesting candidates implicate potential routes for complex carbohydrate synthesis.

7.4.4.2 Production of Hydroxymethylfurfural (HMF) from multiple sources.

HMF is produced from the decompositions of hexose sugars (Glc, Man, Gal, Rha, Fuc) (Jonsson *et al.*, 2013; Palmqvist & Hahn-Hagerdal, 2000). Therefore, HMF can be produced from many CW components. This was reflected in the number of regions associated with HMF release, which was higher than other traits.

At least some of the HMF released after pretreatment was probably released from residual starch – owing to the presence of Isoamylase 2 (AT1G03310) and Pullunase 1 (AT5G04360) genes involved in starch branching and debranching (Delatte *et al.*, 2005, Li *et al.*, 2007).

Other potential sources of Glc-derived HMF include the hydrolysis of cellulose, particularly from PCW tissues, resulting in associated regions within, or near to CesA3 (AT5G05170) (Burn *et al.*, 2002a) and RSW3 (AT5G63840) (Peng *et al.*, 2000) orthologues. Other candidates likely to alter cellulose accessibility to degradation included members of the TBL family 19, 21 and 32 (AT5G15890-900 and AT3G11030), with some more general candidates likely to be involved in Glc provisioning and recycling in growing tissues: UDP-Glycosyltransferases (AT1G22370, AT1G10400), O-GH family 17 protein (AT1G66250), βGs (AT3G24180, AT3G52840 and AT1G66270-80) and a Glc transporter (AT5G16150).

Associated markers in orthologous positions to pectin-related coding regions were also observed, namely: Plant inverase/pectin methylesterase inhibitor proteins (AT4G15750, AT2G31425, 30, 32), pectin lyases (AT1G04680, AT1G56010, AT3G24130, AT1G04680, AT3G24130 and AT3G24239), SAM-dependent methyltransferase (AT4G40000 and potentially AT3G62000) and Pectin acetylases (AT2G46930, AT3G62060). These enzymes could be related to HMF release either by the variable degradation of Rha/Gal-containing RGI, Rha/Fuc-containing RGII side chains (Harholt *et al.*, 2010) or by masking other CW components, preventing their decomposition.

Interestingly, some candidates involved in pentose production were also observed in regions associated with HMF release. For example, reversibly glycosylated polypeptide 3 (AT4G35783) is likely to be involved in the recycling of Ara sugar nucleotides (Bar-Peled & O'Neill, 2011), particularly in pod tissues (Rautengarten *et al.*, 2011). Similarly, L-fucokinase/GDP-L-Fuc pyrophosphorylase (AT1G01220) is required for the salvage of L-Fuc (Kotake *et al.*, 2008). Ara and Fuc will not breakdown to form HMF directly, the apparent associations suggest that the abundance of Ara-containing polymers alters the accessibility of hexoses to degradation.

Other factors that could influence HMF release include those alter the development of the shoot and floral meristem including. These may include Rotundiflora-like 6/DVL 17 (AT4G35783) (Wen *et al.*, 2004), Serrate (AT2G27100)(Grigg *et al.*, 2005) and Wuschel 1 (AT2G17950) (Laux *et al.*, 1996). Whether markers are coincidentally associated with HMF release, or show a genuine association with the trait would require further study.

7.4.4.3 **Furfuraldehyde (2FA) is primarily derived from xylans.**

2FA is the primary breakdown product of pentose sugars found in plant CW namely: Xyl and Ara (Palmqvist & Hahn-Hagerdal, 2000). This perhaps explains why unigenes likely to be involved in the synthesis and recycling of UDP-Ara (UDP-D-Glc/UDP-Gal 4 epimerase 1, Xyl 4 epimerase 1), xylan assembly (IRX7, UDP-Xylosyltransferase 2), xylan branching (fucosyltransferases AT1G14020, AT1G49710 and AT4G24530) and xylan hydrolysis (β-Xylosidase 1 and 3) were all located near to markers associated with 2FA release.

Other candidates associated with 2FA release after pretreatment at 185 °C included those involved in phenylpropanoid synthesis, HCT (AT5G48930) and F5H1 (AT4G36220). Therefore, lignin deposition is likely to affect the degradation of non-carbohydrate components to 2FA, which appears to be particularly important when lower pretreatment severities are used.

Variations in 2FA may also be altered by the availability or structuring of other complex carbohydrates such as pectic components, implicated by pectin-lyases (AT1G19170, AT1G67750, AT3G07010, AT4G22090). Similarly, intraspecific variations in AGP synthesis (for example: Fasciclinin-like AGP9 and 11, and AGP galactosyltransferase 2) might alter 2FA release. These robust components might prevent the hydrolysis of pentoses during pretreatment.

7.4.4.4 Intraspecific differences in carbohydrate cross-linking influence formic acid production during pretreatment.

HMF and 2FA can decompose further to formic acid (Palmqvist & Hahn-Hagerdal, 2000), which is likely to be catalysed by acetic acid release (Chapter 6). Testament to this is that some of the most highly associated markers were for formic acid release, were common to 2FA, HMF and Acetic acid release at 185°C. Unique candidates related to formic acid release could highlight genes related to the abundance of the most delicate polymer groups.

Among the most highly associated SNP markers included those near to UDP-Glc 4 Epimerase (AT1G64440) – responsible for the synthesis of additional UDP-Gal in response to stress (Rosti *et al.*, 2007). The synthesis of Gal is probably used to produce AGPs that covalently link non-cellulosic components *in muro.* This hypothesis is supported by numerous candidates associated with formic acid release related to AGP synthesis. These included AGP18 (AT4G37450), AGP22 (AT5G53250), AGP31 (AT1G28290), Fasciclin-like AGP 2 (AT4G12730), Fasciclin-like AGP precursor-5 (AT4G31370) and APAP 1 (AT3G45230).

Remarkably, recent research has shown that both APAP1 and AGP31 form covalent cross-links between CW polymers (Tan *et al.*, 2013 and Hijazi *et al.*, 2014). One might postulate that other AGPs suggested in this study (AGP18, 22, Fasciclin-like AGP2 and 5) might form similar tethers that limit/promote the breakdown of CW material during pretreatment.

An orthologous region associated with Formic acid release after pretreatment at 185°C included: Galacturonosyltransferase (GAUT 3 - AT4G38270). GAUT3 mutants generate CWs with decreased recalcitrance and has already been patented for this purpose (Mohnen *et al.*, 2013). Mutants deficient in GAUT3 gene contain ≈25% more CW Glc than WT plants and less Rha and Fuc (Caffall *et al.*, 2009). Variations in the orthologous gene in *B. napus* is likely to have a similar effect, changing the composition of the straw and therefore the amount of formic acid released upon pretreatment.

7.5 Conclusions

The aim of this chapter was to observe the patterns of genetic association with each processing trait across the *B. napus* transcriptome. Owing to the small number of cultivars analysed in this pilot-scale GWA study, few markers could be identified with the very high degree of certainty typically needed for GWA studies. Therefore, lone candidates may only provide tentative explanations for the patterns of association observed for each trait.

Nevertheless, the design of this study allowed a unique opportunity to compare areas of the transcriptome that consistently associated with different traits. Comparing and contrasting datasets gave a higher likelihood of revealing the genetic determinants of these traits. In total, 41 areas of the transcriptome associated with more than one saccharification trait. Similarly, 31 areas were associated with one or more fermentation inhibitor.

Corroboration was also seen by comparison with the literature, for example, seventeen markers associated with glucose yields were found in close proximity to endogenous endo-1,4-β-glucosidases and βGs, likely to hydrolyse cellulose *in planta*. External validation was also seen by noting candidates identified in other association studies using similar genotype information, particularly those collected by Wang *et al.*, (2011) and Penning *et al.,* (2014). Also, independent detection of candidates such as APAP1 and AGP31 which have recently been shown to cross-link CW polymers (Tan *et al.*, 2013 and Hijazi *et al.*, 2014) in areas of the transcriptome associated with biomass processing traits were also observed. Together, these suggest that at least some of the associations seen here could be biologically informative.

The results from this pilot-scale GWA study not only concur with previous findings, but also provide a number of 'novel' targets worthy of closer inspection. For example, markers associated with saccharification yields could help clarify routes for D-Xyl recycling *in planta*, which are currently unknown (Geserick & Tenhaken, 2013).

GWA studies provide an excellent overview of the patterns of association related to a trait across the genome. However, extending our understanding beyond this initial survey requires more detailed analysis of candidate genes. Likewise, SNP markers would also benefit from further validation, to test their robustness and suitability for application. Therefore, in the following chapter, a selection of cultivars of known transcriptome sequence, but unknown phenotype was sourced. Phenotype data was collected and robustness of SNPs in predicting saccharification potential was assessed. A selection of candidate genes was also examined in greater detail by collecting phenotype data from mutant *Arabidopsis* plants, deficient in those genes.

8 High-probability marker and candidate gene assessment in *B. napus* and *Arabidopsis*

8.1 Chapter Outline

GWA studies are inherently observational in nature. They reveal areas of the genome that associate most strongly with a particular trait, under particular conditions, thereby indicating likely genetic variants associated with a given phenotype. In theory, this information could be used to select superior germplasm that has a more favourable CW composition for bioethanol production. However, the success of this strategy will depend on the robustness of individual markers at predicting phenotypes after processing. Similarly, candidates would benefit from further validation using reverse genetics. Studies of this kind may highlight mechanisms for biomass and assess the potential of the applicability of particular molecular markers.

Results from this chapter showed that singular SNP alleles could not predict the phenotypes of different cultivars grown under glasshouse conditions when taken in isolation. Small genetic differences attributed to singular loci were probably occluded by differences in growth conditions between the datasets. Nevertheless, analysing stem tissue collected from *Arabidopsis* mutants deficient in a selection of candidate genes showed that associated markers may point to genetic variations related to biomass recalcitrance. Plants deficient in XIFP (AT5G57655), GUX5 (AT1G08990), MYC4 (AT4G17880), TBL8 (AT3G11570) and RWA4 (AT1G29890) showed subtle, but statistically significant differences in saccharification yields compared to wildtype plants when processed under particular conditions. Interestingly, plants deficient in GUX5 showed little difference in overall CW composition from WT plants but achieved higher Glc yields (+6-9%) after enzymatic hydrolysis when untreated or sub-optimally pretreated stem material (185 °C) was hydrolysed. Targeting low abundance components may improve saccharification yields without harming CW integrity or plant functioning.

8.2 Introduction

GWA studies focus our attention on key areas of the genome that are most strongly associated with the variation in a particular trait. This provides an insight into the genetic determinants of that trait and a potential route for application, through marker-assisted selection. However, molecular markers must be sufficiently robust to predict the phenotype of the individual based on these markers alone.

Therefore, high-probability markers generated GWA studies from must be validated before their use (Chanock *et al.*, 2007). Ideally, re-phenotyping should focus on a

smaller selection of marker loci to minimise the effect of multiple-testing errors. However, genotype-phenotype association studies are notoriously difficult to replicate, even when true associations are observed (Liu *et al.*, 2008).

Candidate genes would also benefit from further investigation, to elucidate their roles in greater detail (Wang *et al.*, 2013c). This could validate observations made in *B. napus* and potentially expand our fundamental knowledge of CW recalcitrance across multiple species. Typically, candidate genes are validated using reverse genetic approaches, to determine the potential function of candidate genes and explore the molecular mechanisms responsible for the observed association(s) (Huang & Han, 2013).

The most convenient way of doing this is to analyse transferred DNA (T-DNA) insertion mutants deficient in the selected candidate gene(s) (Huang & Han, 2013). To produce these mutants, a short section of DNA including an antibiotic resistance gene is ligated into the plants genomic DNA using an *Agrobacterium* vector. Insertion of this additional DNA disrupts the function and therefore prevents the synthesis of the protein of interest (O'Malley & Ecker, 2010). The progeny of the infected diploid plants can contain T-DNA insertions on either both (homozygous mutant), one (heterozygous), or no (WT) copies of the gene. Seed from the transformed individual are grown on antibiotic plates to remove WT individuals from the first generation. The resulting mutant lines are therefore segregating and likely to contain homozygous and heterozygous insertion mutants and few WT individuals.

T-DNA insertion mutants for the model plant *Arabidopsis* are widely available and provide an excellent resource of genetically defined germplasm, ideal for exploring CW phenotypes in greater detail (McCann & Carpita, 2005). Moreover, homozygous T-DNA insertion mutants, with insertions located in the protein-coding region of the targeted gene (intron or exon) are very effective (90% success) at disrupting gene function (Wang, 2008). However, a significant hindrance to T-DNA mutant analysis is the need to identify homozygous individuals from segregating lines. Homozygous mutants can be identified by amplifying the region that the insertion was made, using the polymerase chain reaction (PCR) and noting if only larger fragments containing the inserted DNA are produced. Therefore, practical considerations currently limit the scale of reverse genetic studies to a small selection of potentially interesting genes (O'Malley & Ecker, 2010). This is the primary reason why few candidates identified in the GWA study were studied in this chapter. This is also why only phenotype data directly relevant to CW recalcitrance was collected in this section.

After genotyping mutant individuals, comparable phenotype information must be collected using model plant tissues. As *Arabidopsis* and *B. napus* lineages diverged

relatively recently (14-20 MYA)(Yang *et al.*, 1999) and both produce 'type I' CWs (Zablackis *et al.*, 1995, Yokoyama & Nishitani, 2004), they are likely to create similar CW structures. Indeed, the *Arabidopsis* inflorescence stem appear to be an excellent model to study secondary CW development in most woody plants (Strabala & MacMillan, 2013). Therefore, it is likely that comparable saccharification data can be collected from *Arabidopsis* tissues.

Nevertheless, the collection and interpretation of phenotypic data from *Arabidopsis* mutants can be quite challenging. Plants deficient in important CW genes can result in strong phenotypes (such as sterility, dwarfism and collapsed xylem structures) (Endler & Persson, 2011), which can cause phenotypic changes that are not directly related to the actions of that gene *per se*. The interdependance of CW genes increases the risk of these effects. Pertubations in metabolic pathways caused by the removal of a key CW gene can disrupt CW signalling pathways, causing stress responses in mutant plants, leading to cryptic phenotypes in unassociated CW components (Seifert & Blaukopf, 2010).

Conversely, CW mutants may show no phenotype, despite playing important roles in CW functioning. CW genes are normally members of large, functionally redundant, gene families, where other isoforms compensate for the loss of one gene copy (Frankova & Fry, 2013). Gene x environment interactions may also mask the actions of key CW genes that show strong phenotypes only when exposed to particular conditions. Good examples of this are RSW mutants that only show strong phenotypes when grown at elevated temperatures (Peng *et al.*, 2000).

A small selection of potential candidate genes located in areas orthologous to those identified on the *B. napus* transcriptome was investigated using reverse genetics. Mutant plants deficient in genes likely to have different mechanisms of controlling CW recalcitrance were selected and grown alongside WT plants. Candidates included UDP-GlcA decarboxylase 2, XIFP, RWA4, TBL, GUX5 and MYC4. Experiments conducted in this chapter were guided by the following objectives:

- Test a selection of molecular markers using a second diversity panel to assess their suitability and robustness for marker-assisted selection.
- Select a number of interesting candidate genes likely to alter saccharification yields based on the observed association patterns and obtain *Arabidopsis* mutants deficient in those genes.
- Compare plants deficient in those genes, with WT plants in terms of growth and saccharification yields after processing under various conditions.

8.3 Materials and Methods

8.3.1 *B. napus* straw used to test high probability markers

B. napus straw was sourced from an existing diversity panel grown by Martin Broadley and colleagues (University of Nottingham). Three plants of each cultivar were grown in 5 L plant pots at randomised locations in a 30m x 7m polytunnel located at the University of Nottingham, Sutton Bonington Campus, Leicestershire (+52° 49' 56.65", -1° 14' 57.73").

Whole dry plants (stems, empty pods and seed) were collected at maturity (31 July 2013), cutting each stem at approximately 300 mm above ground level by hand. The stems were individually bagged before drying further (RT, 14 d). The bagged stems were transported to Elsoms Seeds Ltd, Lincolnshire and threshed individually by Mark Nightingale and colleagues. Straw (stems and empty pods) from three individual plants from a selection of 24 cultivars were used to test markers for analysis. The cultivars selected for this part of the study were: Altex, Amber x Commanche DH line, Apex, Aphid Resistant Rape, Chuosenshu, Darmour, Dippes, Hansen x Gaspard DH line, Janetzkis schlesischer, JetNeuf, Kromerska, Lesira, Prince, Rafal DH1, Ramses, Roxet, Slovenska Krajova, Tina, Topas, Verona, Wilhelmsburger and York. SNP alleles for each cultivar were identified by aligning the transcriptome sequences for each cultivar in TASSEL; noting the SNP allele associated with each cultivar at each position manually.

8.3.2 Phenotyping *B. napus* straw.

Straw from each cultivar was milled to pass a < 2 mm screen using a sieve mill. A representative sample was taken and cryogenically milled (3 min). Portions of this material (0.75 g) were pretreated at 5% substrate concentrations using a Biotage Initiator^{+} Microwave Synthesiser at 185°C and 210°C (10 min) in duplicate. The concentration of fermentation inhibitors (Formic acid, Acetic acid, 2FA, HMF) in the pretreatment liquor were quantified by HPLC. The water insoluble substrate was rinsed with ultrapure water (MilliQ) until clear and made back up to volume with sodium acetate acetic acid buffer (pH 5) before hydrolysing with the desired dose of Ctec2 (either, 7 or 36 FPU/g material).

8.3.3 Sourcing *Arabidopsis* Mutants

Mutants were ordered from the European *Arabidopsis* Stock Centre (uNASC) (Table 20). All seed sets were segregating T-DNA insertions mutants, with an insertion made within the coding region of the gene of interest. All accessions were of Columbia ecotype background (Col-0).

Mutant	AGI code	SALK Code	Fragment sizes (b)	
			Insertion	No-insertion
GlcA substitiution of xylan 5	AT1G08990	SALK029284	1128	518-818
Reduced Wall Acetylation 4	AT1G29890	SALK142291	1196	531-831
Trichome Birefringence-Like 8	AT3G11570	SALK022312	1127	600-900
Xylose Isomerase Family Protein	AT5G57655	SALK029807	975	434-734
UDP-GlcA decarboxylase 2	AT3G62830	SALK088283	1254	588-888
MYC4	AT4G17880	SALK052158	1106	515-815

Table 20 – List of mutant seed accessions ordered from uNASC. Insertions were mainly in exons, apart from TBL8 where the insertion was in an intron. Predicted diagnostic PCR fragment sizes indicating either WT genotype (no insertion) or T-DNA insertion within the gene of interest (Insertion).

8.3.4 Growing plants for genotyping and bulk seed collection

All mutants were grown from seed in controlled environment rooms located at the John Innes Centre, UK. Seed from each mutant line was initially sown together in 100 mm square pots (3 May 2013) and individual seedling transferred to single cells of a 40 cell tray (p40) after 14 d of growth. The trays were moved to a room equipped with fluorescent lighting with a 10 h/d lighting regime and temperature maintained at 22°C. The plants were watered by hand twice a day until maturity. After identifying homozygous individuals by PCR, the selected plants were allowed to develop flowering stems. The stems were covered before flowering with 380 mm x 900 mm micro-perforated cellophane bags (Focus Packaging & Design Ltd, UK) and sealed to collect ripened seeds. Upon maturity (16 weeks after sowing) plants were removed from the soil, dried (RT, 3d) before collecting the seed.

8.3.5 Identifying homozygous mutants

Seed provided by (uNASC) were segregating lines, potentially containing individuals with either a single insertion (heterozygous) with one active copy of the gene, or an insertion for both copies (homozygous). As only homozygous insertion mutants contain no functioning copy of the gene of interest, these individuals are likely to show the effect of a particular gene by their absence.

8.3.5.1 DNA extraction

Approximately 1 cm^3 of leaf tissue was removed from four week-old individuals placed inside a 2 mL Eppendorf® tube using forceps. The tubes were immersed in liquid nitrogen before grinding with an Eppendorf® micropestle. The ground tissue

was suspended in extraction buffer (400 µL, 200 mM Tris-HCl, 250 mM NaCl, 25 mM EDTA, 0.5% SDS, pH 7.5). The solution was vortexed and incubated (RT, 90 min) before separating the cell components from solubilised DNA by centrifugation (13000 RPM, 1 min). DNA was precipitated from the supernatant (300 µL) using cold isopropanol (300µL, ≈4 °C, 3 min). The precipitated DNA was collected by centrifugation (13000 RPM, 10 min) before discarding the supernatant and drying the pellet (RT, 1 h). The DNA pellet was reconstituted in 100 µL of autoclaved ddH_2O and stored at 4 °C.

8.3.5.2 **Primer design and Polymerase chain reaction (PCR) conditions**

Genomic primers flanking the gene of interest were designed for each mutant using the SALK T-DNA Primer Design (signal.salk.edu/tdnaprimers.2.html)(Table 21). A common T-DNA border primer (5'-ATTTTGCCGATTTCGGAAC-3') matching a sequence found in the insertion was used on all lines. Genomic primers were selected to contain a 40-60% G-C content to ensure stable binding of the primer to the template DNA (Table 21).

The PCR was conducted by using 1 µL of the extracted DNA solution, 0.2µL of each primer (RP+LP or RP+BP), 0.8µL of deoxynucleotide triphosphate solution (2.5 mM) with 1 µL of 10x PCR buffer (containing 15 mM $MgCl_2$), 0.05 µL of TaKaRa Taq™ DNA polymerase (5 U/µL) and 7.55 µL of sterile H_2O per reaction. The solutions were exposed to 40 cycles of: denaturation (94 °C, 30 sec), annealing (56 °C, 30 sec) and extension (72 °C, 1 min) before a final elongation (72°C, 7 min) and storing at 10 °C. PCR products were separated using gel electrophoresis – adding the PCR product (7µL) to Blue gel loading dye (2µL, New England Biolabs, UK) and separating the DNA along a 1% agarose gel containing ethidum bromide (2 µL/100 mL). The gel was held in a tank containing 0.5 x Tris-borate EDTA solution and separated by passing an electric current through the liquid (0.175 A, 150 V, ≈ 15 min). Bands of DNA were imaged under UV light using an Alphaimager® EP.

Mutant name	Left genomic primer			Right genomic primer		
	Sequence (5'-3')	G-C (%)	TM	Sequence (5'-3')	G-C (%)	TM
GlcA substitiution of xylan 5	GGAAAACCATTGAGCTTAGCC	47.6	63.6	GCATGCATGATGACATAATCG	42.8	63.9
Reduced Wall Acetylation 4	TCTGCGATGAAAGACGAATG	45.0	60.0	GAAAACACAGAGTGGCTGAGG	52.4	59.9
Trichome Birefringence-Like 8	CTCTCGCAAAAGAAGATCGC	50.0	60.2	TAGAACATCAGAGCCGACCC	50.0	60.2
Xylose Isomerase Family Protein	GTAATACATCACGGTCGTCGC	52.3	64.2	TTTAACTTGGTTTCGGTGTGG	42.8	63.5
UDP-GlcA decarboxylase 2	ATCAATTGCCGTAACAACAGG	42.8	63.5	GCAATCTTATGAAGACGGCTG	47.6	63.5
MYC4	TTTATGCGATTCCGAAATGTC	38.0	63.5	TGAAAGCTTCCATTACCGATG	42.8	63.7

Table 21 – Genomic primer designs used to identify homozygous/heterozygous mutants from segregating lines.

8.3.6 Growing mutants for phenotyping

Seed-lots collected from individual plants confirmed to contain only homozygous T-DNA mutants were dried (RT, 4 weeks) before re-sowing. Seed from each lot was sown on Scott Levington F1 low nutrient growing media (Scotts, UK) held in 50 mm square pots. The seeds were incubated in darkness (4 °C, 4 d) to increase uniformity of germination. The plants were grown for the remainder of the experiment in a temperature and light controlled glasshouse (22°C, 16 h light / 8 h dark lighting regime).

Plants were grown to maturity under identical conditions, keeping each individual in isolation from adjacent plants using ARASYSTEM growth equipment (ARASYSTEM, Belgium) (Figure 43). Individual plants were pricked out after 7 d of growth into individual Arabaskets containing Scott Levington F2 compost (Scotts, UK) with added 4 mm grit (6:1). The soil was treated with an insecticide (Intercept®, Everris, UK) prior to pricking out. In total, ≈25 plants of each mutant and WT (Col-0) were grown at a randomised position across three ARASYSTEM growth trays.

Aracon bases and tubes were added after a > 10 cm high flowering stem had formed on the majority of plants (≈26 days after planting, DAP). All plants were harvested 50 DAP. At this point, all plants were in their final growth phase when the first seed pods were beginning to dehist (Boyes *et al.*, 2001). Whole plants were cut at the base, transferred to perforated polythene bags, dried (40 °C, 72 h) and milled (< 1 mm) before analysis.

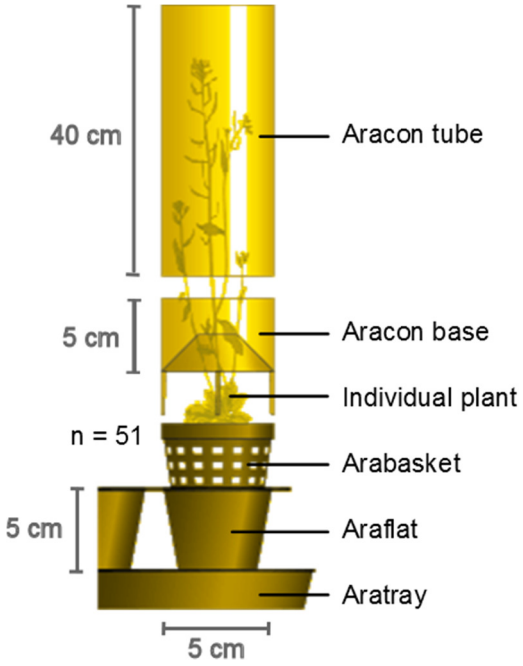


Figure 43 - Diagram of ARASYSTEM growth system adapted from 'www.arasystem.com'.

8.3.7 Microscopy of mutant stems

Stems isolated from mutant plants were initially fixed in 4% formaldehyde in PBS solution (1%) using vacuum infiltration (15 min) followed by soaking in the solution overnight (4°C). The tissue was manually rinsed in 50%, 60% and 70% ethanol (30 min each, RT). Tissue samples were placed in Tissue-Tek® Mesh Biopsy Cassettes (Sakura, Japan), loaded into a vacuum infiltration processor (Tissue-TEK® VIP150F, Sakura, Japan) and automatically rinsed with 70%, 80%, 90% ethanol (RT, 1 h 30 min each) followed by 100% ethanol (RT, 2 h). The tissue was rinsed with 100% ethanol (1 h, 35°C) before vacuum infiltrating the tissues with 100% ethanol (1 h, 35 °C). The tissues were vacuum infiltrated thrice with 100% xylene (3 h 30 min in total, 35 °C) and vacuum infiltrated with hot paraffin wax four times (2 h each, 60 °C).

The wax-infused tissues were set in paraffin wax blocks using a Tissue-Tek® TEC® 5 Tissue Embedding Console System. The embedded tissue was sectioned using a microtome (10 µm thickness) and mounted on SuperFrost Ultra Plus® slides (Menzel-Gläser, ThermoScientific, UK). The mounted tissue sections were de-waxed twice using Histo-Clear™ (10 min ea.), rinsed twice with 100% ethanol (2 min ea.) and dried (RT, 30 min). Sections were stained for 15 min (Safarin O and Alcian Blue, in 0.1M sodium acetate / acetic acid buffer, pH 5), rinsed in ddH_2O and dried (1 h, RT). Tissues were viewed under a light microscope (Leica Microsystems).

8.3.8 Pretreatment and hydrolysis of mature *Arabidopsis* stems

Samples of sieve-milled tissue (< 1 mm) were made into 5 mL slurries containing 2.5% substrate (FWB) in ultrapure H_2O (Milli Q) and sealed in 2-5 mL Biotage Microwave Vials. The *Arabidopsis* tissues were pretreated in a Biotage Initiator+ Microwave Synthesiser at either 185 °C or 210 °C for 10 min. The water soluble pretreated substrates were extensively rinsed by centrifuging (2000 RPM, 2 min) and replacing the supernatant with ultrapure water (MilliQ) until clear, followed by sodium acetate / acetic acid buffer (0.1M, pH 5) to remove soluble material and equilibrate the solutions to pH 5. Microbial activity was prevented using thiomersal (0.01% w/w). The pretreated and rinsed substrates were hydrolysed using 7 or 36 FPU/g original material of Accellerase 1500. The concentration of sugars released was quantified using DNS and GOPOD, subtracting the abundance of sugar derived from the cellulase.

8.3.9 Soluble D-Xylose in stem material

Pooled stem tissue collected from XIFP deficient mutants and WT plants was freeze-milled and soluble sugars extracted from ≈20 mg FW stem tissue in duplicate using 70% ethanol (2 mL, 100°C, 5 min). Supernatants (1 mL) were dried over CaO in a

vacuum desiccator (3 d) and sugars reconstituted in 100 µL of ultrapure water (MilliQ). The abundance of D-Xyl in the reconstituted extract was quantified colorimetrically using a D-Xyl assay kit (Megazyme, IRL). Briefly, 10 µL of each sample as added to buffered mixtures containing NAD+ and ATP (pH 7.5, 280 µL) in a microtitre plate. Glc present in the sample was removed using a Hexokinase solution (2 µL, 6 min, RT), before background sample absorbance was taken (340 nm). A mixture of xylose mutarotase (converts α-Xyl to β-form) and β-xylose dehydrogenase (converts β-D-Xyl to D-xylonic acid + NADP) was then added. The amount of NADP released, which is proportional to the amount of D-Xyl in the sample is quantified (340 nm) after a second incubation period (7 min, RT).

8.4 Results and discussion

8.4.1 Can highly associated SNP markers predict traits relevant to bioethanol production in *B. napus?*

GWA studies are an excellent way of detecting the position of potentially important genetic associations related to the trait. The genetic variants identified could potentially be used to identify superior germplasm, with more favourable genetic composition, without phenotyping the individuals again (marker-assisted selection) (Collard *et al.*, 2005). Typically this technique focusses on a small number of robust markers that accurately predict favourable phenotypes.

Predictive markers should be able to segregate higher and lower yielding cultivars based on genotype alone. Therefore, phenotype data was collected from 20 cultivars with known transcriptome sequence were sourced from an existing diversity panel to see if their phenotypes also segregated base on the alleles present in those cultivars.

Phenotype data similar to those used in the previous chapter (7) was collected from the straw (stems and empty pods) from a second set of *B. napus* cultivars (Table 22). Phenotype data included fermentation inhibitors and saccharification yields produced after pretreatment at 185°C, or 210°C for 10 min and hydrolysing the remaining material with a low or high cellulase dose as before (Table 22).

Similar product yields were gained for straw collected from the test dataset (Table 22) compared to the training dataset, despite being grown under very different conditions (Chapter 6). Nevertheless, plants grown indoors tended to produce higher fermentation inhibitors after pretreatment at 185°C and generally produced less Glc after enzymatic saccharification than field-grown samples when processed using optimal pretreatment conditions (210H).

Table 22 - Summary of phenotype data collected from the test dataset.

Cultivar Name	Pretreated at 185°C, Yields (% FW)								Pretreated at 210°C, Yields (% FW)							
	Acetic	Formic	2FA	HMF	Glc, Low	Glc, High	RS, Low	RS, High	Acetic	Formic	2FA	HMF	Glc, Low	Glc, High	RS, Low	RS, High
Apex	1.36	1.50	0.03	0.04	8.72	15.93	10.31	20.77	2.79	2.32	0.23	0.10	12.28	19.56	13.68	21.48
Rocet	0.97	1.03	0.01	0.02	7.43	13.74	10.13	18.51	2.36	1.66	0.20	0.09	11.71	18.99	11.69	20.94
Verona	1.10	1.37	0.02	0.03	7.36	14.97	10.72	20.61	2.20	1.74	0.20	0.10	10.80	18.90	11.56	20.90
Prince	1.05	1.12	0.02	0.04	6.55	15.88	10.51	19.88	2.27	1.60	0.24	0.13	11.24	20.28	12.30	22.50
Amber x Commanche	1.31	1.22	0.02	0.02	8.53	16.67	11.66	20.61	2.84	2.06	0.19	0.07	12.11	21.23	13.16	22.57
Hansen x Gaspard	1.13	1.12	0.02	0.05	7.69	14.40	11.41	20.23	2.48	1.95	0.29	0.16	12.07	19.99	13.11	21.74
Rafal	1.17	1.18	0.03	0.04	8.05	13.74	11.11	17.86	2.50	2.01	0.26	0.12	12.09	17.59	13.10	18.56
Janetzkis schlesischer	1.25	1.25	0.02	0.04	7.59	14.54	11.06	19.21	2.83	2.08	0.23	0.12	12.32	18.60	13.20	19.94
Darmour	1.15	1.18	0.02	0.07	6.64	13.78	10.33	18.68	2.50	1.89	0.29	0.22	10.87	17.79	13.02	19.41
JetNeuf	1.07	1.25	0.02	0.03	8.33	16.15	10.71	20.40	2.42	1.83	0.29	0.12	14.12	19.48	14.18	21.05
Dippes	1.12	1.31	0.03	0.07	7.81	15.89	9.65	20.54	2.53	1.72	0.31	0.19	12.77	19.75	13.49	21.24
Kromerska	1.14	1.17	0.03	0.05	8.03	15.32	11.08	20.83	2.58	2.13	0.29	0.16	12.63	19.99	12.93	22.49
Ramses	1.09	1.26	0.02	0.04	7.74	18.03	11.73	24.14	2.41	2.21	0.23	0.13	10.99	22.12	14.83	23.72
Slovenska Krajova	1.14	1.07	0.03	0.03	7.25	13.98	10.99	19.34	2.56	1.61	0.29	0.14	10.10	17.87	11.40	20.38
Aphid Resistant Rape	1.12	1.73	0.02	0.02	7.32	14.31	11.20	18.15	2.65	2.53	0.25	0.10	11.65	17.12	12.86	18.80
Chuosenshu	1.08	1.02	0.03	0.09	7.05	14.47	11.73	20.35	2.40	1.48	0.26	0.27	11.56	19.78	13.45	21.94
Topas	1.32	1.32	0.03	0.02	8.62	17.74	13.36	23.23	2.67	2.08	0.21	0.06	12.81	21.08	14.91	22.89
Altex	1.11	1.20	0.02	0.03	7.43	15.02	11.77	20.50	2.69	1.90	0.24	0.12	11.55	18.22	13.31	20.01
Wilhelmsburger	1.16	1.06	0.03	0.06	6.88	14.59	10.62	19.97	2.62	1.63	0.32	0.19	10.97	20.17	13.47	21.16
York	1.14	0.96	0.02	0.04	6.95	13.59	10.48	18.53	2.56	1.44	0.29	0.16	11.14	19.42	12.93	20.23

* Acetic = Acetic acid, Formic = Formic acid, 2FA = Furfuraldehyde, HMF = Hydroxymethylfurfural, Glc = Glucose , RS = Reducing sugar, Low = low cellulase dose, High = high cellulase dose

High probability SNP markers identified in the GWA study (Chapter 7) could not predict the product yields relevant to bioethanol production for different cultivars, grown under different conditions, in a different growing season (For example, see Figure 44). Separating cultivars in the test dataset by single SNP alleles did not show the same segregation patterns as those in the test dataset.

The most likely reason for this is that the growth conditions of the training and test dataset were very different. The training dataset was grown under field conditions whereas the test dataset was grown in irrigated growing media indoors. Many disparities in growth conditions from exposure to pathogens (Underwood, 2012) to water and nutrient availability (Marschner & Rimmington, 1996; Tardieu *et al.*, 2014) were likely to effect the composition of the different materials. Other key differences may have included those related to extreme temperatures, such as low temperatures in the field during winter and above ambient temperatures when grown in a polytunnel in summer. These differences are known to impact *B. napus* CW composition (Kubacka-Zebalska & Kacperska, 1999; Solecka *et al.*, 2008). Biomass used in the test dataset will have dried more rapidly than the material collected from the field, being in warmer conditions and controlled irrigation. Indeed, many of the candidates observed in this study were involved in the reclamation of sugars from the cell wall, a process that is initiated in response to sugar starvation (Lee *et al.*, 2007b). If plants were dried rapidly, it is unlikely that the plant would have time to respond.

It is also likely that the effect of each individual gene on the overall phenotype is small. This would make it difficult to predict the phenotypes of the test dataset based on a select number of markers alone. In this respect, genome-wide selection may be more effective for biomass-related traits, rather than focussing on singular loci (Massman *et al.*, 2013).

In conclusion, it is likely that traits relevant to bioethanol production and the patterns in genetic associations that they causes are likely to be influenced by interactions between genotype and the environment. Moreover, each gene may only have a small but cumulative effect on the phenotype, which cannot be easily predicted from singular markers. Further work aimed at determining the effect of growth conditions on biomass recalcitrance is required and a study with greater replication may be needed to ascertain robust markers suitable for selection.

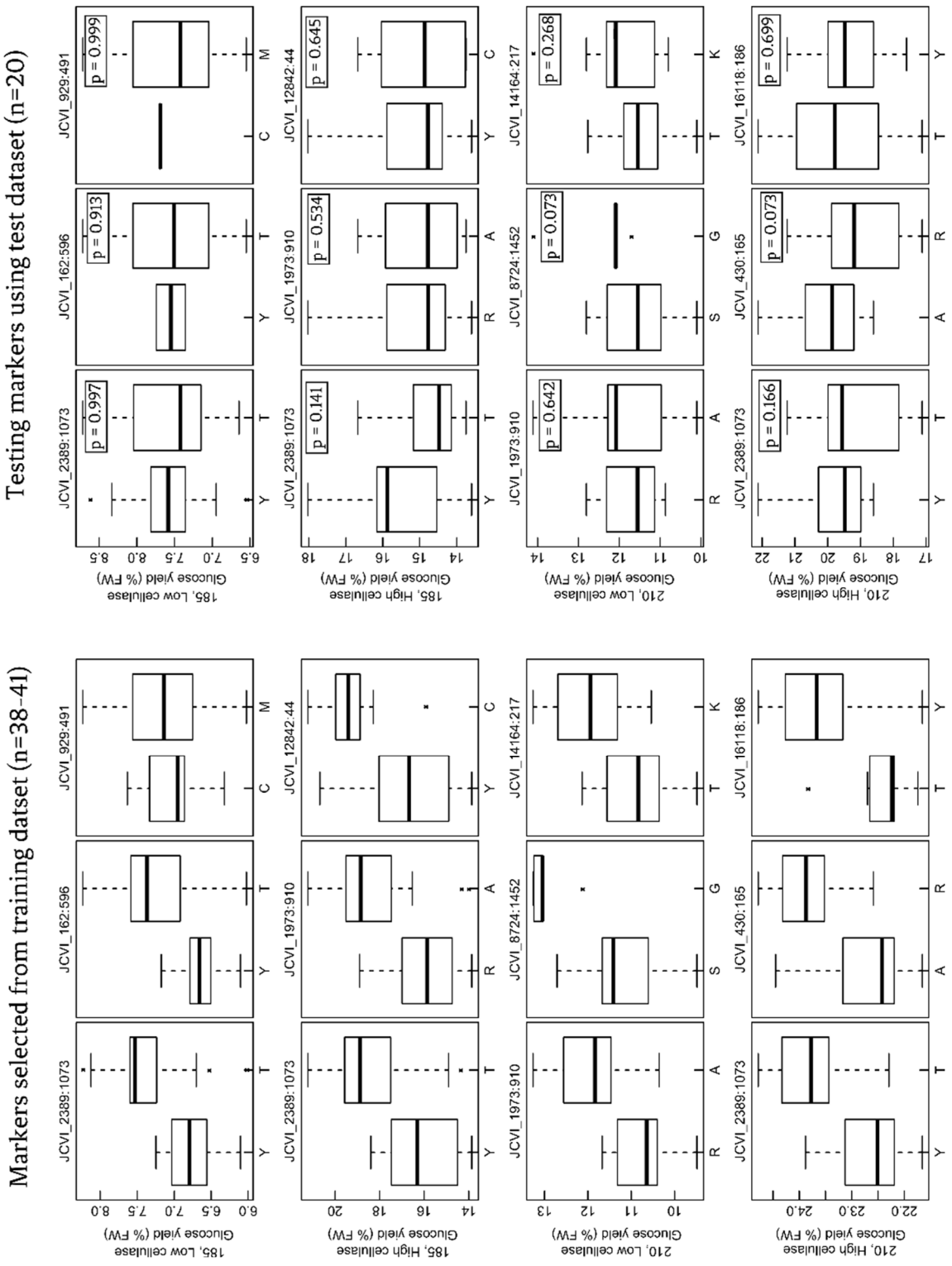


Figure 44 – Segregation of phenotypes based on singular alleles in both the test and training datasets. Even the most highly associated SNPs could not segregate individuals by singular markers - the top three alleles are given as examples.

8.4.2 Validation of candidate genes by reverse genetics in *Arabidopsis*

Although the SNP markers identified in the association study could not be used to predict the phenotypes of cultivars grown under different conditions, the patterns of association observed in the GWA study may still be biologically informative. Therefore, a selection of interesting candidate genes were chosen, loss-of-function mutants obtained and their impact on biomass recalcitrance determined at a range of processing conditions.

8.4.2.1 Confirming homozygous T-DNA insertion mutants.

A number of individual plants containing homozygous T-DNA insertions were identified from each accession isolating mutants deficient in GUX5, MYC4, RWA4, TBL8 and XIFP by analysing PCR fragments of varying sizes, depending on the genotype of the mutant (Figure 45). Genotype was confirmed by the presence of a ≈500 b product (indicating an insertion) and the absence of a ≈1 Kb fragment (indicative of a functional copy of the gene). Seed was only taken from individuals that produced a 500 b amplification product and no larger 1 Kb product – indicating a homozygous T-DNA insertion for both copies of that gene (Figure 45). As only lines containing an insertion in an exon or intron were selected, homozygous mutants are highly likely to be deficient in the gene of interest (Wang, 2008).

Unfortunately, homozygous individuals could not be isolated for UDP-GlcA decarboxylase 2 (UXS2) as all individuals tested (n = 30) appeared to be heterozygous mutants (containing both a T-DNA insertion and an uninterrupted copy of the gene). This suggests that homozygous mutants either did not survive to maturity or contamination of the DNA samples with WT DNA resulted in the production of a 1 Kb band in these wells.

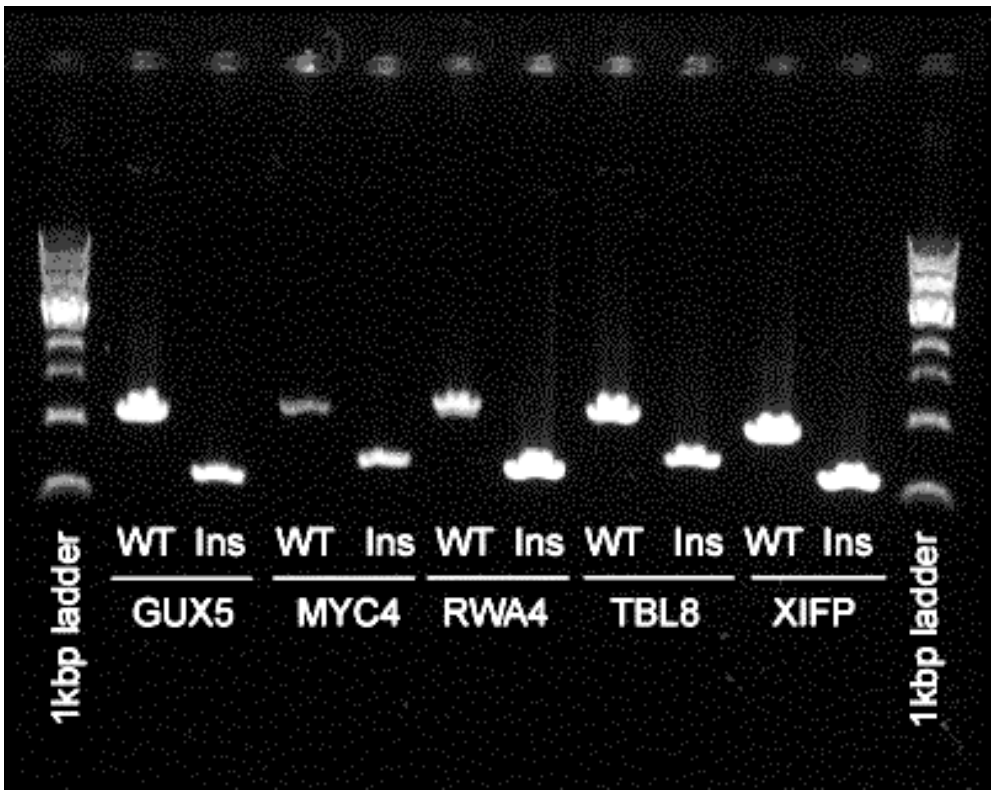


Figure 45 - Gel illustrating PCR products generated by amplification of DNA from wildtype (WT) and homozygous insertion (Ins) in the presence of gene-specific primers. Homozygous T-DNA insertion mutants were identified for AT1G08990 (GUX5), AT4G17880 (MYC4), AT1G29890 (RWA4), AT3G11570 (TBL8), AT5G57655 (XIFP) mutants.

8.4.2.2 **Changes to whole plant architecture**

Seedlings from each mutant (n=22-29) were grown in a randomised layout in individual cells in ARASYSTEM growth equipment. Photographs and plant measurements were taken for mature plants (50 DAP) to establish any differences in whole plant height, weight and stem branching. All mutant plants looked visibly similar to WT plants, with no obvious growth defects (Figure 46). Most mutant plants (XIFP, MYC4, TBL8 and GUX5) were slightly (≈5 cm), but significantly taller than WT individuals (Table 23). GUX5 mutants had slightly fewer main stem branches on average, compared to WT, but all had a similar number of basal tillers (Table 23). RWA4 mutants showed no significant difference in plant architecture for any measurement (Table 23).

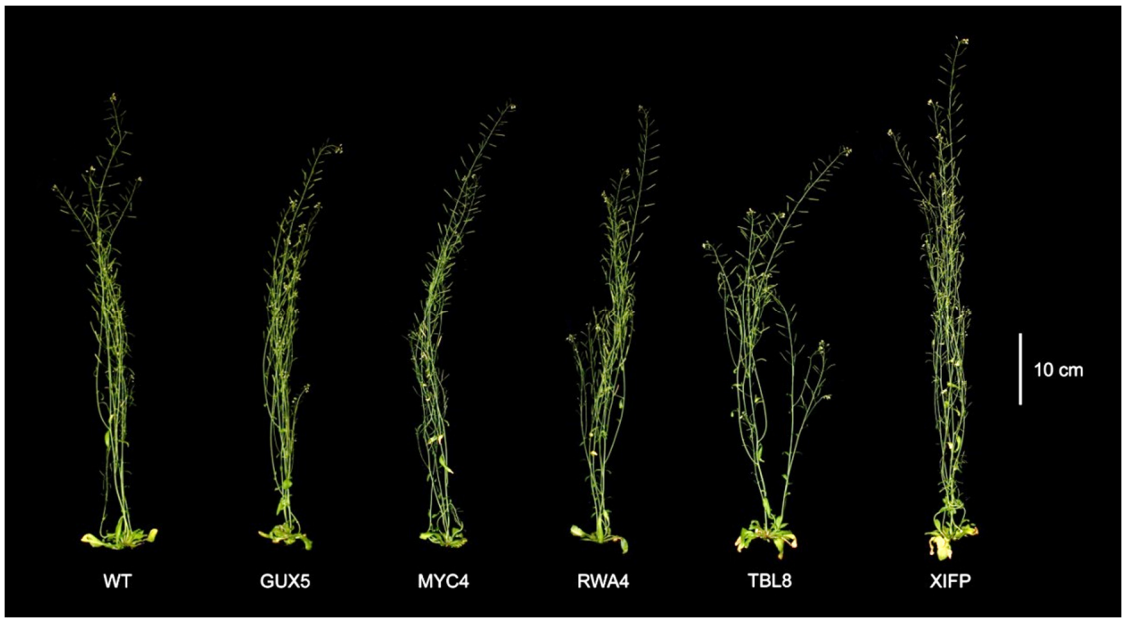


Figure 46 - Photographs of mature *Arabidopsis* mutants (50 DAP). Including GUX5 (AT1G08990), MYC4 (AT4G17880), RWA4 (AT1G29890), TBL8 (AT3G11570), XIFP (AT5G57655) mutants and a WT control (Col-0).

Genotype	Plants (n)	Height (cm)				Main stem branches (n)				Tillers (n)			
		Mean	±	Stdev	*p*	Mean	±	Stdev	*p*	Mean	±	Stdev	*p*
WT	24	57	±	7	-	3.5	±	1.4	-	3.4	±	1.2	-
RWA4	26	60	±	10	ns	3.0	±	0.9	ns	3.2	±	0.9	ns
MYC4	25	**63**	±	**3**	***	3.1	±	0.9	ns	3.7	±	0.8	ns
XIFP	27	**64**	±	**4**	***	3.3	±	1.0	ns	3.9	±	1.2	ns
TBL8	29	**63**	±	**6**	***	2.8	±	1.0	ns	4.0	±	1.0	ns
GUX5	22	**62**	±	**6**	**	**2.6**	±	**0.9**	**	3.6	±	1.5	ns

Table 23 – Differences in heights and branching patterns of *Arabidopsis* mutants (50 DAP).

8.4.2.3 Changes to stem microstructure of mutant stems

Mutant plants deficient in essential CW genes may also show other phenotypes that are not easily detectable to the naked eye. These phenotypes can be caused by compromised CW integrity, or plant stress responses caused by disrupted CW signalling pathways (Seifert & Blaukopf, 2010). Few consistent differences were observed between stem cross-sections taken from mutant stems. Only GUX5 mutants appeared to have less well developed small vascular bundles. This demonstrates that all mutants produced similar stem tissue to WT plants.

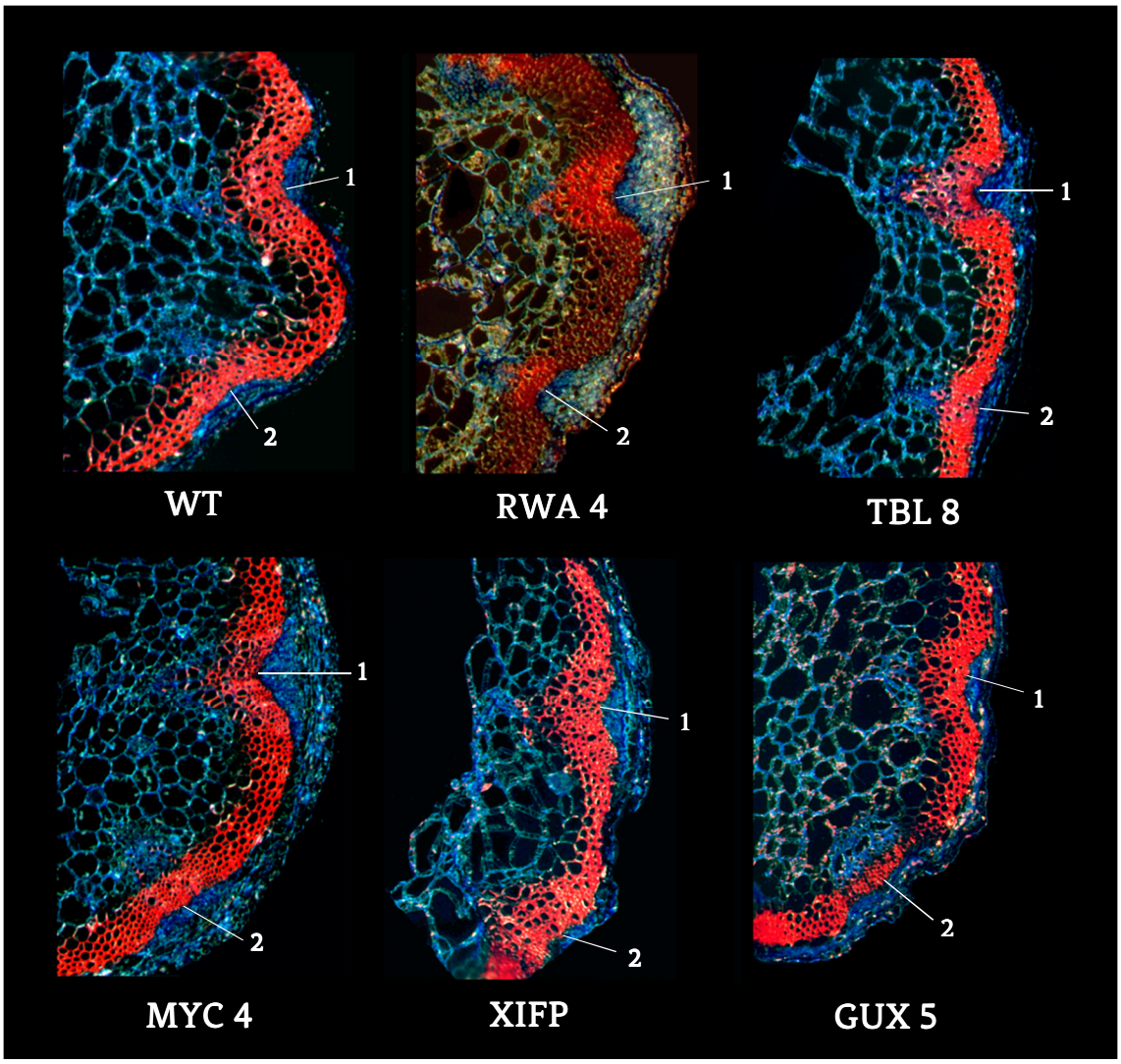


Figure 47 - Cross-section taken from base of the stem of *Arabidopsis* (10 µm thick). CWs were stained with Alcian Blue and lignified tissues stained red with Safarin O. Few consistent differences were observed between mutants and WT stems. However, it was noted that the development of the large (1) and small (2) vascular bundles differed between mutants.

8.4.2.4 FT-IR spectra of mutant stems

Genetic alterations that alter a single CW constituent is likely have consequences for other CW components. For example, altering lignin composition can change non-cellulosic polysaccharide compositions (Van Acker *et al.*, 2013), altering non-cellulosic polysaccharide compositions can change cellulose abundance (Bischoff *et al.*, 2010) and changing cellulose abundance can alter lignin deposition (Cano-Delgado *et al.*, 2003). FT-IR was therefore selected to gain a non-specific overview CW composition, which can be used to elucidate the main changes in CW composition (Chen *et al.*, 1998).

Inflorescent stems collected from multiple *Arabidopsis* plants of the same genotype (*n* = 24, 22, 23, 25, 25 and 21 for WT, GUX5, MYC4, RWA4, TBL8, XIFP, respectively) were milled together (< 2 mm) and a subsample freeze-milled into a fine powder. Spectra collected from each mutant give an indication of the largest chemical changes in stem chemistry between mutant and WT stem chemistry (Figure 48). Interestingly, some mutants had remarkably similar spectral changes compared to WT plants. For example, RWA4 and TBL8 showed similar spectral shapes, as did MYC4 and XIFP mutants.

Both RWA4 and TBL8 showed greater absorbance at pectin-related wavenumbers 1630 and 1550 cm^{-1} (Szymanska-Chargot & Zdunek, 2013) and a decrease in absorbance at ≈940 cm^{-1} (Figure 48) (α-1,4-glycosidic bonds between Xyl) (Kacurakova *et al.*, 1998). This suggests that some alterations to pectic components and/or their interactions with xylan moieties may be different to those of WT plants. This could include changes to hydration properties of CW components (Kacurakova *et al.*, 1998).

Loss-of-function mutants of MYC4 and XIFP also produce similar spectra-types, although the difference in spectra was stronger for MYC4 mutants (Figure 22). Main spectral changes common to both mutants suggest an enrichment in non-cellulosic polysaccharides, shown by increased absorbance at 1015 and 985 cm^{-1} compared to WT (Kacurakova *et al.*, 2000). However, these changes were matched by a decrease in absorbance at 1350 and 1390 cm^{-1} (bonds present in most matrix polysaccharides), consistent with non-specific shifts in carbohydrate structuring between mutants and WT plants (Figure 48).

Spectra of MYC4 differed from other mutants as they showed an increase in absorbance at 1720-40 cm^{-1} (C=O stretching vibration of alkyl ester), matched by a decrease in absorbance in a broad peak at ≈1600 cm^{-1} (COO- antisymmetric stretching) (Figure 48) – indicative of increased pectin methylation compared to WT plants (Szymanska-Chargot & Zdunek, 2013). It is also possible that the increase in absorbance at 1720 cm^{-1} indicate an increased cross-linking of xylans with lignin (Kacurakova *et al.*, 1999).

GUX5 mutants showed the smallest differences in overall spectra compared to WT amongst all the mutants (Figure 48). Therefore, GUX5 is likely to influence low-abundance components that only contribute slightly to the overall spectra. Stem material from GUX5 mutants showed a slight decreased absorbance at 1060 cm^{-1}, possibly implicating changes to Gal residue abundances (Kacurakova *et al.*, 2000) and small increase in absorbance at 1545, 1645 and 1390 cm^{-1}, likely to indicate an increase in pectic-like substances compared WT plants (Szymanska-Chargot & Zdunek, 2013).

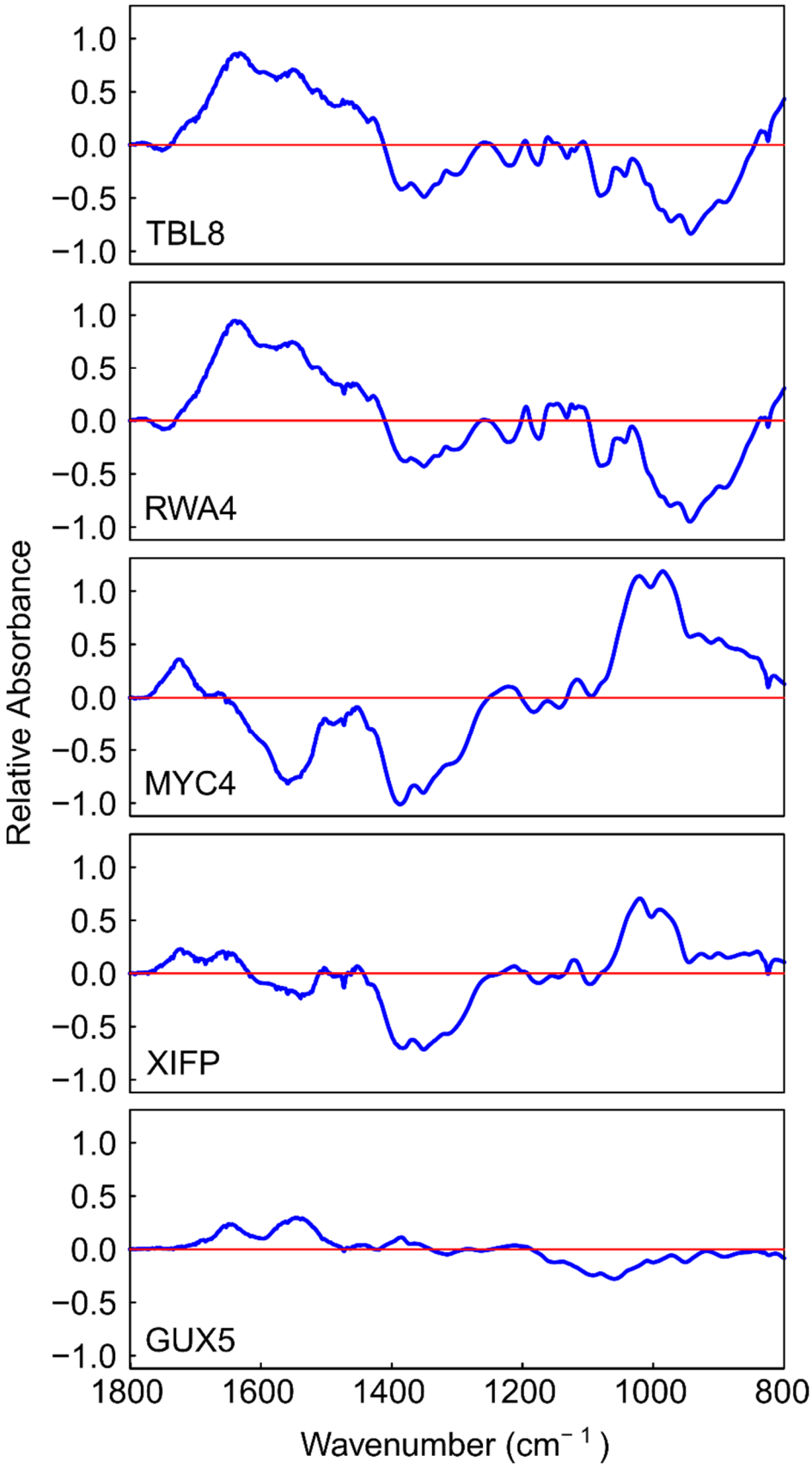


Figure 48 - Differences in mean relative absorbance between infra-red spectra of each mutant compared to WT.

8.4.2.5 Can suitable saccharification data be collected from Arabidopsis tissues?

WT *Arabidopsis* plants were grown in a growth cabinet (12 h light / 12 h day lighting regime) for 60 days to generate material to test saccharification methods. After this time, the plants were dried (40°C, overnight) and threshed. The three tissues analogous to those found in *B. napus* straw (the main stem, tillers and siliques) were separated. The tissues were pretreated and hydrolysed using the same conditions used to screen OSR straw (185°C, 210°C) to see if equivalent phenotype data could be generated using these conditions.

When processing *Arabidopsis* tissues under similar conditions to *B. napus* straw, comparable increases in sugar yields were observed after pretreatment and hydrolysis at varying severities (Figure 23). As *Arabidopsis* tissues contain a higher proportion of PCWs and produce less woody stems compared to those found in *B. napus* straw (Tofanica *et al.*, 2011), the proportion of sugar that can be hydrolysed from the material without pretreatment was considerably greater. Nevertheless, pretreatment still significantly increased the amount of Glc that could be derived from each tissue (Figure 23). The most significant improvement could be seen in the hydrolysis of the main stem, therefore this tissue was selected to phenotype the *Arabidopsis* mutant lines.

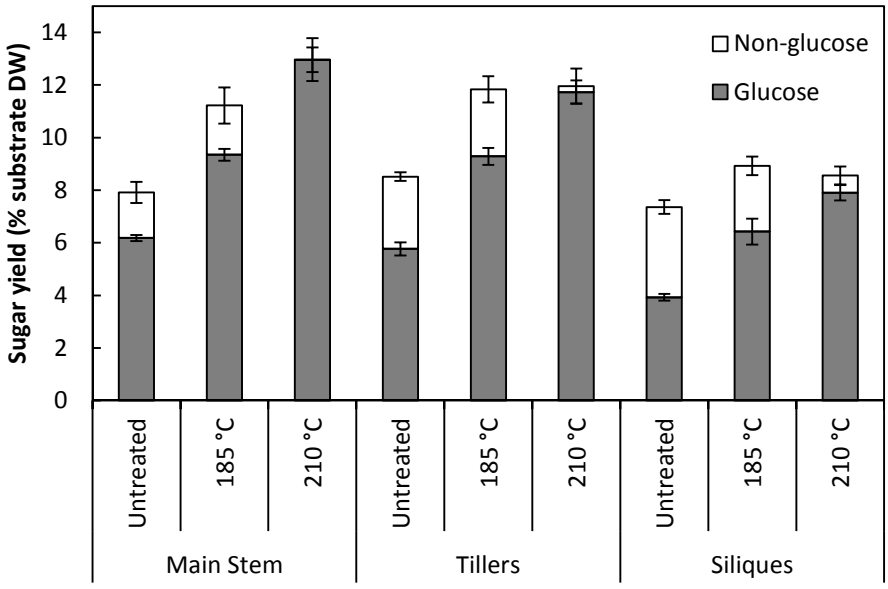


Figure 49 - Glucose and reducing sugar yields released from WT *Arabidopsis* tissues pretreated under various conditions (2.5% [substrate], 36 FPU/g). Non-glucose abundance is estimated from reducing sugar yield via DNS (RS-Glc).

8.4.2.6 Saccharification yields from *Arabidopsis* mutants, pretreated at varying severities

Stem material from each mutant was pretreated at 2.5% substrate concentration at 185°C or 210°C for 10 min, rinsed and hydrolysed at a low or high cellulase dose (7 and 36 FPU/g original material respectively, 96 h, 50°C). Untreated material was hydrolysed under the same conditions. All mutants showed significant differences in saccharification yields under different pretreatment conditions, with the majority of the loss-of-function mutants having reduced Glc and RS yields after processing, while GUX5 mutants showed increased Glc yields after processing at low severities (Table 24).

Pretreatment	Glucose (% DM)				Reducing sugar (% DM)			
	Low Cellulase		High Cellulase		Low Cellulase		High Cellulase	
Wild-type								
Untreated	3.5 ± 0.2	-	6.2 ± 0.2	-	5.9 ± 0.7	-	7.8 ± 0.7	-
185 °C	7.8 ± 0.1	-	10.8 ± 0.3	-	10.6 ± 0.3	-	12.6 ± 0.7	-
210 °C	7.2 ± 0.4	-	15.1 ± 0.6	-	10.0 ± 0.3	-	18.7 ± 0.2	-
RWA4 (AT1G29890)								
Untreated	3.3 ± 0.1	ns	6.0 ± 0.1	ns	5.8 ± 0.7	ns	7.2 ± 0.5	ns
185 °C	7.7 ± 0.1	ns	10.9 ± 0.4	ns	10.7 ± 0.4	ns	12.5 ± 0.9	ns
210 °C	7.2 ± 0.1	ns	15.2 ± 0.5	ns	9.4 ± 0.5	ns	**18.1 ± 0.2**	**↓***
TBL8 (AT3G11570)								
Untreated	**3.2 ± 0.1**	**↓***	5.9 ± 0.2	ns	5.4 ± 0.5	ns	7.3 ± 0.7	ns
185 °C	**7.5 ± 0.1**	**↓****	10.4 ± 0.5	ns	10.2 ± 0.4	ns	12.1 ± 1.0	ns
210 °C	7.4 ± 0.3	ns	15.0 ± 0.4	ns	**8.9 ± 0.4**	**↓****	17.7 ± 0.8	ns
MYC4 (AT4G17880)								
Untreated	**3.2 ± 0.1**	**↓***	6.1 ± 0.0	ns	5.5 ± 0.6	ns	7.5 ± 0.8	ns
185 °C	7.8 ± 0.1	ns	**10.2 ± 0.2**	**↓***	11.2 ± 0.6	ns	11.8 ± 0.7	ns
210 °C	6.5 ± 0.8	ns	14.8 ± 0.1	ns	**8.7 ± 0.5**	**↓****	**17.8 ± 0.3**	**↓****
XIFP (AT5G57655)								
Untreated	3.4 ± 0.3	ns	6.3 ± 0.2	ns	5.3 ± 0.6	ns	7.2 ± 0.4	ns
185 °C	7.8 ± 0.2	ns	10.2 ± 0.6	ns	10.8 ± 1.3	ns	**11.3 ± 0.6**	**↓***
210 °C	6.9 ± 0.4	ns	15.2 ± 0.5	ns	**8.8 ± 0.6**	**↓***	**17.3 ± 0.3**	**↓*****
GUX5 (AT1G08990)								
Untreated	**3.8 ± 0.1**	**↑***	**6.7 ± 0.2**	**↑****	6.3 ± 0.6	ns	8.3 ± 0.9	ns
185 °C	**8.3 ± 0.2**	**↑****	**11.5 ± 0.4**	**↑***	10.8 ± 0.7	ns	12.4 ± 1.0	ns
210 °C	7.7 ± 0.5	ns	15.8 ± 0.2	ns	10.0 ± 0.6	ns	19.0 ± 1.3	ns

*p < 0.05, **p < 0.01, *** p < 0.001

Table 24 – Glucose and reducing sugar yields derived from the stem tissue of different *Arabidopsis* mutants, hydrolysed after no pretreatment (Untreated) or pretreatment at 185°C or 210°C. The water insoluble material was hydrolysed using 7 and 36 FPU/g original material of Accellerase 1500.

8.4.3 Potential roles of each gene in determining CW recalcitrance

8.4.3.1 Glucuronic acid substitution of xylan 5 (GUX5)

GUX family members (1-5) have already been targeted as interesting candidates, likely to result in improved CW compositions for bioprocessing, based on observations from reverse genetic screens (Dupree & Miles, 2009). Previous work has demonstrated that single T-DNA insertion mutants of GUX1 and GUX2 show significant alterations to the degree of xylan substitution, with no significant change in growth (Mortimer *et al.*, 2010). More recent papers have demonstrated that GUX 1, 2 and 3 genes control the decoration of xylans with GlcA and secondary CW thickening (Lee *et al.*, 2012a) with isoform-specific patterns of decoration for each (Bromley *et al.*, 2013).

Early work conducted on GUX5 (AT1G08990) revealed its similarity to plant glycogenin-like starch initiation proteins (PGSIP), naming it PGSIP5 (Chatterjee *et al.*, 2005). Analysis of its sequences revealed that the "D x D" active site needed for UDP-Glc binding was lost in this isoform, leading the authors to state that it was unlikely to retain the GT family 8 activity (Chatterjee *et al.*, 2005). Since then, no study has looked specifically at GUX5, with Lee *et al.,* (2012) simply noting that GUX5 was most highly expressed in leaves and flowers, but did not investigate this gene further (Lee *et al.*, 2012a). Analysis of the structure of closely related GUX genes, suggest that GUX5 may have some similarity in function, but is unlikely to transfer GlcA to xylans (Rennie *et al.*, 2012). Rennie *et al.,* (2012) summarised that gene-knockout trials would be required to elucidate its precise role (Rennie *et al.*, 2012).

Results presented here showed that stem tissue collected from mutants deficient in GUX5 were slightly taller ($p < 0.01$), with slightly fewer stem branches ($p < 0.01$), but otherwise indistinguishable from WT plants (Figure 46, Table 23, pg. 200). This is consistent with previous observations (Lee *et al.*, 2012a). Stems from GUX5 T-DNA insertion mutants showed the most similar spectra to the WT out of all the mutant plants, indicating only a slight change in overall CW composition (Figure 48, pg. 203). Nevertheless, stem tissues showed improved Glc yields (+6-9%) compared to WT when tissue was insufficiently pretreated (No PT, or 185°C) (Table 24). In absolute terms, this increase in Glc yield (≈3.2-6.7 mg/g stem) would equate to an approximately 0.7-1.5% increase in total Glc yield, assuming stem material contained a similar amount of Glc to other *Arabidopsis* plants (44% Glc)(Van Acker *et al.*, 2013). This suggests that GUX5 may play a role in increasing plant CW residence to hydrolysis, particularly under low severities. Genetic targets that cause phenotypes such as this would be excellent for bioethanol producers wishing to increase saccharification yields in crop residues.

Overall, the results presented here add confidence that *At*GUX5 and its ortholog in *B. napus* could be important in determining CW recalcitrance. GUX5 is likely to act on low-abundance CW components, owing to its slight change in spectra-type compared to other mutants. GUX5 may be naturally more highly expressed in the flower and leaves (Lee *et al.*, 2012a), however this does not rule out the possibility of GUX5 being involved in additional strengthening of other tissues upon perturbation. It would be interesting to conduct sequential extractions of mutant CWs with varying expression of GUX5, to identify what CW component this gene targets. Further categorisation of this gene under various environmental conditions would also be useful to elucidate its role.

8.4.3.2 Reduced wall acetylation 4 (RWA4) and Trichome Birefringence 8 (TBL8)

B. napus markers (JCVI_29436:531 & JCVI_2480:239) found in neighbouring regions orthologous to RWA4 (AT1G29880) and TBL8 (AT3G11570), were the most highly associated SNPs associated with Glc yields (185L and 185H respectively) before genome assignment. After genome assignment, these previously highly associated markers were no longer present. This suggests that relatively rare polymorphisms for these markers exist across both *B. napus* genomes. Nevertheless, these polymorphisms could indicate rare but important variants related to CW synthesis. Also, RWA and TBL genes are likely to acetylate CW components (Gille & Pauly, 2012b, Pauly & Scheller, 2000, Rennie & Scheller, 2014). Therefore, these genes provide a potentially interesting mechanism of altering CW recalcitrance. Therefore these genes were selected for further investigation.

Modifying biomass acetylation has been targeted by other researchers as a method of improving saccharification yields after processing (Gille & Pauly, 2012a and b). This has included patents of the main genes thought to be involved in acetylation of plant polysaccharides, for the purpose of improved bioprocessing properties (Gille & Pauly, 2012a). Reducing biomass acetylation is thought to be beneficial, as acetyl groups hinder the actions of cellulases through increased steric hindrance and potentially inhibit downstream processing (Gille & Pauly, 2012a and b, Xiong *et al.*, 2013). Indeed, when hydrolysing untreated plant materials, reducing CW acetylation increases CW recalcitrance, leading some to dismiss CW acetylation as a target for breeding (Xiong *et al.*, 2013).

However, this simplistic view does not take into account pretreatments that are likely to be applied to biomass. As shown previously (Section 6.4.6, pg. 143), cultivars that produced more acetic acid under low pretreatment conditions also yielded more Glc after saccharification. Contrary to the view of the aforementioned researchers, higher degrees of CW acetylation may be beneficial to saccharification after processing, by

increasing the efficiency of autocatalytic pretreatment at lower severities. Therefore, noting saccharification yields produced after pretreatment at different severities may be needed to observe a phenotype.

CW acetylation can also alter CW dynamics indirectly. For example, an o-acetyltransferase in poplar known as "Gene-Y", orthologous to Reduced wall acetylation 2 (RWA2) in *Arabidopsis* was recently patented as a gene that improves CW saccharification and alters lignin composition (Chen *et al.*, 2014). Therefore, modifications to genes involved in acetylation may influence of other CW components, thereby altering saccharification yields.

Two labs have previously analysed *Arabidopsis* plants deficient in RWA4 (Lee *et al.*, 2011b; Manabe *et al.*, 2011). Both found, as this study did, that single RWA4 mutants were morphologically indistinguishable from WT. Neither studies could detect any change to the whole-tissue acetylation in leaf or stem (Manabe *et al.*, 2011). Likewise, the monosaccharide compositions of RWA4 leaf (Manabe *et al.*, 2011) and stem (Lee *et al.*, 2011b) tissue also did not differ from the WT. Both studies identified RWA4 was most highly expressed in SCWs, particularly at the base of stems, although the relative expression levels compared to other RWA mutants were contradictory (Lee *et al.*, 2011b; Manabe *et al.*, 2011). Results from Manabe *et al.,* (2011) showed that RWA4 as the most highly expressed RWA gene, while Lee *et al.,* (2011) showed it as third most highly expressed after RWA1 and 2. It is therefore likely that RWA4 expression is sensitive to environment and/or growth stage. Assessment of double, triple and quadruple mutants showed stronger phenotypes, with stunted growth, greater acetylation suggesting partial functional redundancy and overlapping functions of RWA family members and confirmed their integral role in CW recalcitrance (Manabe *et al.*, 2013).

No functional categorisation of TBL8 (AT3G11570) has been conducted so far. However, TBL8 shares significant homology to TBL29/ESKIMO1 (ESK1) (Xin *et al.*, 2007), a gene involved specifically in the acetylation of particular xylosyl-residues, which effects SCW thickening (Xiong *et al.*, 2013, Yuan *et al.*, 2013). Other TBL mutants such as TBL22 and 27 are involved in XG acetylation in specific tissues (Gille *et al.*, 2011). Is the TBL gene family is large (45 members in all) (TAIR, 2014), this suggests that members of this family are tissue-specific variants or partial functional redundancy in this family allows them to proliferate.

FT-IR spectra collected from mutants deficient in TBL8 and RWA4 showed remarkably similar spectra-types. Indicating an increase the absorbance of pectic components and decreased the absorbance corresponding to other CW (largely carbohydrate) polymers. Gille *et al.,* (2011) noted the similarity in acetylation of XG

between mutants deficient in RWA genes, compared to TBL mutants, suggesting that they may function in tandem, with RWA proteins functioning as less-specific acetyl activators (Gille *et al.*, 2011). Indeed, in this study, TBL8 had a stronger spectra-type compared to RWA4 (Figure 48), suggesting a similar role for both genes, with TBL8 having a slightly greater influence.

Previous attempts to hydrolyse tissue from RWA and TBL mutants conducted using untreated or alcohol insoluble material showed no differences compared to WT (Lamesch *et al.*, 2012; Xiong *et al.*, 2013). However, using untreated CW material only hydrolysed < 20% of the overall Glc and the change in CW composition may only be observed after pretreatment. The results presented here therefore extend those by previous researchers to include the effect following pretreatment.

Mutants deficient in RWA4 showed almost no change in CW recalcitrance under most conditions, with only slightly lower RS yields compared to WT under the most intensive pretreatment conditions (210H) (Table 24). On the other hand, TBL8 deficient mutants showed a reduction in Glc yield under cellulase-deficient conditions, when sub-optimal pretreatment and hydrolysis conditions were used (Untreated L, 185L) (Table 24). Similarly, mutant stem tissue pretreated at 210L, produced less RS compared to WT tissue. This suggests that WT plants with a functioning TBL8 gene produce more readily hydrolysable CW material. Whether this is caused by changes in acetylation of specific compounds is not known and would need to be tested in further detail. However, these results show that the actions of TBL8 is likely to decrease biomass recalcitrance.

8.4.3.3 **MYC4 transcription factor (MYC4)**

One of the most probable SNP markers associated with Glc yields after pretreatment at 210°C and hydrolysis using a low cellulase dose was found in a unigene (JCVI_14164), adjacent to a MYC4 (AT4G17880) ortholog on the *B. napus* pseudomolecule. Previous reverse genetics studies have shown that MYC4 performs similar roles to MYC3, as jasmonate-responsive transcription factors in response to pathogens and herbivores (Fernandez-Calvo *et al.*, 2011). MYC3 expression is ubiquitously expressed in all *Arabidopis* tissues, as MYC4 is expressed exclusively in vascular tissues. Plants deficient in both MYC3 and 4 were more susceptible to herbivores (Fernandez-Calvo *et al.*, 2011), suggesting a role in CW strengthening. However, it is not known what effect these genes will have on altering biomass recalcitrance after pretreatment and hydrolysis.

Interestingly, T-DNA mutants deficient in MYC4 showed the largest differences in spectra-type of all the mutants explored in this chapter (Figure 48, pg. 203). A

particularly interesting feature was the decrease in IR-absorbance regions related to the abundance of ionised carboxyl groups (COO^-) compared to WT (Figure 48, pg. 203). These groups are important in forming ionic cross-links between CW compounds mediated by Ca^{2+} (Carpita & Gibeaut, 1993). However, other spectral shifts related to non-specific changes in non-cellulosic components were also observed.

It is therefore likely that the response to herbivory observed by Fernandez-Calvo *et al.,* (2011), coordinated by MYC4 in WT plants, results in an enrichment of cross-linked pectins (HG). Other CW components respond accordingly, possibly through changes in carbon partitioning, as resources are routed away from UDP-Xyl synthesis, in favour of UDP-GlcA synthesis for pectins. This might explain the similar spectral features between MYC4 and XIFP mutants. Alternatively, changes in pectic properties could alter the interactions of other CW components.

MYC4 mutant showed decreased saccharification yields, with no obvious pattern related to pretreatment severity or enzyme dosing (Table 24, pg. 205). The changes in CW composition coordinated by MYC4 may cause unpredictable changes to CW functioning, as the plant will respond to a deficiency of those polymers by other means. Further work would be needed to establish the influence that this gene has on biomass recalcitrance, possibly under sub-optimal growth conditions. It would be interesting to follow the fate of sugar radioisotopes in mutant plants, to establish the effect of MYC4 and other mutants (particularly XIFP) on carbon partitioning between CW polymers (Sharples & Fry, 2007).

8.4.3.4 Xylose isomerase family protein (XIFP)

In the previous chapter, the frequent coincidence of SNPs associated with saccharification yields (all apart from 185L) were located within and around a unigene with similar sequence to an *Arabidopsis* putative XIFP (AT5G57655) ortholog. This suggested that this gene might be important in determining biomass recalcitrance. The most likely function of this gene, based on sequence similarity, is the conversion of D-Xyl to D-Xylulose (Jackson & Nicolson, 2002). Although the function of this gene has not been explored previously *in planta* and recent evidence expressing this gene in a fungi *Flammulina velutipes* shows that it functions as a xylose isomerase with specific activity to D-Xyl (Maehara *et al.*, 2013).

The majority of NDP-sugar recycling pathways have already been elucidated for other sugars (Bar-Peled & O'Neill, 2011) but the pathway, responsible for the recycling of D-Xyl is currently unknown (Geserick & Tenhaken, 2013). Therefore, it is possible that XIFP is involved in the recycling of D-Xyl for the production of new polymers with

different recalcitrance. This may be particularly important in the field, where abiotic factors are likely to limit sugar availability.

In the previous chapter, two recycling routes were proposed which could recycle D-Xyl, based on the candidates revealed in the GWA study. These were the conversion of D-Xyl back to UDP-Xyl, via a sugar kinase (possibly AT2G21370), or converting D-Xyl to a F6P, a generic precursor suitable for the production of other NDP-sugars (via XIFP), followed by the initial stages of the pentose phosphate pathway. This pathway would be similar to those seen elsewhere in nature (Jackson & Nicolson, 2002).

The latter pathway would entail the conversion of free D-Xyl to D-Xylulose (by XIFP), followed by conversion of D-Xylulose to Xylulose-5-P (by Xylulose kinase 2) (Hemmerlin, *et al.,* 2006). Xylulose-5-P can be converted to F6P via central metabolic pathways. This plausible route for the recycling of D-Xyl would be worthy of further investigation. Developing this hypothesis and identifying the target genes with which to test it was only made possible by noting areas of the *B. napus* transcriptome associated with saccharification yields. As only one loci in the *Arabidopsis* genome has been annotated with xylose isomerase activity (TAIR, 2014), XIFP - AT5G57655, it is possible that this route is governed by this gene alone. Therefore, one might expect mutants deficient in XIFP to show a D-Xyl deficient phenotype and consequently, have altered stem recalcitrance.

All *Arabidopsis* plants were grown with sufficient light, water and nutrients until their collection as green plants 50 DAP. Under these conditions, it is likely that sugar recycling within the plant will not be as important as plants grown under field conditions. This perhaps explains why XIFP deficient mutants showed no physical differences and Glc yields after processing were not significantly different from WT plants. Spectra collected from stem tissue of XIFP deficient mutants showed non-specific changes of IR absorbance similar to that of MYC4 mutants. These changes are consistent with XIFP alters CW composition in an indirect manner, possibly through perturbations in NDP-sugar abundances. These non-specific changes to non-cellulosic polysaccharide abundances may explain why RS yields released using conditions that yielded a large proportion of CW sugars (185H, 210L and 210H) were lower than WT plants while Glc yields remained unchanged.

Under conditions where plant sugars were sufficient to maintain CW synthesis, CW recalcitrance may not be hindered. However, if XIFP is involved in the reclamation of D-Xyl, one would expect XIFP deficient mutants to accumulate extracellular D-Xyl as long as other potential recycling routes cannot fully compensate for this increase. To directly test this hypothesis, extracellular sugars from freeze-milled XIFP deficient

mutants and WT stems were extracted using 70% ethanol (boiling 5 min) and quantified. XIFP deficient mutants appeared to contain ≈32% more soluble D-Xyl than WT plants (Figure 50), but this difference was not signficant (Student's t-test, $n = 4$, $p = 0.116$). Only ≈40 mg of stem material was available to test this hypothesis, therefore extractions could only be conducted in duplicate. Ideally, mutants deficient in XIFP could be grown in sufficient qualities to repeat this experiment to determine if XI deficient plants accumulate significantly more soluble D-Xyl than WT plants.

Here, plants were grown in near optimal conditions in terms of light, water and nutrient availability. Growing plants with sufficient daylight will probably limit the need to recycle D-Xyl, as sufficient photosynthetic carbon is available for CW synthesis. Using a short day length, or moving plants into dark conditions before harvesting would reveal greater reclamation of CW sugars (Lee *et al.,* 2007b). It would be particularly interesting to see if multiple T-DNA insertion mutants, deficient in both XIFP and other genes identified in the GWA study (Chapter 7) – namely sugar kinase AT2G21370 – would accumulate extracellular D-Xyl. This would not only confirm that these genes are important in recycling D-Xyl, but also the relative importance of the two potential D-Xyl recycling could be ascertained under various conditions.

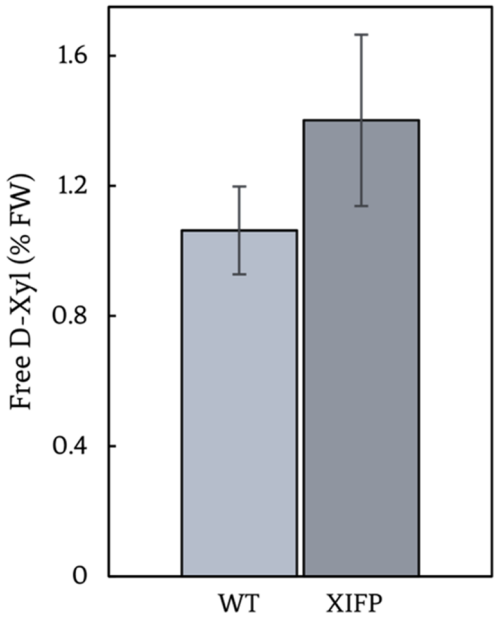


Figure 50 - Extracellular D-Xyl extracted from of WT and XIFP deficient mutants. Although more D-Xyl was released from XIFP deficient stems, the difference between the two groups was not significant (Student's t-test, $n = 2$ each, $p = 0.116$).

8.4.4 **Conclusion**

GWA studies present observational data, under conditions set in the study and are therefore notoriously difficult to replicate (Liu *et al.*, 2008). This perhaps explains why molecular markers identified in the original GWA study could not predict the phenotype of plants grown under glasshouse conditions. The disparate growth conditions used in the training and test datasets made replication particularly unlikely. This is probably why similar marker distributions were not observed between the two groups. Ideally, phenotype data would be collected from material grown under similar, field conditions. Also, testing singular marker associations may not be the best approach. Independent replication of a whole GWA study would be more informative. Singular genes associated with biomass recalcitrance appear to have relatively low effect sizes (small quantitive change to the phenotype). Therefore, large differences in growth conditions could have masked the effect of individual genes.

Arabidopsis mutants deficient in selected candidate genes were visibly similar to WT plants, but some showed spectral differences and subtle changes in saccharification yields. This suggests that at least some of the candidate genes identified in the GWA study may be potentially useful targets for crop residue improvement as they do not necessarily hinder plant functioning but can alter saccharification yields.

The results from this chapter serve as an example of the usefulness of the GWA study, not only to generate testable hypotheses, but also provide specific targets with which to conduct experiments. For example, this study began to explore the possibility of XIFP being involved in D-Xyl recycling *in planta.* Although significant increases in extracellular D-Xyl accumulation could not be demonstrated, the mean amount of extracellular D-Xyl was higher than that of the control. Further experimentation would be required to confirm or reject this hypothesis, ideally under different growth conditions to those used here.

Although the results in this section, only scratch the surface of the potential roles in which these genes may play in altering CW recalcitrance, they highlight specific avenues of research, which could be beneficial to our understanding of plant CW recalcitrance. Further work would be required to elucidate the roles of these genes in sufficient detail. However, these initial results suggest that promising candidates that alter saccharification yields by various mechanisms could be targeted for biomass improvement.

9 General Discussion and Conclusions

9.1 Summary of findings and key contributions

The general aims of this thesis were to identify strategies to improve the use of *B. napus* straw for the purpose of cellulosic bioethanol production and expand our knowledge of the process-specific and genotypic determinants of biomass recalcitrance. To do this, a series of objective were set to guide the work (Figure 51).

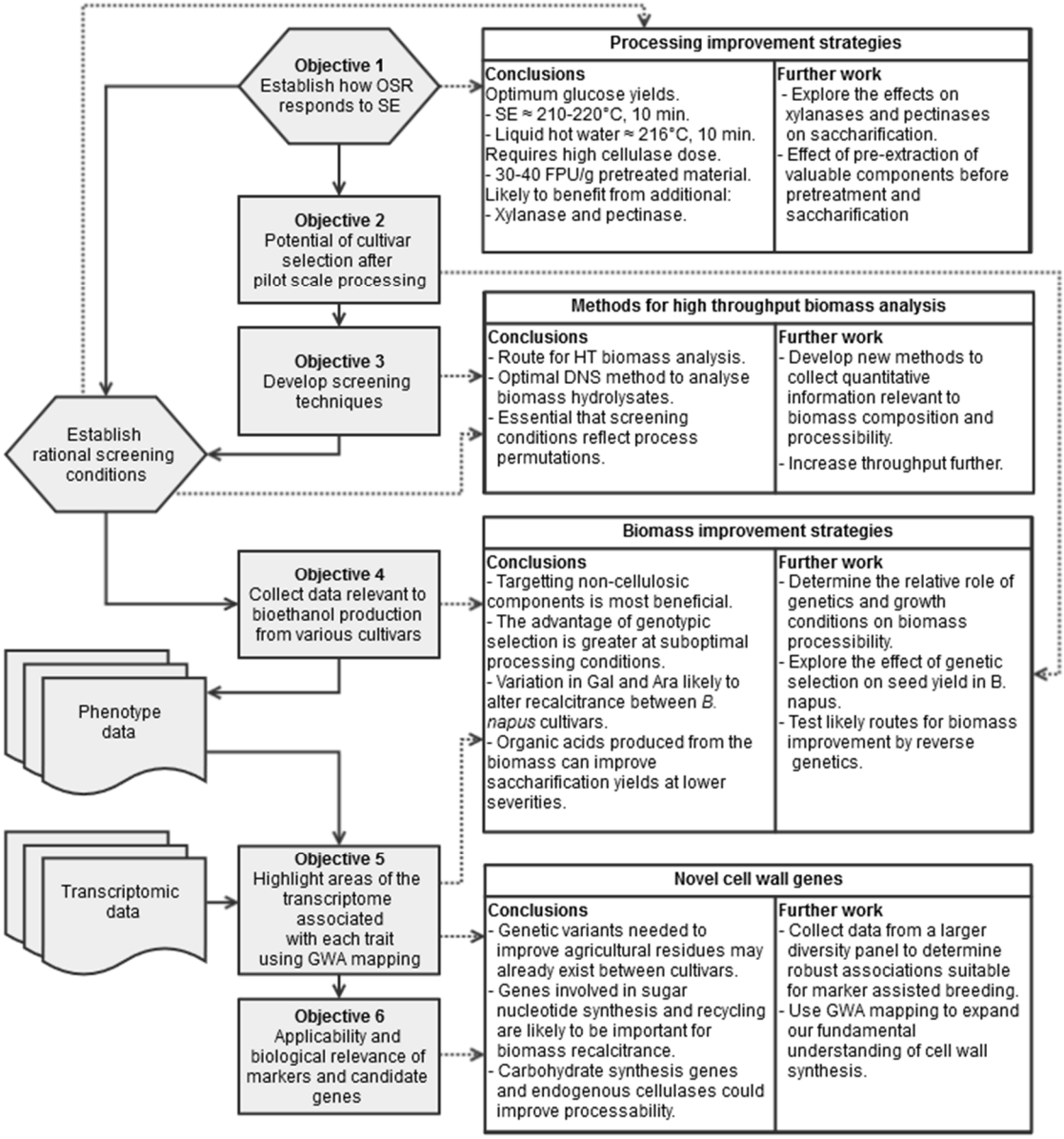


Figure 51 - Flow diagram illustrating the main objectives addressed in this study. The main outputs of each stage are displayed as boxes with dotted lines.

In this chapter, the main findings and key conclusions of this work will be summarised. The main limitations, existing gaps in knowledge and proposals of future research will also be outlined (Figure 51).

9.1.1 *Brassica napus* straw responds well to autocatalytic pretreatment, but saccharification performance could be improved by further removal of non-cellulosic components from pretreated substrates

This study explored how *B. napus* straw responds to autocatalytic pretreatment at varying severities, at both pilot (Chapter 4) and laboratory scales (Chapter 5). Pretreatment using SE (Figure 18, pg. 90) and liquid hot water (Figure 33, pg. 134) caused similar trends in saccharification yields, despite being conducted using different equipment and at different scales. Both showed that optimal pretreatment conditions in terms of maximum Glc yield (86-88% original Glc) was between 210 to 220°C, 10 min (Ro = 4.24 to 4.52). Under these conditions, extensive removal of non-cellulosic components was achieved which increased saccharification yields. Therefore, OSR straw may be particularly amenable to autocatalytic pretreatments compared to other biomass sources with a lower proportion of non-cellulosic β-1,4-D-glycan (e.g. gymnosperms), or those with less highly acetylated CWs (grasses) (Sarkar *et al.*, 2009).

This study focussed mainly on how to maximise the extraction of cellulosic-sugars, which was achieved primarily by the removal of non-cellulosic carbohydrates. However, non-cellulosic components could also be exploited to improve whole-crop value (Zhang *et al.,* 2008). As *B. napus* straw is particularly rich in non-cellulosic components, it would be interesting to see how pre-extraction of these components could improve saccharification performance of the remaining material, while creating a useable product.

Results from Chapter 4 demonstrated that commercial cellulases, optimised for the hydrolysis of other (mainly monocotyledonous) agricultural residues, perform relatively poorly when hydrolysing *B. napus* straw. A high cellulase dose (30-40 FPU/g original material) was required to achieve maximum Glc yields, compared to the dose needed to hydrolyse SE wheat straw (Horn *et al.,* 2011). The main limitations to saccharification of OSR straw steam exploded at varying severities were identified by comparing three key hydrolysis parameters: initial hydrolysis rate, proportion of carbohydrate hydrolysed at this rate and the final sugar yield (Chapter 4). Initial cellulase binding and hydrolysis rate was most closely associated with the amount of UA remaining in the pretreated residue. This potentially implicated residual pectic/glucuronoxylan side-chains retained within the material as limiting cellulase

binding. Final sugar yields correlated most strongly with Xyl removal, consistent with and extending previous observations (Ryden *et al.,* 2014).

To our surprise, the proportion of rapidly hydrolysable carbohydrate correlated most closely, and positively with lignin abundance. This suggests that the increasing proportion of klason lignin (most likely pseudo-lignin) in the pretreated residue allowed a greater proportion of sugars to be hydrolysed rapidly. This trend was unexpected as lignin and pseudo-lignin are typically described as a hindrance to hydrolysis (Akin, 2008, Kumar *et al.*, 2013). The most likely reason for this observation is that (pseudo)lignin helps maintain the accessibility of the material to hydrolysis, either by preventing the aggregation of cellulose microfibrils (Ishizawa *et al.*, 2009), or potentially by the breakdown of particular non-carbohydrate moieties that limit access to the remaining material to an acid-insoluble material. It is also possible that non-specific binding of cellulases to lignin increases the proportion of rapidly hydrolysed biomass by keeping cellulase and substrate in close contact, in a dilute system (1% [substrate]).

Together results provide targets for cellulase cocktail improvement for the hydrolysis of steam exploded OSR straw, which could be achieved by adding supplementary xylanase and pectinase. This may also be applicable to other dicotyledonous biomass sources, which also benefit from addition of these accessory enzymes (Zhang *et al.,* 2013).

Cellulases and accessory enzymes work synergistically to hydrolyse biomass (Hu *et al.,* 2011, Kim *et al.,* 2014). Therefore, specific combinations of accessory enzymes might be needed to improve saccharification. Cellulase cocktail optimisation could be most easily achieved by producing 96-well plates containing steam exploded OSR straw slurries. Accessory enzymes could be applied in a matrix across a plate, testing different xylanases in one dimension (rows A-H), and pectinases in the other (columns 1-12). A sub-optimal background dose of commercial cellulase could be added to all wells based on work conducted in Chapter 4 (Figure 20, pg. 93). Similarly, dilution series could be applied across a plate to identify the likely cellulase proportions needed. The resulting solutions could be analysed using the automated DNS assay developed in Chapter 6 (Section 6.4.3., pg. 128). In this way, 96 combinations of accessory enzymes could be tested simultaneously, yielding optimum cellulase cocktail compositions. Pectic domains that limit saccharification could also be inferred from the result.

Further process improvements could include optimisation at higher substrate concentrations, which will be needed for efficient cellulosic ethanol production. Batch feeding of substrate and recycling cellulases is an attractive solution (Elliston *et al.*

2013), but may be difficult to achieve with current cellulases which bind strongly to the recalcitrant pellet (Várnai *et al.* 2013). Therefore, the main limitations to saccharification efficiency may differ when processed using high substrate concentrations and multiple rounds of cellulase recycling. Work presented in this thesis showed that plate-based systems may be limited to relatively low substrate concentrations (<10% substrate, Figure 28, pg. 127), therefore larger-scale work may be required for further cellulase improvement.

9.1.2 Variations in non-cellulosic polysaccharides is the main difference between *B. napus* cultivars when processed using pilot-scale steam explosion

B. napus straw derived from various cultivars varied in composition before pretreatment (Chapter 5). These variations were mainly isolated to non-cellulosic components (UA, Man and Ara). The largest difference between cultivars after SE was the Glucan and Xylan retained in the pretreated material. Interestingly, straw that contained more glucan before pretreatment did not necessarily yield more Glc after processing. At a pilot-scale, Galactan content in the original straw correlated with Glc yields, with some indication that Arabinan content might also be important. Spectral shifts associated with variations in pectin and xylan structuring were most closely related to final saccharification and ethanol yields. Using pilot-scale methodology was useful as it illustrated that the potential of cultivar selection was not limited to bench-top methodologies (Lindedam *et al.*, 2010). However, exploring these correlations further was limited by logistical problems related to the incomplete recovery of material during pilot-scale processing (Rocha *et al.*, 2012).

This provided a strong motive to scale down activities, not only to analyse more samples, but to improve the quantitative precision in which phenotype data could be collected. Despite the differences in scale, results collected using LHW pretreatment at a smaller scale complement and corroborate these observations many ways. For example, many of the likely candidate genes identified in close proximity SNP markers associated with saccharification yields were involved in glucuronoxylan, XG and AGP synthesis and modification. Many potential candidates were also related to pectin-methylesterification between cultivars (Chapter 7). These mirror pilot-scale observations, showing correlations between Gal abundance in original straw, spectral shifts associated with xylan structuring and pectin methylation, and ethanol yields after processing (Chapter 5).

9.1.3 Optimising biomass screening methods improves output and precision

Producing phenotypic datasets that complement genomic datasets is a growing necessity and is only made feasible by the development of HT techniques. Developing methods for HT biomass analysis can also provide answers to key questions in a more timely and cost effective manner (Decker *et al.*, 2009). However, gathering phenotype information of this kind is challenging.

Here, smaller-scale, higher-throughput approaches to biomass analysis were developed to allow more cultivars to be compared. Biomass was also pretreated in a fully closed system thereby improving quantification of products. Therefore, this work contributes to the growing development of biomass analysis techniques suitable for screening biomass. Adapting existing methods specifically for the purpose of producing and analysing biomass hydrolysates allows data to be gathered to answer questions that would be logistically impossible by other methods. A key contribution of this work was to tailor the DNS method specifically for the analysis of biomass hydrolysates (Chapter 6). This provides a route for rapid determination of RS in biomass hydrolysates over a wide range of concentrations and allows other researchers to tailor the method to their own applications with greater ease.

9.1.4 The advantage of genotypic selection is greater at suboptimal processing conditions

Screening platforms typically limit processing regimes to those that yield near maximum Glc yields (Lindedam et al., 2014). This potentially limits subsequent interpretation of the results as they may only reflect the variations in straw composition relevant to specific conditions. Indeed, the most interesting genotypic variations might be those that increase saccharification yields using less expensive conditions (lower cellulase dose, or less severe pretreatment conditions). Taking observations at various levels of biomatrix opening, aimed to provide a fuller impression of how cultivar variation interacts with processing parameters resulting in the final yield. Exploration of this data shows the relative effects of screening conditions effect the outcome of HT methods (Decker *et al.*, 2009).

Cultivars that produced more Glc using one set of conditions also produced more Glc using other conditions (Figure 34, pg. 137). This partially supports the conclusions of Lindedam *et al.,* (2014), who showed that relative differences in cultivar recalcitrance were comparable between analysis methods with subtlety different process parameters. Nevertheless, some differences were observed between processing conditions, both through differences in the strengths of correlations (Figure 35, pg. 139) and associated areas of the *B. napus* transcriptome (Chapter 7) were observed. This was

particularly apparent for organic acid yields (formic and acetic acid), which did not correlate between pretreatment severities (Figure 37, pg. 141). Indeed, fermentation inhibitors produced after pretreatment at 185 °C showed much stronger associations with genotype that the same cultivars pretreated at 210 °C.

Processing cultivars at sub-optimal pretreatment conditions and sufficient cellulase (185H) showed the largest variation in saccharification yields between cultivars (Chapter 6). Indeed, OSR straw SE using high pretreatment conditions (210°C, 10 min) produced much more uniform substrates compared to the variation in the original material (Chapter 5). This shows that variations in CW composition may have greatest influence on product yields using sub-optimal pretreatment conditions, where compounds are retained in the pretreated material to varying degrees.

The datasets collected in this study clearly show that intraspecific variations in organic acids released from OSR straw following pretreatment can influence Glc yields. This confirms the previously suspected, but otherwise untested assumption that endogenous organic acids released during pretreatment catalyse depolymerisation (Hendriks & Zeeman, 2009). Cultivars that released more acetic and formic acid during pretreatment at 185°C yielded more Glc after saccharification, while those pretreated at 210°C did not (See Figure 36, pg. 123). This provides direct evidence that organic acids released from biomass catalyse the breakdown of biomass, but will only have a significant effect on Glc yields when straw is pretreated at suboptimal severities. Similarly, cultivars that produced more 2FA or HMF did not release more sugar after processing, giving additional confidence that the correlations were not associated with general increases in biomatrix opening, but rather caused by differences in pretreatment liquor pH. Exploiting these differences could increase the quality of biomass for cellulosic ethanol production using lower energy inputs.

9.1.5 **Association mapping is a useful tool to help elucidate the likely genetic determinants of biomass recalcitrance but is sensitive to environmental conditions.**

GWA studies are not only adept for the generation of testable hypotheses related to CW synthesis and deconstruction, but they also provide specific targets worthy of closer inspection. This makes them an attractive method to clarify potential routes for biomass improvement.

Here a pilot-scale GWAs study was conducted using trait data relevant to biomass recalcitrance. Trait data was collected using varying conditions in an attempt to reveal potential genes associated with biomass recalcitrance.

The design of this study provided a novel opportunity to see if areas of the transcriptome associated with biomass recalcitrance differed depending on process conditions. Indeed seventeen regions of the *B. napus* transcriptome were common to multiple saccharification traits (RS and Glc), and multiple processing conditions were highlighted. These regions are likely to harbour genes associated with the variance in biomass recalcitrance between cultivars processed using various conditions.

Key candidates likely to be involved in determining variations in recalcitrance between cultivars included: those governing UDP-Xyl/Ara provisioning and recycling (UDP-GlcA decarboxylase 2, XIFP and nucleotide sugar transporters), cellulose properties (CesLA15, RSW4, KOR2), xylan/ pectin modifications (GUX1, GUX5, IRX14 and pectin lyases) and CW acetylation (TBL family members). Common candidate genes included those involved in the endogenous hydrolysis of cellulose (at least six endo-1,4-βGs and seven βGs). Overall, associated SNP markers could pinpoint candidate genes involved in determining biomass recalcitrance and lay the foundations for further work in this area.

Nevertheless, later results also show the potential difficulties in establishing robust markers associated with biomass utilisation, particularly when using comparatively small datasets as used here (Chapter 6). To gain sufficient confidence that associated regions are related to the trait in question and identify robust markers common to multiple growth conditions, much higher replication (200-1000 cultivars) may be needed. Although some studies conducted in *Arabidopsis* and *B. napus* have successfully assigned areas of the genome related to simple traits controlled by few loci (Korte and Farlow, 2013, Harper *et al.,* 2013), associations studies typically require much higher replication to isolate statistically significant associations (Korte and Farlow, 2013). GWA studies also have the added complication that causative genetic determinants of a phenotype cannot be distinguish from non-causative relationships (Penning *et al.,* 2014).

Nevertheless, GWA studies are most powerful when used to generate specific and testable hypotheses which can be pursued in greater depth, thereby accelerating gene discovery (Lambert & Black, 2012). Therefore, candidate genes found in associated regions were collated, grouping candidates by potential mechanisms of biomass resistance. These included:

H_1 = Sugar nucleotide provisioning (particularly those producing UDP-Xyl / Ara) could influence biomass recalcitrance. Recycling of sugars and sucrose provisioning might also be important in mature, field grown agricultural residues. Exploiting natural control points, either unidirectional synthesis or transport across membranes, could help modify CW synthesis.

H_2 = Endogenous endo-1,4-βGs and βGs are likely to influence saccharification yields between cultivars.

H_3 = Genes involved in cellulose, xylan, pectin and glycoprotein biosynthesis are very likely to work in concert to determine product yields between cultivars. Particular targets might include those altering cellulose deposition, xylan and XG branching, pectin methylesterification and glycoprotein synthesis. The concerted variations in all these genes, which have a small effect on the overall genotype (Chapters 7 and 8), may significantly improve the processibility of biomass.

H_4 = Novel biomass improvement routes might involve altering biomass acetylation, provisioning of Ca^{2+} via Annexins and targeting regulatory genes.

Specific candidates required to test these mechanisms can be drawn from those in associated areas of the transcriptome. For example, CW cross-linking AGPs, AGP31 and APAP1, were located in areas of the transcriptome associated with the variation in formic acid release between cultivars after pretreatment at 210 °C. Therefore, other AGPs identified in regions associated with other AGPs such as AGP18, AGP22, Fasciclin-like AGP5 and 17 precursors might also covalently link CW structures. Together, the results from this GWA study clarify potential avenues worthy of further study and give the targets needed to pursue them in the short term.

In theory, areas of the *B. napus* transcriptome associated with bioethanol traits could provide markers with which to breed. However, it is likely that the markers identified in this study result in small changes in phenotype, which were occluded by environmental conditions. It is therefore likely that these markers could not be used for marker assisted breeding in isolation (Chapter 8). The low effect sizes of individual markers (Appendix Tables 7-14) and the small but significant differences in saccharification yields from *Arabidopsis* mutants deficient in particular candidate genes (Chapter 8) suggest that singular markers will not explain large differences in recalcitrance between cultivars. Consequently, other stratagies such as genome-wide selection, selecting germplasm with an enriched amount of favourable alleles, would be more successful for biomass improvement (Massman *et al.,* 2013).

Together, the results from Chapter 6 and 7 show the strengths of using large datasets, combined with genetic information to illuminate potential avenues worthy of closer inspection. However, the results in these chapters also show the weaknesses of GWA studies when exploring the genetic basis of traits likely to be controlled by many loci that individually control a small proportion of the overall phenotype. With increased replication, validation and stratagies for application, these results could take us a step closer to understanding what genes are the key bottlenecks to saccharification.

9.2 Research limitations and future research

9.2.1 Establish the role of agronomy in altering biomass recalcitrance

Plant CWs are incredibly responsive to changes in environment. To maintain the integrity in response to environmental stimuli, plant cells must provide additional reinforcement when needed, drawing on a limited genetic toolbox. Furthermore, plants have a limited supply of building materials to work with. Ultimately, plant fitness will be determined by how successfully it partitions these resources. Therefore, interactions between genotype and environment will be critically important determining overall saccharification potential.

In GWA studies, the effect to agronomy can be partially controlled for through randomisation and statistical correction (Oakey *et al.,* 2014). However, interactions between the environment and genetic background are likely to have a strong impact on final yields and associations seen. Indeed, the effect of agronomic conditions can never be fully removed from an association mapping exercise (Brachi *et al.*, 2011). This may explain why when comparing plants grown indoors (Chapter 8) to those collected under field conditions (Chapter 6 and 7), genotype x environment interactions may have disrupted associations between genotype and saccharification yield. Moreover, growing *Arabidopsis* plants under greenhouse conditions might underestimate the effect of particular mechanisms such as sugar nucleotide recycling, which may be important in the field (Chapter 8).

Therefore, a central question that remains unanswered is the effect of agronomy on determining biomass recalcitrance. More specific questions could include: What is the relative effect of environmental compared to genotypic variation in the field? What agronomic conditions produce more easily hydrolysable CW material? Can straw quality be improved by adding or limiting the availability of particular nutrients?

These questions could be answered using small-scale experiments growing plants in media of known composition, or as field trials, adding various amounts of supplementary nutrients to particular areas. Material produced from field trials would be more directly applicable to crop species and would establish the capacity of this

strategy for biomass improvement. However, the former approach would be more cost effective and would aid in the fundamental understanding of how plant nutrition change saccharification yields. If this approach was taken, model plant species could be used to maximise the application of the results. The interaction between genotype and environmental conditions could be pursued in greater depth by noting changes in saccharification yields between *Arabidopsis* accessions of known genotypes (Weigel and Mott, 2009) grown under different environmental conditions.

9.2.2 Explore the function of candidate genes in greater depth

In the preceding chapter (8), a selection of markers was chosen based on their association with saccharification traits. Candidates were deliberately narrowed to make a reverse-genetic study feasible, while still maintaining an overview of the potential mechanisms that could be involved. However, this study, only scratched the surface of the potential candidates that could have been explored and could not explore any candidates associated other traits, such fermentation inhibitor production.

Therefore, future work could include the exploration of the candidate genes located in areas of the transcriptome associated with traits relevant to bioethanol production. Undoubtedly, more sophisticated approaches than the ones used here would be needed to elucidate the particular roles that these candidate genes might play in altering CW recalcitrance. These could include mutants deficient in multiple genes, plants with altered gene expression.

Particularly intriguing possibilities include candidates likely to control one of the few recycling pathways yet-to-be elucidated *in planta*, namely the recovery of D-Xyl back to UDP-Xyl (Geserick & Tenhaken, 2013). Candidates revealed in this study suggest that the sugar kinase AT2G21370, could convert D-Xyl to Xyl-1-P while Xylose isomerase might recycle D-Xyl back to produce other NDP-sugars via the pentose phosphate pathway. It is likely that these recycling routes will have implications for biomass recalcitrance. Plant deficient in these genes should accumulate extra-cellulase D-Xyl, which could be easily tested. Artificially depriving the plants of photosynthetic carbon may show larger differences in CW composition and extra-cellulase Xyl between mutant and WT plants.

9.2.3 Conduct larger-scale GWA studies to establish areas of the transcriptome that are consistently associated with bioethanol processing traits

GWA mapping provide an insight into the nature of biomass recalcitrance beyond those available by other tools. The results are valuable as they allow us to observe a previously unobservable connection between the 'natural' diversity in genetic

composition between cultivars and biomass recalcitrance. However, without further corroboration from other perspectives they are subject to the same caveats as all observational data. Single GWA studies provide a single set of observations, which would benefit greatly by replication under different conditions (Brachi *et al.*, 2011). This way, areas of the genome/transcriptome that consistently associate with saccharification yields can be determined. Here partial replication was achieved by comparing similar traits to elucidate a core set of genes associated with saccharification efficiency across the conditions used. Also, some candidates appeared to be similar to those found by other recent GWA studies collecting similar saccharification traits (Penning *et al.*, 2014, Wang *et al.*, 2013c), this gives greater confidence that at least some of the associations observed here are indicative of the underlying genetic determinants of traits relevant to biomass recalcitrance.

However, fully independent replication conducted with many more cultivars may be needed to show areas of the transcriptome that consistently associate with saccharification yields in different environments. Comparing results from two GWA studies collected from different species would be particularly useful. The results presented here suggest that genetic differences in CW composition between *B. napus* cultivars had the greatest influence on straw recalcitrance. However, this may not be the case of other species. For example, initial investigations conducted using wheat straw suggest that the whole-plant properties such as leaf/stem ratios will have a large difference on saccharification potential in species such as wheat (Zhang *et al.*, 2014, Collins *et al.*, 2014). In wheat, the composition of each tissue appears to be very highly conserved and so differences in leaf/stem ratio are the main differences cultivars (Collins *et al.*, 2014). However, leaf and stem ratios are likely to be sensitive to environmental variables (Lindedam *et al.*, 2012). Capturing differences in CW compositions in these species may require diversity panels containing much a larger genetic diversity (Wu *et al.* 2014).

Only by comparing large datasets gathered under various conditions, in various species and by various methods, might we gain a fuller understanding of plant biomass and how to exploit them. Results presented in this thesis offer an initial glimpse into the likely determinants of biomass recalcitrance, and establish a solid foundation for further inquiry in this field.

9.2.4 Determine the key genetic determinants of CW polysaccharide synthesis and regulation

Here GWA mapping was used to elucidate the key bottlenecks to saccharification, for the purpose of bioethanol production. However, this is not the only potential use for biomass. Understanding the fundamental, genetic control of plant CW synthesis would

create a molecular tool-kit that could be exploited to improve plants for a range of purposes. Cell wall genes could be manipulated to improve the nutritional value of food, increase the tensile strength of timber or create organic polymer substitutes for petrochemicals.

Using association genetics could be one of the fastest ways to reveal the main genes involved in plant CW synthesis and expand our fundamental knowledge needed for their exploitation. For example, conducting reverse genetic studies on each of the circa. 2700 genes thought to be involved in CW synthesis (Wang *et al.*, 2012), is likely to take decades. Even elucidating the main enzyme families involved in pectin biosynthesis (> 67 in total) (Mohnen, 2008) would be a significant challenge. However, GWA studies have the capacity to identify multiple regions of the genome associated with these polymers rapidly, if only suitable phenotype data could be collected.

If CW material could be accurately extracted from many cultivars and quantified, patterns of genetic association related to each polymer group could be determined. Extraction and quantification of singular functional groups could show the main genes controlling the abundance of these compounds. For example, all regions of the genome associated with the relative abundances of extracted pectin between cultivars could first be ascertained. These extracts could be sequentially hydrolysed using various GHs with specific activities and the products quantified (Frankova & Fry, 2013). Analysing the amount of solubilised products in these extracts should provide more specific marker associations altering the abundance of particular polymer groups. Association patterns are also likely to show the main metabolic pathways that underpin the synthesis of these compounds.

To achieve this, a large diversity panel would be required and quantification would have to be accurate enough to resolve these associations. Moreover, quantitative methods suitable for CW extractions and quantification would need to be developed to make this feasible. Although technically challenging, this approach has many benefits over other methods. If a study of this nature was successful, it could considerably accelerate our understanding of CW synthesis and potentially open new opportunities for biomass exploitation.

10 Appendices

UI	Site	Year	Cultivar	185RSL	185RSH	185GlcL	185GlcH	210RSL	210RSH	210GlcL	210GlcH
1	KWS	2	1	9.970	27.419	4.325	16.314	14.533	25.306	10.749	19.897
1	KWS	2	1	11.907	26.778	6.519	17.010	14.473	31.527	10.952	24.091
1	KWS	2	1	11.428	28.060	6.422	17.271	14.092	31.718	10.445	25.139
1	KWS	2	1	12.572	34.599	7.202	20.012	13.288	33.667	9.838	26.292
2	KWS	2	2	10.800	27.570	6.470	14.705	13.308	26.039	10.142	19.583
2	KWS	2	2	12.904	29.115	7.836	16.662	13.409	27.499	10.344	20.945
2	KWS	2	2	10.376	25.553	5.983	15.705	12.545	25.758	8.926	21.784
2	KWS	2	2	13.181	29.888	6.422	18.620	13.067	27.069	9.635	21.889
7	ROTH	1	7	12.683	35.994	7.592	21.099	15.558	33.677	11.560	23.252
7	ROTH	1	7	15.045	32.959	8.714	21.447	16.823	30.510	11.560	23.357
7	ROTH	1	7	13.735	32.978	8.031	20.621	16.481	33.400	11.965	24.300
7	ROTH	1	7	14.953	31.791	8.080	22.187	14.634	32.669	12.269	26.397
11	KWS	2	11	11.686	30.736	5.690	12.530	16.481	30.475	11.357	18.953
11	KWS	2	11	12.516	28.361	6.666	16.749	13.569	28.966	10.445	22.204
11	KWS	2	11	12.553	28.229	5.398	16.227	14.393	30.420	10.749	22.833
11	KWS	2	11	13.162	28.945	6.275	16.706	13.750	29.313	10.344	-
15	KWS	2	15	12.203	26.194	5.398	14.487	13.710	23.705	8.622	18.010
15	KWS	2	15	13.439	30.321	7.738	16.184	13.831	30.393	8.825	24.720
15	KWS	2	15	11.261	28.531	6.178	15.749	13.409	32.529	7.406	27.865
15	KWS	2	15	12.092	22.688	5.690	13.400	12.104	24.998	9.129	-
16	ROTH	1	16	14.842	36.823	8.275	21.230	19.333	32.616	14.194	20.421
16	ROTH	1	16	15.396	35.372	8.470	23.100	19.293	36.669	14.801	24.615
16	ROTH	1	16	15.303	36.088	8.129	22.752	19.734	35.914	13.586	25.663
16	ROTH	1	16	14.897	33.751	7.885	21.012	16.280	33.456	13.181	27.131
23	KWS	2	23	13.476	32.526	5.300	17.837	17.044	27.736	11.661	21.889
23	KWS	2	23	15.063	32.206	6.568	19.229	15.497	34.186	11.560	25.034
23	KWS	2	23	13.919	32.469	8.616	19.011	16.260	31.089	12.370	25.768
23	KWS	2	23	15.691	26.363	7.446	16.184	14.754	31.864	11.560	28.180
24	KWS	2	24	11.963	34.053	6.470	18.750	16.722	32.867	11.762	23.776
24	KWS	2	24	14.547	35.428	6.275	21.708	17.445	34.675	13.181	23.776
24	KWS	2	24	14.621	35.033	6.032	20.969	18.289	38.833	12.877	26.712
24	KWS	2	24	13.568	29.850	6.422	17.706	15.517	29.831	11.864	28.704
37	KWS	2	37	13.532	31.810	7.251	16.662	17.566	29.221	12.978	21.470
37	KWS	2	37	14.916	33.468	6.861	19.098	17.284	33.326	12.471	21.889
37	KWS	2	37	15.100	31.885	6.812	18.315	17.325	31.150	-	22.204
37	KWS	2	37	17.407	34.203	7.543	20.316	16.361	30.867	12.573	22.937
40	KWS	2	40	12.812	28.418	6.714	15.966	14.513	27.505	9.635	17.066
40	KWS	2	40	14.381	32.959	6.666	19.142	12.867	26.874	9.331	19.373
40	KWS	2	40	12.646	27.475	7.007	17.489	14.574	25.129	11.661	20.316
40	KWS	2	40	13.753	26.458	5.934	15.488	13.329	28.258	9.838	21.994
42	KWS	2	42	12.959	33.336	7.982	17.706	17.204	31.555	10.851	23.357
42	KWS	2	42	13.513	30.114	7.153	18.968	15.537	34.069	11.560	24.405
42	KWS	2	42	13.126	34.599	7.251	19.620	17.184	36.036	10.445	25.139
42	KWS	2	42	14.362	29.492	7.056	17.967	15.778	34.031	10.344	25.139
43	KWS	2	43	12.996	28.795	7.495	15.053	14.955	29.626	10.243	20.631
43	KWS	2	43	13.421	27.739	7.007	16.532	14.292	30.139	10.445	22.204
43	KWS	2	43	13.181	28.587	6.275	16.358	16.280	33.481	11.155	25.454
43	KWS	2	43	12.461	23.348	6.373	13.661	14.112	31.960	10.851	25.873
47	KWS	2	47	13.255	30.981	6.909	16.836	16.260	30.417	11.357	23.042
47	KWS	2	47	13.070	26.929	6.422	16.575	14.935	33.502	12.066	23.252
47	KWS	2	47	12.572	28.474	6.519	17.141	15.798	36.076	12.674	23.986
47	KWS	2	47	13.808	26.703	7.641	16.358	13.590	28.738	12.775	27.865
48	KWS	2	48	10.394	25.063	4.276	13.443	14.875	29.240	10.445	17.905
48	KWS	2	48	11.834	24.158	6.617	13.008	16.461	34.245	12.066	21.050
48	KWS	2	48	12.424	24.893	5.007	12.791	14.634	31.272	10.547	21.155
48	KWS	2	48	11.741	21.539	5.593	13.748	10.557	24.883	10.142	25.768
49	KWS	2	49	10.357	30.340	5.593	17.010	16.381	32.307	11.458	21.679
49	KWS	2	49	12.240	27.325	6.470	17.576	14.835	33.541	11.458	22.413

Appendix Table 1, (1 of 6) – Summary of saccharification yields for each digest using various process conditions.

49	KWS	2	49	12.018	33.129	7.787	18.663	17.365	27.238	12.370	24.405
49	KWS	2	49	12.978	30.340	5.641	18.837	13.750	31.749	10.648	26.397
50	KWS	2	50	13.255	29.171	5.641	14.270	14.955	23.493	10.142	17.591
50	KWS	2	50	13.015	27.155	5.007	15.444	12.565	33.189	9.939	20.421
50	KWS	2	50	11.871	27.551	4.812	15.183	13.610	31.110	10.040	24.196
50	KWS	2	50	14.454	27.588	5.739	14.531	14.011	33.993	10.243	25.349
52	KWS	2	52	11.944	30.981	5.154	15.183	14.574	28.373	10.142	20.421
52	KWS	2	52	11.963	26.137	6.080	15.270	13.810	32.583	11.661	24.196
52	KWS	2	52	13.329	29.435	5.788	15.879	15.055	38.124	9.736	25.034
52	KWS	2	52	14.694	30.736	5.690	16.532	13.630	32.209	10.851	27.970
53	KWS	2	53	12.203	24.912	6.080	12.530	14.995	29.433	10.547	16.647
53	KWS	2	53	11.760	21.840	6.032	13.487	12.606	29.611	10.749	20.212
53	KWS	2	53	13.458	27.607	6.958	13.878	16.361	32.812	9.331	21.575
53	KWS	2	53	11.538	20.521	5.495	12.443	10.838	27.165	8.926	24.091
62	KWS	2	62	13.273	31.998	7.007	17.315	17.365	34.429	12.370	22.937
62	KWS	2	62	13.255	28.493	6.324	17.358	15.819	36.454	12.877	23.357
62	KWS	2	62	14.270	32.526	6.617	18.054	17.345	37.536	10.445	24.720
62	KWS	2	62	15.673	31.584	6.909	17.924	14.815	32.190	11.864	27.341
65	KWS	2	65	14.473	39.782	8.421	19.838	16.461	33.638	10.445	24.510
65	KWS	2	65	15.525	36.559	8.080	20.403	15.899	35.574	11.155	24.510
65	KWS	2	65	16.097	35.673	7.690	19.533	16.782	33.846	11.155	24.615
65	KWS	2	65	16.522	32.469	8.177	19.533	14.955	32.574	10.851	25.139
70	KWS	2	70	13.439	34.674	6.812	18.924	16.702	29.858	11.357	19.687
70	KWS	2	70	15.433	34.411	7.592	19.968	16.662	33.599	11.762	22.937
70	KWS	2	70	14.011	31.546	7.787	19.794	16.060	31.738	12.269	23.042
70	KWS	2	70	17.112	34.147	7.738	19.098	15.537	35.163	11.256	24.196
73	KWS	2	73	12.609	30.246	5.007	15.879	16.200	31.651	9.939	23.357
73	KWS	2	73	11.963	26.250	6.666	17.663	13.409	31.527	10.648	23.776
73	KWS	2	73	12.793	30.566	5.641	17.315	16.260	37.495	11.560	24.825
73	KWS	2	73	12.646	27.758	5.934	17.489	13.569	33.168	11.155	26.502
78	KWS	2	78	11.852	26.005	6.129	12.704	13.971	26.830	9.331	19.687
78	KWS	2	78	10.689	23.687	7.056	14.618	11.561	24.899	10.344	20.631
78	KWS	2	78	10.966	26.703	5.446	14.618	13.248	27.420	9.635	21.155
78	KWS	2	78	12.923	26.250	6.519	14.792	12.204	26.494	9.230	21.470
86	KWS	2	86	12.572	31.282	7.348	15.879	15.778	29.916	11.155	21.575
86	KWS	2	86	12.683	28.380	7.885	18.272	13.690	30.549	11.155	23.042
86	KWS	2	86	14.436	32.846	7.251	18.576	15.196	31.576	11.053	23.776
86	KWS	2	86	16.171	32.526	6.812	18.837	15.537	32.554	10.749	24.929
91	KWS	2	91	13.458	31.508	7.202	16.619	16.923	31.266	11.762	22.413
91	KWS	2	91	14.547	29.831	6.958	17.924	16.381	34.675	11.661	23.671
91	KWS	2	91	13.938	29.115	7.104	17.054	15.638	30.745	11.661	23.986
91	KWS	2	91	14.824	30.170	8.275	17.315	15.417	31.173	11.256	24.300
99	KWS	2	99	14.547	33.921	7.153	17.924	15.718	33.349	10.243	23.147
99	KWS	2	99	13.033	26.439	6.958	16.880	14.252	30.667	10.142	23.147
99	KWS	2	99	14.491	30.001	6.861	17.010	15.778	34.252	10.547	23.357
99	KWS	2	99	16.005	31.150	7.933	17.184	14.252	31.710	10.142	27.760
100	KWS	2	100	12.609	26.740	6.519	13.052	14.112	25.865	9.534	20.212
100	KWS	2	100	12.720	27.080	6.324	14.922	13.549	29.709	10.243	20.212
100	KWS	2	100	12.793	27.230	6.666	14.487	14.815	30.461	9.331	20.841
100	KWS	2	100	11.594	24.724	6.080	14.748	10.879	26.839	9.432	21.365
106	KWS	2	106	12.683	29.737	6.519	14.313	14.634	25.344	9.939	21.155
106	KWS	2	106	13.679	27.871	5.885	15.053	14.513	30.373	10.445	21.994
106	KWS	2	106	12.609	24.516	6.470	13.574	15.216	30.785	10.952	23.252
106	KWS	2	106	12.812	23.895	6.422	13.922	12.947	27.491	10.749	-
110	KWS	1	110	13.808	33.581	7.397	18.272	18.309	30.726	11.762	24.929
110	KWS	1	110	16.245	35.146	7.738	20.186	17.987	35.633	12.978	25.873
110	KWS	1	110	15.451	33.751	7.397	19.098	17.546	36.157	12.471	25.873
110	KWS	1	110	15.876	34.166	7.836	19.490	17.104	33.398	13.586	26.397
111	KWS	2	111	11.723	31.000	4.910	16.097	15.417	28.681	10.344	21.260

Appendix Table 1 continued... (2 of 6) – Summary of saccharification yields for each digest using various process conditions.

111	KWS	2	111	13.698	31.942	5.056	18.228	15.156	34.460	10.648	23.147
111	KWS	2	111	13.845	33.996	5.788	18.098	15.276	34.272	9.230	24.720
111	KWS	2	111	13.126	31.056	6.519	18.011	11.943	29.371	10.040	25.663
115	KWS	1	115	12.775	31.508	4.861	17.402	15.558	28.006	10.243	21.679
115	KWS	1	115	13.070	30.905	6.714	18.228	14.112	29.689	10.648	22.623
115	KWS	1	115	13.495	32.432	7.251	18.533	15.618	30.481	9.635	22.937
115	KWS	1	115	14.233	30.566	7.056	17.489	14.272	32.478	11.053	-
125	KWS	2	125	13.144	28.041	6.519	13.748	14.915	28.469	10.040	20.107
125	KWS	2	125	14.141	29.077	5.983	16.706	14.393	31.801	10.648	22.204
125	KWS	2	125	14.251	30.038	6.812	16.836	15.116	32.549	9.230	23.357
125	KWS	2	125	14.713	28.983	7.543	15.444	13.790	31.135	9.939	23.986
130	KWS	2	130	12.996	27.362	6.519	13.661	15.457	25.132	10.851	14.026
130	KWS	2	130	12.424	24.328	6.763	15.227	13.911	26.620	10.851	19.792
130	KWS	2	130	12.553	27.080	9.299	15.314	14.694	23.264	11.155	20.002
130	KWS	2	130	15.580	28.681	6.958	15.575	15.859	30.713	11.357	20.526
131	ROTH	1	131	15.802	34.128	8.275	19.838	20.076	35.779	13.383	24.405
131	ROTH	1	131	15.451	33.694	10.177	20.795	18.349	34.362	14.497	24.929
131	ROTH	1	131	15.303	35.108	9.201	20.577	18.289	34.150	14.194	25.244
131	ROTH	1	131	17.057	33.016	9.299	21.012	17.425	30.905	14.700	25.454
133	KWS	2	133	11.741	25.685	4.812	13.095	14.373	27.678	9.331	20.945
133	KWS	2	133	12.683	26.684	6.470	14.748	13.449	30.217	9.939	22.518
133	KWS	2	133	12.480	27.400	8.665	14.792	14.152	26.265	10.142	22.623
133	KWS	2	133	11.760	24.215	5.398	13.922	11.622	29.045	9.736	23.776
140	KWS	2	140	12.110	27.796	5.836	15.270	16.461	29.240	11.762	20.945
140	KWS	2	140	14.048	30.905	6.861	17.184	16.702	31.058	12.471	23.147
140	KWS	2	140	12.830	30.510	5.934	17.184	16.963	31.819	12.978	23.462
140	KWS	2	140	14.676	30.189	6.714	17.097	15.537	31.921	12.269	-
141	KWS	2	141	12.037	29.718	6.666	16.401	14.734	30.359	10.040	21.889
141	KWS	2	141	13.181	27.965	8.177	17.967	12.967	29.005	9.939	22.833
141	KWS	2	141	12.609	29.209	5.690	16.575	13.228	29.326	9.230	24.825
141	KWS	2	141	13.347	29.190	6.422	15.270	13.088	33.628	9.939	-
143	KWS	2	143	12.978	27.494	6.227	13.748	16.080	26.849	11.053	20.945
143	KWS	2	143	13.421	28.757	5.934	16.923	14.272	29.161	10.547	21.050
143	KWS	2	143	14.104	29.077	6.470	16.532	15.457	30.664	9.838	21.050
143	KWS	2	143	13.698	25.591	6.861	14.835	13.730	26.110	10.851	22.308
150	KWS	2	150	14.584	34.448	7.641	18.054	16.341	33.156	11.458	24.720
150	KWS	2	150	15.968	33.468	6.861	20.012	16.762	35.027	11.661	24.720
150	KWS	2	150	15.709	35.485	7.251	19.446	16.160	34.718	11.661	24.825
150	KWS	2	150	15.525	31.791	7.348	18.968	14.634	31.250	10.749	25.768
151	KWS	2	151	11.907	28.154	5.836	14.748	13.449	28.257	9.635	20.212
151	KWS	2	151	11.871	26.231	7.251	16.662	11.501	26.913	10.142	21.155
151	KWS	2	151	14.196	31.621	7.543	17.010	14.734	32.569	10.243	21.470
151	KWS	2	151	15.340	31.716	8.129	17.576	12.666	31.327	9.635	24.929
155	KWS	2	155	13.568	31.866	5.739	15.444	14.433	31.651	10.142	21.679
155	KWS	2	155	12.129	28.380	4.910	16.445	13.168	30.667	10.142	23.147
155	KWS	2	155	11.594	28.097	5.544	16.575	13.790	31.819	10.142	23.567
155	KWS	2	155	13.070	28.342	5.007	15.575	13.750	29.850	10.040	24.300
164	KWS	2	164	12.590	31.395	6.617	15.923	15.176	28.797	10.040	20.107
164	KWS	2	164	12.387	28.964	6.373	18.272	11.762	28.751	10.648	21.679
164	KWS	2	164	13.310	33.110	5.739	17.576	15.176	35.955	10.547	23.776
164	KWS	2	164	14.860	32.168	5.885	19.794	13.308	31.576	10.749	24.615
166	KWS	2	166	13.089	31.376	6.470	16.923	17.325	28.257	12.471	20.736
166	KWS	2	166	13.661	30.095	6.470	17.837	16.120	30.999	11.965	20.945
166	KWS	2	166	11.815	28.908	6.032	17.445	16.682	30.481	12.674	22.099
166	KWS	2	166	12.572	27.061	6.861	16.488	13.349	26.053	12.066	23.147
168	ROTH	1	168	15.266	36.126	8.031	20.055	20.718	32.365	14.700	25.139
168	ROTH	1	168	15.359	34.109	8.909	22.187	18.509	33.756	13.586	25.978
168	ROTH	1	168	15.599	35.089	8.519	22.100	19.132	34.819	12.269	26.292
168	ROTH	1	168	16.392	35.221	7.787	21.404	17.807	36.026	13.687	26.397

Appendix Table 1 continued... (3 of 6) – Summary of saccharification yields for each digest using various process conditions.

169	KWS	2	169	13.882	31.263	7.251	16.836	18.630	32.519	13.181	21.679
169	KWS	2	169	14.805	31.848	7.495	18.707	18.570	30.960	13.586	22.413
169	KWS	2	169	15.654	32.733	7.885	18.968	18.911	33.562	10.952	24.405
169	KWS	2	169	14.731	31.131	7.007	19.185	15.758	30.771	12.775	24.615
177	KWS	2	177	12.295	26.608	6.080	13.574	15.015	31.381	10.952	22.413
177	KWS	2	177	12.830	26.175	5.690	14.966	14.172	31.195	10.243	22.518
177	KWS	2	177	13.199	27.513	6.470	15.183	14.333	30.522	9.635	23.147
177	KWS	2	177	13.033	26.552	5.836	15.488	11.923	27.913	9.534	23.881
184	KWS	2	184	13.329	28.813	6.470	15.009	16.682	32.616	10.648	18.220
184	KWS	2	184	13.827	27.268	6.958	15.792	15.497	32.172	10.749	20.945
184	KWS	2	184	13.015	28.587	6.763	15.966	14.574	30.765	10.040	22.413
184	KWS	2	184	14.473	27.513	6.324	14.574	14.353	30.483	9.838	22.623
185	KWS	2	185	12.147	26.740	5.398	12.921	15.798	31.092	10.547	21.889
185	KWS	2	185	11.852	23.178	6.080	13.791	12.927	29.885	9.838	23.252
185	KWS	2	185	11.317	25.176	6.275	13.139	14.553	25.555	10.344	23.567
185	KWS	2	185	14.067	23.951	6.470	12.921	13.831	29.236	10.344	24.825
100A	KWS	1	100	12.350	25.798	6.080	12.834	15.758	33.156	10.851	20.841
100A	KWS	1	100	12.184	25.214	6.032	12.921	14.714	32.837	11.458	21.050
100A	KWS	1	100	12.277	27.419	7.641	14.748	16.160	25.636	11.155	21.365
100A	KWS	1	100	12.701	25.365	6.568	12.486	15.276	28.258	10.547	23.881
111A	KWS	1	111	13.162	31.414	5.641	15.488	17.867	32.326	11.661	21.155
111A	KWS	1	111	15.469	30.076	7.446	17.532	15.317	31.429	12.066	22.204
111A	KWS	1	111	14.547	32.959	6.714	18.272	16.582	35.792	12.269	24.300
111A	KWS	1	111	14.030	30.641	6.373	17.402	15.397	31.615	11.560	27.026
125A	KWS	1	125	11.280	28.663	6.861	14.661	14.172	23.609	10.445	18.744
125A	KWS	1	125	11.483	23.499	6.519	15.575	12.826	25.055	10.445	19.373
125A	KWS	1	125	13.033	29.209	8.470	15.966	15.798	27.096	11.458	19.792
125A	KWS	1	125	12.092	23.800	6.032	13.965	14.915	28.412	10.851	22.099
130B	ROTH	1	130	14.399	31.866	8.372	17.141	19.413	32.539	12.674	24.091
130B	ROTH	1	130	15.340	33.883	6.617	21.317	17.807	34.557	14.903	24.300
130B	ROTH	1	130	13.052	34.938	7.885	18.446	19.333	33.218	14.396	25.349
130B	ROTH	1	130	15.857	30.886	8.714	19.751	17.626	32.689	14.903	26.083
133A	KWS	1	133	11.963	25.006	6.617	12.660	14.935	24.650	10.749	18.220
133A	KWS	1	133	12.941	27.155	5.446	14.922	15.738	29.494	11.256	18.220
133A	KWS	1	133	11.686	25.986	6.032	14.270	15.738	27.521	10.851	20.421
133A	KWS	1	133	12.074	24.328	5.788	13.182	13.891	28.623	11.155	21.365
141A	KWS	1	141	14.731	34.561	6.763	18.837	18.831	35.702	13.079	25.034
141A	KWS	1	141	15.451	33.525	10.128	19.446	19.413	36.571	13.181	25.139
141A	KWS	1	141	14.916	33.845	7.592	19.446	18.550	36.502	11.864	25.663
141A	KWS	1	141	14.270	31.000	6.666	17.271	17.425	33.724	10.851	26.607
141B	ROTH	1	141	15.322	34.203	7.836	18.272	19.032	33.329	13.484	23.776
141B	ROTH	1	141	15.728	33.732	8.324	19.229	17.244	33.717	13.788	25.349
141B	ROTH	1	141	15.709	30.717	8.080	18.576	17.325	33.299	13.079	26.083
141B	ROTH	1	141	16.097	33.468	8.372	18.707	17.505	35.815	13.079	-
143B	ROTH	1	143	12.000	30.736	5.544	16.488	14.995	30.996	9.635	20.631
143B	ROTH	1	143	12.664	29.435	6.763	18.968	13.650	31.019	10.243	20.841
143B	ROTH	1	143	12.240	30.378	6.909	18.185	14.112	30.055	9.939	22.413
143B	ROTH	1	143	13.827	27.777	6.227	15.879	13.128	31.154	9.838	23.671
151A	KWS	1	151	14.085	32.696	7.495	15.444	17.706	32.770	11.458	21.784
151A	KWS	1	151	15.359	32.771	-	17.532	16.160	35.261	11.762	22.937
151A	KWS	1	151	12.129	26.966	7.056	16.575	15.337	28.880	11.965	25.349
151A	KWS	1	151	13.864	26.703	6.324	14.792	14.674	31.135	10.749	26.817
155A	KWS	1	155	12.793	32.469	5.690	16.227	17.686	28.604	13.181	22.833
155A	KWS	1	155	14.454	35.937	6.032	19.925	18.590	37.158	13.181	23.881
155A	KWS	1	155	14.491	35.033	5.788	19.011	19.232	31.738	12.775	24.510
155A	KWS	1	155	13.845	29.096	5.544	16.532	16.642	29.869	12.066	24.510
164A	KWS	1	164	13.975	34.090	6.568	17.315	19.293	33.291	13.079	21.994
164A	KWS	1	164	14.584	31.094	6.422	18.402	16.562	36.552	13.079	23.881
164A	KWS	1	164	14.934	35.070	6.422	18.533	17.787	38.691	11.155	26.083

Appendix Table 1 continued... (4 of 6) – Summary of saccharification yields for each digest using various process conditions.

164A	KWS	1	164	14.602	28.625	6.178	16.401	15.919	29.773	11.864	27.446
166B	ROTH	1	166	13.052	28.493	6.861	14.922	16.200	28.450	11.256	19.792
166B	ROTH	1	166	14.824	29.209	9.153	16.532	15.899	34.401	11.560	20.841
166B	ROTH	1	166	13.236	26.627	8.763	-	15.357	26.933	12.066	24.300
166B	ROTH	1	166	14.565	26.740	7.592	14.792	14.634	30.636	11.458	25.768
184A	KWS	1	184	13.624	28.455	7.690	14.966	17.867	30.996	12.269	23.252
184A	KWS	1	184	14.214	31.244	6.032	18.054	17.787	34.264	13.079	24.300
184A	KWS	1	184	13.845	31.433	7.348	18.228	18.369	33.806	11.357	24.615
184A	KWS	1	184	12.886	24.328	6.812	15.009	15.879	29.812	12.877	24.615
23A	KWS	1	23	13.679	34.467	6.617	17.967	17.907	33.677	12.573	19.268
23A	KWS	1	23	14.473	31.904	8.470	20.490	16.481	29.982	11.864	22.413
23A	KWS	1	23	14.713	33.374	7.056	18.663	15.919	31.535	11.762	24.196
23A	KWS	1	23	15.322	33.186	7.202	19.664	15.578	33.034	11.458	25.454
23B	ROTH	1	23	14.565	31.395	9.104	17.315	16.983	31.632	12.168	23.881
23B	ROTH	1	23	15.008	30.792	8.275	19.968	15.919	32.954	13.383	24.929
23B	ROTH	1	23	15.045	33.732	8.275	19.881	16.762	33.603	12.066	26.083
23B	ROTH	1	23	16.152	32.017	8.275	19.446	16.220	32.957	11.965	27.026
24A	KWS	1	24	13.236	32.903	6.470	17.271	18.068	31.439	12.573	18.115
24A	KWS	1	24	13.975	34.825	4.325	20.055	17.044	37.041	12.674	23.776
24A	KWS	1	24	12.886	32.846	7.592	18.750	16.983	31.576	12.168	27.131
24A	KWS	1	24	14.417	34.882	6.714	19.490	16.461	33.859	12.573	-
24B	ROTH	1	24	13.310	32.545	6.568	18.185	19.172	26.907	12.978	19.583
24B	ROTH	1	24	13.919	31.640	6.275	19.620	18.650	35.457	13.991	24.091
24B	ROTH	1	24	13.550	32.469	7.982	20.577	17.947	35.752	14.396	25.663
24B	ROTH	1	24	15.303	31.923	7.738	19.533	17.987	34.626	13.890	25.768
2A	KWS	1	2	11.760	24.837	6.714	12.834	15.618	24.014	10.952	17.171
2A	KWS	1	2	10.246	20.012	6.275	14.661	11.843	23.511	11.560	19.163
2A	KWS	1	2	11.594	24.196	6.470	14.574	15.256	28.089	11.458	21.155
2A	KWS	1	2	12.055	23.989	6.666	14.052	13.188	27.338	10.749	21.679
37B	ROTH	1	37	13.735	32.658	7.543	17.793	19.433	33.542	13.788	22.728
37B	ROTH	1	37	14.436	31.621	5.202	19.751	18.690	36.043	14.295	24.196
37B	ROTH	1	37	13.993	30.038	8.080	18.794	18.590	33.116	14.599	24.405
37B	ROTH	1	37	14.381	30.679	7.885	21.621	15.176	29.256	13.687	24.929
49A	KWS	1	49	14.307	31.979	7.495	17.141	19.052	33.098	14.092	24.929
49A	KWS	1	49	15.285	31.414	8.324	19.968	18.590	34.792	14.396	25.244
49A	KWS	1	49	15.469	32.394	7.933	19.446	19.915	35.468	12.168	25.873
49A	KWS	1	49	14.528	32.733	7.836	18.707	17.003	34.645	13.890	-
50A	KWS	1	50	12.406	25.986	6.032	12.704	15.256	27.620	10.445	19.373
50A	KWS	1	50	12.332	23.216	7.446	12.617	13.088	26.189	10.344	20.841
50A	KWS	1	50	12.812	24.648	6.080	12.878	14.453	27.258	10.142	22.623
50A	KWS	1	50	12.092	22.764	5.641	11.877	12.987	25.842	10.344	-
52A	KWS	1	52	10.615	23.800	4.081	12.356	15.337	28.122	10.243	21.155
52A	KWS	1	52	11.114	23.819	5.788	13.661	14.232	30.315	10.445	21.365
52A	KWS	1	52	11.391	24.328	5.934	13.704	14.393	26.852	10.040	22.204
52A	KWS	1	52	12.147	22.406	6.129	13.052	11.561	25.535	9.736	22.413
52B	ROTH	1	52	10.394	23.254	5.300	12.225	13.790	26.810	9.129	18.010
52B	ROTH	1	52	10.726	21.633	6.470	13.791	11.662	26.717	10.040	19.478
52B	ROTH	1	52	12.738	23.800	6.275	13.835	13.389	26.771	9.838	21.994
52B	ROTH	1	52	13.661	23.367	6.812	13.748	12.385	26.436	9.331	-
70A	KWS	1	70	13.107	30.905	6.568	16.314	17.044	29.298	12.674	21.470
70A	KWS	1	70	13.310	29.775	7.007	18.011	16.321	35.222	12.877	22.623
70A	KWS	1	70	14.048	30.830	6.909	17.271	17.686	29.690	13.079	22.937
70A	KWS	1	70	13.938	29.548	5.885	18.576	12.545	26.110	11.661	26.817
73A	KWS	1	73	12.904	28.154	6.812	15.966	17.606	32.519	11.458	20.002
73A	KWS	1	73	14.473	29.643	7.738	18.359	17.405	36.180	12.674	21.784
73A	KWS	1	73	14.547	31.546	8.177	18.359	18.509	35.042	9.635	24.196
73A	KWS	1	73	13.975	30.396	6.568	17.924	16.039	34.127	12.370	24.720
80A	KWS	1	80	13.495	37.049	6.373	18.663	20.477	36.300	13.687	21.050
80A	KWS	1	80	15.912	36.013	9.250	21.099	19.574	39.758	13.890	24.405

Appendix Table 1 continued... (5 of 6) – Summary of saccharification yields for each digest using various process conditions.

80A	KWS	1	80	16.023	38.331	8.226	20.751	20.919	38.712	11.965	25.978
80A	KWS	1	80	16.411	34.448	6.861	19.577	18.088	35.470	13.586	26.292
80B	ROTH	1	80	14.860	37.445	8.567	21.491	20.116	30.822	14.396	21.679
80B	ROTH	1	80	16.429	35.617	7.348	22.361	18.309	32.055	14.497	22.099
80B	ROTH	1	80	16.208	36.126	8.860	22.317	20.417	31.697	15.105	26.292
80B	ROTH	1	80	16.614	36.974	8.567	22.491	17.887	32.152	13.282	26.607
91B	ROTH	1	91	13.439	29.680	7.641	17.054	17.646	32.635	11.458	21.365
91B	ROTH	1	91	13.273	28.060	6.812	17.315	16.160	32.641	11.965	22.937
91B	ROTH	1	91	14.122	29.002	7.543	17.271	16.863	33.988	11.458	23.986
91B	ROTH	1	91	15.709	26.382	7.641	16.227	15.256	31.116	11.155	24.825
99A	KWS	1	99	14.713	35.372	7.836	19.185	19.373	33.619	13.079	23.881
99A	KWS	1	99	15.857	32.959	7.933	19.490	18.349	33.541	12.775	24.196
99A	KWS	1	99	12.978	32.319	7.007	20.142	16.823	30.420	13.687	24.510
99A	KWS	1	99	14.233	28.003	7.543	19.403	14.292	30.445	12.471	24.929

Appendix Table 1 continued... (6 of 6) – Summary of saccharification yields for each digest using various process conditions.

UI	Year	Site	Cultivar	185 °C				210 °C			
				Acetic	Formic	2FA	HMF	Acetic	Formic	2FA	HMF
1	2	KWS	1	0.8149	0.7222	0.0144	0.0049	1.5328	3.5647	0.2449	0.0330
1	2	KWS	1	0.7984	0.6680	0.0153	0.0047	1.5250	3.6652	0.2435	0.0330
2	2	KWS	2	0.7903	0.7109	0.0104	0.0036	1.2920	2.9935	0.1563	0.0240
2	2	KWS	2	0.8014	0.7448	0.0107	0.0039	1.2746	2.8140	0.1559	0.0236
7	1	ROTH	7	0.7583	0.7403	0.0115	0.0041	1.6264	4.5253	0.1611	0.0259
7	1	ROTH	7	0.7547	0.7324	0.0111	0.0040	1.6048	4.4565	0.1621	0.0271
11	2	KWS	11	0.6932	0.4660	0.0041	0.0018	1.6435	4.7285	0.1653	0.0274
11	2	KWS	11	0.6977	0.4671	0.0040	0.0018	1.6570	4.7918	0.1651	0.0267
15	2	KWS	15	0.8665	0.6996	0.0055	0.0019	1.2057	2.8964	0.0726	0.0129
15	2	KWS	15	0.8617	0.6917	0.0057	0.0021	1.1952	2.8275	0.0717	0.0123
16	1	ROTH	16	0.8092	0.9988	0.0129	0.0061	1.3553	3.8525	0.1452	0.0237
16	1	ROTH	16	0.8098	0.9999	0.0131	0.0062	1.2848	3.5252	0.1453	0.0245
23	2	KWS	23	0.8797	0.7911	0.0077	0.0026	1.1340	2.3862	0.0855	0.0139
23	2	KWS	23	0.8794	0.7877	0.0079	0.0025	1.0884	1.9978	0.0847	0.0125
24	2	KWS	24	0.8098	0.7945	0.0109	0.0044	1.5760	4.4452	0.1671	0.0262
24	2	KWS	24	0.8137	0.8193	0.0104	0.0042	1.4429	3.7984	0.1662	0.0261
37	2	KWS	37	1.0003	1.2099	0.0066	0.0031	1.8552	8.3420	0.0969	0.0227
37	2	KWS	37	0.9964	1.1918	0.0068	0.0031	1.7361	7.3780	0.0941	0.0210
40	2	KWS	40	0.8578	0.6556	0.0076	0.0039	1.3661	4.4870	0.1007	0.0222
40	2	KWS	40	0.8128	0.7956	0.0078	0.0041	0.9250	1.7269	0.0968	0.0215
42	2	KWS	42	0.7927	0.6037	0.0079	0.0030	1.6444	4.7037	0.1630	0.0276
42	2	KWS	42	0.7918	0.6026	0.0080	0.0029	1.6486	4.7635	0.1604	0.0268
43	2	KWS	43	0.8329	0.7369	0.0086	0.0036	1.7577	4.9758	0.1919	0.0317
43	2	KWS	43	0.8335	0.6026	0.0088	0.0036	1.6969	4.5886	0.1888	0.0309
47	2	KWS	47	0.8623	0.6929	0.0123	0.0044	1.3958	3.4800	0.1359	0.0219
47	2	KWS	47	0.8170	0.6951	0.0126	0.0046	1.2237	2.1039	0.1342	0.0212
48	2	KWS	48	0.7846	0.7967	0.0117	0.0055	1.2000	2.3478	0.1391	0.0242
48	2	KWS	48	0.7807	0.7832	0.0119	0.0057	1.2198	2.4245	0.1387	0.0243
49	2	KWS	49	0.7433	0.5687	0.0106	0.0039	0.7262	0.1194	0.0889	0.0165
49	2	KWS	49	0.7409	0.5574	0.0104	0.0040	1.0941	2.3116	0.0888	0.0166
50	2	KWS	50	0.7181	0.4456	0.0054	0.0022	1.4054	3.8830	0.1127	0.0182
50	2	KWS	50	0.7217	0.4479	0.0056	0.0023	1.3358	3.6211	0.1128	0.0183
52	2	KWS	52	0.6890	0.3723	0.0034	0.0012	1.8585	6.2175	0.1253	0.0217
52	2	KWS	52	0.6818	0.3599	0.0034	0.0015	1.5823	4.4102	0.1248	0.0216
53	2	KWS	53	0.7340	0.5901	0.0077	0.0036	1.5208	4.7387	0.1534	0.0285
53	2	KWS	53	0.7340	0.5901	0.0080	0.0036	1.5253	4.7500	0.1507	0.0285
62	2	KWS	62	0.7855	0.7143	0.0106	0.0036	1.6750	5.0706	0.1978	0.0277
62	2	KWS	62	0.7888	0.7166	0.0115	0.0038	1.4081	3.4405	0.1989	0.0268
65	2	KWS	65	0.9016	0.7414	0.0047	0.0018	1.3631	3.8266	0.0600	0.0129
65	2	KWS	65	0.9010	0.7403	0.0045	0.0017	1.3394	3.6618	0.0586	0.0129
70	2	KWS	70	0.8092	0.6466	0.0053	0.0020	1.5436	4.2047	0.0958	0.0178
70	2	KWS	70	0.8101	0.6421	0.0053	0.0020	1.5001	3.8808	0.0943	0.0175
73	2	KWS	73	0.8287	0.7651	0.0106	0.0044	1.2261	2.6345	0.1106	0.0187
73	2	KWS	73	0.8296	0.7651	0.0108	0.0043	1.2336	2.7045	0.1108	0.0182
78	2	KWS	78	0.7457	0.5856	0.0071	0.0032	1.3769	4.1551	0.1302	0.0216
78	2	KWS	78	0.7463	0.5879	0.0072	0.0032	1.4219	4.4508	0.1298	0.0221
86	2	KWS	86	0.8125	0.6443	0.0056	0.0021	1.3004	2.7384	0.0836	0.0150
86	2	KWS	86	0.7951	0.5890	0.0053	0.0023	1.3631	3.1278	0.0830	0.0147
91	2	KWS	91	0.7846	0.7820	0.0080	0.0035	1.5319	5.4758	0.0913	0.0170
91	2	KWS	91	0.7810	0.7753	0.0079	0.0039	1.2207	3.1786	0.0911	0.0173
99	2	KWS	99	0.8521	0.6929	0.0057	0.0024	1.5241	3.4518	0.0932	0.0179

Appendix Table 2, (1 of 3) – Summary of organic acid yields for each accession using various process conditions.

99	2	KWS	99	0.8551	0.7008	0.0058	0.0021	1.6057	3.8243	0.0924	0.0174
100	2	KWS	100	0.7735	0.6105	0.0052	0.0020	1.2770	3.5184	0.0860	0.0152
100	2	KWS	100	0.7732	0.6082	0.0053	0.0020	1.2114	3.0680	0.0859	0.0149
106	2	KWS	106	0.7801	0.6398	0.0062	0.0024	1.1694	2.8354	0.0832	0.0141
106	2	KWS	106	0.7801	0.6432	0.0059	0.0021	1.1304	2.5916	0.0809	0.0146
110	1	KWS	110	0.7421	0.6195	0.0060	0.0036	1.2333	2.8264	0.1072	0.0205
110	1	KWS	110	0.7403	0.6138	0.0061	0.0037	1.1667	2.3862	0.1065	0.0199
111	2	KWS	111	0.7277	0.4366	0.0084	0.0028	1.4006	3.4405	0.1949	0.0262
111	2	KWS	111	0.7340	0.4276	0.0084	0.0031	1.4735	3.9270	0.1915	0.0259
115	1	KWS	115	0.7762	0.8532	0.0103	0.0069	1.3472	3.7826	0.1353	0.0276
115	1	KWS	115	0.7750	0.8441	0.0104	0.0069	1.3658	3.9327	0.1333	0.0277
125	2	KWS	125	0.7921	0.7493	0.0094	0.0056	1.1262	2.4121	0.1255	0.0209
125	2	KWS	125	0.7936	0.7516	0.0094	0.0041	1.1253	2.4166	0.1278	0.0209
130	2	KWS	130	0.7571	0.6421	0.0066	0.0044	1.2333	4.1302	0.0819	0.0202
130	2	KWS	130	0.7589	0.6714	0.0066	0.0045	1.2246	4.0557	0.0807	0.0203
131	1	ROTH	131	0.8815	1.2618	0.0108	0.0067	1.6567	6.4309	0.1696	0.0323
131	1	ROTH	131	0.8749	1.2381	0.0108	0.0068	1.6462	6.3857	0.1676	0.0330
133	2	KWS	133	0.7091	0.5202	0.0057	0.0033	1.1064	1.9549	0.1005	0.0206
133	2	KWS	133	0.7001	0.5168	0.0057	0.0034	1.1901	2.5905	0.1036	0.0206
140	2	KWS	140	0.7382	0.6703	0.0074	0.0046	1.2057	3.3694	0.1083	0.0218
140	2	KWS	140	0.7445	0.6951	0.0075	0.0045	1.1598	3.0364	0.1057	0.0217
141	2	KWS	141	0.6236	0.1307	0.0097	0.0040	1.1847	2.7327	0.0928	0.0174
141	2	KWS	141	0.7693	0.6409	0.0091	0.0041	1.2390	3.1459	0.0937	0.0176
143	2	KWS	143	0.7939	0.7719	0.0074	0.0035	1.1028	2.8716	0.0899	0.0159
143	2	KWS	143	0.7891	0.7595	0.0076	0.0034	1.1088	2.8693	0.0868	0.0160
150	2	KWS	150	0.7607	0.5721	0.0055	0.0023	1.3112	3.5477	0.1054	0.0180
150	2	KWS	150	0.7538	0.5461	0.0054	0.0025	1.2881	3.4247	0.1054	0.0180
151	2	KWS	151	0.7705	0.5258	0.0048	0.0020	1.1103	2.1299	0.0745	0.0131
151	2	KWS	151	0.7699	0.5156	0.0048	0.0019	1.1145	2.1141	0.0744	0.0125
155	2	KWS	155	0.7043	0.5540	0.0108	0.0057	1.3589	2.9709	0.2004	0.0346
155	2	KWS	155	0.7004	0.5314	0.0110	0.0061	1.2467	2.6424	0.1996	0.0348
164	2	KWS	164	0.8347	0.7380	0.0084	0.0040	1.2833	3.3863	0.1195	0.0208
164	2	KWS	164	0.8392	0.7572	0.0082	0.0038	1.2893	3.3773	0.1194	0.0209
166	2	KWS	166	0.8863	1.0349	0.0108	0.0057	1.4642	5.4002	0.1398	0.0268
166	2	KWS	166	0.8866	1.0270	0.0112	0.0056	1.4567	5.3302	0.1386	0.0268
168	1	ROTH	168	0.8092	1.1354	0.0128	0.0103	1.1088	3.2046	0.1152	0.0267
168	1	ROTH	168	0.7987	1.0993	0.0127	0.0105	1.1232	3.3017	0.1148	0.0265
169	2	KWS	169	0.8332	0.7843	0.0099	0.0047	1.4711	5.4871	0.1511	0.0249
169	2	KWS	169	0.8302	0.7741	0.0098	0.0048	1.4657	5.3878	0.1502	0.0247
177	2	KWS	177	0.7319	0.5834	0.0073	0.0031	1.2264	2.8580	0.1191	0.0207
177	2	KWS	177	0.7301	0.5721	0.0074	0.0030	1.2123	2.7372	0.1182	0.0203
184	2	KWS	184	0.8320	0.5710	0.0061	0.0027	1.3901	3.6539	0.0965	0.0215
184	2	KWS	184	0.8014	0.6105	0.0056	0.0025	1.3928	3.6674	0.0938	0.0207
185	2	KWS	185	0.7136	0.5156	0.0075	0.0046	1.5571	4.2792	0.1890	0.0349
185	2	KWS	185	0.7178	0.5472	0.0076	0.0043	1.5751	4.4903	0.1881	0.0370
100A	1	KWS	100	0.7238	0.7008	0.0065	0.0044	1.3685	4.7184	0.1184	0.0237
100A	1	KWS	100	0.7406	0.7200	0.0066	0.0035	1.3706	4.7195	0.1184	0.0237
111A	1	KWS	111	0.7238	0.6996	0.0053	0.0032	1.4069	3.3513	0.1450	0.0269
111A	1	KWS	111	0.7262	0.7109	0.0054	0.0032	1.4129	3.3716	0.1452	0.0253
125A	1	KWS	125	0.7804	0.8024	0.0090	0.0085	1.3802	4.1833	0.1170	0.0306
125A	1	KWS	125	0.7816	0.8058	0.0089	0.0085	1.3883	4.2160	0.1162	0.0307
130B	1	ROTH	130	0.8164	1.0338	0.0129	0.0209	1.4771	5.3257	0.1788	0.0597
130B	1	ROTH	130	0.8083	1.0078	0.0131	0.0209	1.4792	5.2941	0.1830	0.0605

Appendix Table 2, (2 of 3) – Summary of organic acid yields for each accession using various process conditions.

133A	1	KWS	133	0.7897	0.8746	0.0112	0.0133	1.3865	4.3842	0.1354	0.0405
133A	1	KWS	133	0.7900	0.8803	0.0108	0.0134	1.3790	4.3515	0.1345	0.0401
141A	1	KWS	141	0.9112	1.1455	0.0122	0.0071	1.4851	4.2465	0.1202	0.0258
141A	1	KWS	141	0.9328	1.1354	0.0124	0.0070	1.5313	4.5874	0.1197	0.0258
141B	1	ROTH	141	0.8578	0.9311	0.0093	0.0057	1.3499	3.3581	0.1123	0.0224
141B	1	ROTH	141	0.8356	0.9322	0.0092	0.0051	1.3136	3.1346	0.1119	0.0223
143B	1	ROTH	143	0.7568	0.6590	0.0084	0.0046	1.2836	3.1910	0.0884	0.0199
143B	1	ROTH	143	0.7690	0.7030	0.0084	0.0041	1.3226	3.3378	0.0865	0.0197
151A	1	KWS	151	0.7735	0.6376	0.0059	0.0027	1.1712	2.9619	0.0638	0.0151
151A	1	KWS	151	0.7873	0.6353	0.0058	0.0027	1.2857	3.8277	0.0631	0.0143
155A	1	KWS	155	0.7517	0.8182	0.0123	0.0094	1.1124	0.8893	0.1474	0.0299
155A	1	KWS	155	0.7607	0.8554	0.0124	0.0092	1.1043	0.8836	0.1457	0.0300
164A	1	KWS	164	0.8752	0.8882	0.0104	0.0043	1.4842	3.6110	0.1271	0.0230
164A	1	KWS	164	0.8746	0.8848	0.0101	0.0044	1.2971	2.5036	0.1268	0.0225
166B	1	ROTH	166	0.8236	1.0327	0.0134	0.0072	1.3133	3.6087	0.0959	0.0251
166B	1	ROTH	166	0.8281	1.0372	0.0136	0.0075	1.3055	3.5218	0.0952	0.0253
184A	1	KWS	184	0.8290	1.0699	0.0097	0.0055	1.4717	4.3063	0.1316	0.0288
184A	1	KWS	184	0.8350	1.0812	0.0098	0.0058	1.4111	3.9903	0.1303	0.0290
23A	1	KWS	23	0.8236	0.9333	0.0125	0.0070				
23A	1	KWS	23	0.8278	0.9480	0.0126	0.0073				
23B	1	ROTH	23	0.7864	0.8306	0.0093	0.0055	1.4300	4.6168	0.1604	0.0286
23B	1	ROTH	23	0.8224	0.9514	0.0094	0.0049	1.3712	4.2149	0.1598	0.0292
24A	1	KWS	24	0.8254	1.0868	0.0092	0.0071	1.1424	2.7429	0.0836	0.0205
24A	1	KWS	24	0.8344	1.1083	0.0092	0.0071	1.1373	2.8445	0.0855	0.0197
24B	1	ROTH	24	0.7951	0.9683	0.0122	0.0098	1.4714	4.3267	0.1493	0.0351
24B	1	ROTH	24	0.8332	1.0914	0.0124	0.0098	1.3523	3.5839	0.1476	0.0359
2A	1	KWS	2	0.8251	1.1670	0.0121	0.0084	1.3112	3.5985	0.1495	0.0270
2A	1	KWS	2	0.8248	1.1670	0.0123	0.0093	1.3580	3.9835	0.1480	0.0272
37B	1	ROTH	37	0.7732	0.9006	0.0121	0.0084	1.1571	3.1481	0.0979	0.0247
37B	1	ROTH	37	0.7741	0.9130	0.0121	0.0079	1.1616	3.2023	0.0935	0.0234
49A	1	KWS	49	0.8005	0.9074	0.0110	0.0065	1.3991	3.4123	0.1528	0.0287
49A	1	KWS	49	0.8011	0.9107	0.0110	0.0069	1.3997	3.4123	0.1514	0.0284
50A	1	KWS	50	0.7373	0.6804	0.0070	0.0052	1.4342	4.3312	0.1307	0.0325
50A	1	KWS	50	0.7352	0.6850	0.0068	0.0055	1.4258	4.2533	0.1307	0.0323
52A	1	KWS	52	0.7121	0.5992	0.0067	0.0042	1.1286	2.9720	0.0825	0.0202
52A	1	KWS	52	0.7160	0.6116	0.0065	0.0041	1.1292	2.9540	0.0860	0.0203
52B	1	ROTH	52	0.7517	0.8103	0.0097	0.0137	1.3184	4.8437	0.1153	0.0412
52B	1	ROTH	52	0.7493	0.8024	0.0099	0.0141	1.3124	4.8369	0.1138	0.0415
70A	1	KWS	70	0.8257	0.9085	0.0073	0.0056	1.4195	4.8990	0.1263	0.0258
70A	1	KWS	70	0.8311	0.9288	0.0073	0.0055	1.4168	4.9125	0.1258	0.0254
73A	1	KWS	73	0.7562	0.7369	0.0070	0.0031	1.3118	2.9291	0.0941	0.0195
73A	1	KWS	73	0.7319	0.6556	0.0070	0.0033	1.3130	2.9303	0.0938	0.0198
80A	1	KWS	80	0.8065	0.7990	0.0081	0.0043	1.1580	2.6492	0.0865	0.0165
80A	1	KWS	80	0.8395	0.9074	0.0080	0.0040	1.1493	2.6040	0.0856	0.0164
80B	1	ROTH	80	0.7933	0.9277	0.0103	0.0102	1.1265	2.7734	0.0794	0.0225
80B	1	ROTH	80	0.7837	0.8927	0.0106	0.0104	1.1037	2.6706	0.0793	0.0224
91B	1	ROTH	91	0.7475	0.7866	0.0098	0.0080	1.4498	4.1923	0.1422	0.0367
91B	1	ROTH	91	0.7472	0.7741	0.0098	0.0086	1.4510	4.1844	0.1414	0.0374
99A	1	KWS	99	0.7912	0.8599	0.0071	0.0045	1.3061	4.0365	0.0783	0.0199
99A	1	KWS	99	0.7900	0.8566	0.0069	0.0048	1.3391	4.2713	0.0780	0.0199

Appendix Table 2, (3 of 3) – Summary of organic acid yields for each accession using various process conditions.

IBTI Code	Cultivar Name	Glucose yield (% original material)			
		Pretreated at 185°C		Pretreated at 210°C	
		Low Cellulase	High Cellulase	Low Cellulase	High Cellulase
1	Aberdeenshire Prize	6.77 ± 0.32	18.06 ± 0.62	11.21 ± 0.37	23.88 ± 0.78
2	Abukuma Natane	6.98 ± 0.26	15.81 ± 0.46	10.61 ± 0.28	21.66 ± 0.65
7	Apex-93_5 X Ginyou_3 line	7.46 ± 0.32	20.58 ± 0.61	11.22 ± 0.37	23.62 ± 0.76
11	Bienvenu	6.69 ± 0.32	16.14 ± 0.62	11.41 ± 0.37	22.70 ± 0.83
15	Brauner Schnittkohl	6.86 ± 0.32	15.59 ± 0.62	9.49 ± 0.37	23.67 ± 0.83
16	Bravour	7.51 ± 0.32	21.21 ± 0.61	13.04 ± 0.37	23.69 ± 0.76
23	Canard	7.59 ± 0.21	18.95 ± 0.36	11.84 ± 0.23	24.33 ± 0.54
24	Canberra x Courage	6.70 ± 0.21	19.51 ± 0.36	12.67 ± 0.23	24.12 ± 0.56
37	Columbus X Nickel	7.14 ± 0.24	18.97 ± 0.43	13.24 ± 0.29	23.19 ± 0.62
40	Couve nabica	7.09 ± 0.32	17.48 ± 0.62	10.88 ± 0.37	21.74 ± 0.78
42	Darmor	7.63 ± 0.32	18.90 ± 0.62	11.47 ± 0.37	24.22 ± 0.78
43	Devon Champion	7.23 ± 0.32	16.00 ± 0.62	11.37 ± 0.37	23.72 ± 0.78
47	Doral	7.29 ± 0.32	17.21 ± 0.62	12.70 ± 0.37	24.23 ± 0.78
48	Drakkar	6.26 ± 0.32	14.02 ± 0.62	11.47 ± 0.37	22.66 ± 0.78
49	Duplo	7.41 ± 0.26	18.87 ± 0.46	12.55 ± 0.28	24.33 ± 0.67
50	Dwarf Essex	6.32 ± 0.26	14.34 ± 0.46	10.36 ± 0.28	22.43 ± 0.67
52	English Giant	6.02 ± 0.21	14.38 ± 0.36	10.00 ± 0.23	22.63 ± 0.56
53	Erglu	6.79 ± 0.32	13.87 ± 0.62	10.69 ± 0.37	22.23 ± 0.78
62	Fortin Family	7.18 ± 0.32	18.07 ± 0.62	12.41 ± 0.37	24.26 ± 0.78
65	Groene Groninger Snijmoes	8.13 ± 0.32	20.06 ± 0.62	11.56 ± 0.37	24.31 ± 0.78
70	Hansen x Gaspard	7.33 ± 0.26	18.94 ± 0.46	12.13 ± 0.28	23.38 ± 0.65
73	Huguenot	6.95 ± 0.26	17.86 ± 0.46	11.27 ± 0.28	23.84 ± 0.65
78	Jaune a Collet Vert	6.89 ± 0.32	14.88 ± 0.62	10.47 ± 0.37	22.28 ± 0.78
80	Judzae	7.72 ± 0.25	20.98 ± 0.44	13.08 ± 0.28	23.96 ± 0.63
86	Lembkes Malchower	7.60 ± 0.32	18.28 ± 0.62	11.67 ± 0.37	23.61 ± 0.78
91	Licrown x Express	7.35 ± 0.24	17.11 ± 0.43	11.55 ± 0.27	23.42 ± 0.62
99	Madrigal x Recital	7.63 ± 0.26	18.85 ± 0.46	11.69 ± 0.28	24.33 ± 0.65
100	Major	6.89 ± 0.26	14.43 ± 0.46	10.47 ± 0.28	22.20 ± 0.65
106	Moana, Moana Rape	6.91 ± 0.32	14.91 ± 0.62	11.23 ± 0.37	23.05 ± 0.83
110	N01D-1330	7.58 ± 0.33	19.62 ± 0.63	12.08 ± 0.39	24.78 ± 0.79
111	N02D-1952	6.53 ± 0.26	17.89 ± 0.46	11.08 ± 0.28	23.87 ± 0.65
115	Ningyou 7	6.81 ± 0.33	18.39 ± 0.63	10.09 ± 0.39	23.11 ± 0.84
125	POH 285, Bolko	7.17 ± 0.26	15.94 ± 0.46	10.53 ± 0.28	22.19 ± 0.65
130	Quinta	7.54 ± 0.24	17.07 ± 0.43	12.56 ± 0.27	22.29 ± 0.62
131	Rafal	8.24 ± 0.32	19.87 ± 0.61	13.26 ± 0.37	23.97 ± 0.76
133	Ramses	6.61 ± 0.26	14.59 ± 0.46	10.54 ± 0.28	22.06 ± 0.65
140	Sarepta	6.92 ± 0.32	17.17 ± 0.62	12.83 ± 0.37	23.22 ± 0.83
141	Sensation NZ	7.56 ± 0.21	18.16 ± 0.36	11.61 ± 0.23	24.45 ± 0.58
143	Shannon x Winner	6.50 ± 0.24	16.49 ± 0.43	10.34 ± 0.27	22.19 ± 0.62
150	Slapska Slapy	7.57 ± 0.32	19.41 ± 0.62	11.98 ± 0.37	24.47 ± 0.78
151	Slovenska Krajova	7.39 ± 0.27	16.83 ± 0.46	10.82 ± 0.28	23.46 ± 0.65
155	Stellar	6.10 ± 0.26	17.48 ± 0.46	11.52 ± 0.28	23.78 ± 0.65
164	Taisetsu	6.71 ± 0.26	18.25 ± 0.46	11.47 ± 0.28	23.88 ± 0.65
166	Tapidor	7.25 ± 0.24	16.53 ± 0.46	11.91 ± 0.27	22.59 ± 0.62
168	Temple	7.60 ± 0.32	20.67 ± 0.61	12.71 ± 0.37	24.46 ± 0.76
169	Tequila x Aragon	7.66 ± 0.32	18.77 ± 0.62	13.05 ± 0.37	23.58 ± 0.78
177	Victor	6.70 ± 0.32	15.45 ± 0.62	10.86 ± 0.37	23.44 ± 0.78
184	York	7.14 ± 0.26	16.51 ± 0.46	11.43 ± 0.28	23.15 ± 0.65
185	Yudal	6.73 ± 0.32	13.97 ± 0.62	11.02 ± 0.37	23.64 ± 0.78

Appendix Table 3 – Summary of glucose yields released for each cultivar under various process conditions, controlling for differences between years and sites.

IBTI Code	Cultivar Name	Reducing sugar yield (% original material)			
		Pretreated at 185°C		Pretreated at 210°C	
		Low Cellulase	High Cellulase	Low Cellulase	High Cellulase
1	Aberdeenshire Prize	12.25 ± 0.47	29.44 ± 1.07	15.00 ± 0.61	31.18 ± 1.11
2	Abukuma Natane	12.09 ± 0.36	26.06 ± 0.80	13.66 ± 0.47	27.36 ± 0.88
7	Apex-93_5 X Ginyou_3 line	13.75 ± 0.46	32.90 ± 1.06	15.28 ± 0.61	31.83 ± 1.09
11	Bienvenu	13.03 ± 0.47	29.31 ± 1.07	15.35 ± 0.61	30.67 ± 1.11
15	Brauner Schnittkohl	12.86 ± 0.47	27.43 ± 1.07	14.35 ± 0.61	29.40 ± 1.11
16	Bravour	14.54 ± 0.46	34.73 ± 1.06	17.47 ± 0.61	33.24 ± 1.09
23	Canard	14.71 ± 0.29	32.00 ± 0.64	15.92 ± 0.38	31.91 ± 0.72
24	Canberra x Courage	13.81 ± 0.29	33.12 ± 0.64	17.06 ± 0.38	33.25 ± 0.72
37	Columbus X Nickel	14.57 ± 0.34	31.92 ± 0.76	17.33 ± 0.45	31.93 ± 0.84
40	Couve nabica	13.75 ± 0.47	29.10 ± 1.07	14.78 ± 0.61	28.74 ± 1.11
42	Darmor	13.82 ± 0.47	31.80 ± 1.07	16.83 ± 0.61	33.45 ± 1.11
43	Devon Champion	13.45 ± 0.47	27.59 ± 1.07	15.64 ± 0.61	31.69 ± 1.11
47	Doral	13.58 ± 0.47	28.61 ± 1.07	15.82 ± 0.61	32.28 ± 1.11
48	Drakkar	12.35 ± 0.47	24.76 ± 1.07	15.03 ± 0.61	30.75 ± 1.11
49	Duplo	13.65 ± 0.36	31.28 ± 0.80	16.81 ± 0.47	32.75 ± 0.88
50	Dwarf Essex	13.10 ± 0.36	26.40 ± 0.80	13.95 ± 0.47	29.31 ± 0.88
52	English Giant	12.24 ± 0.29	25.55 ± 0.64	13.51 ± 0.38	29.37 ± 0.72
53	Erglu	12.85 ± 0.47	24.58 ± 1.07	14.69 ± 0.61	30.64 ± 1.11
62	Fortin Family	14.31 ± 0.47	31.15 ± 1.07	16.76 ± 0.61	34.28 ± 1.11
65	Groene Groninger Snijmoes	15.50 ± 0.47	35.55 ± 1.07	16.51 ± 0.61	33.44 ± 1.11
70	Hansen x Gaspard	14.44 ± 0.36	32.00 ± 0.80	15.89 ± 0.47	31.52 ± 0.88
73	Huguenot	13.51 ± 0.36	29.51 ± 0.80	15.94 ± 0.47	33.64 ± 0.88
78	Jaune a Collet Vert	12.36 ± 0.47	26.30 ± 1.07	13.94 ± 0.61	28.39 ± 1.11
80	Judzae	15.39 ± 0.35	36.10 ± 0.78	18.26 ± 0.46	33.75 ± 0.86
86	Lembkes Malchower	14.19 ± 0.47	31.25 ± 1.07	15.75 ± 0.61	31.58 ± 1.11
91	Licrown x Express	14.11 ± 0.34	29.27 ± 0.76	16.22 ± 0.45	32.11 ± 0.84
99	Madrigal x Recital	14.59 ± 0.36	31.34 ± 0.80	15.92 ± 0.47	32.26 ± 0.88
100	Major	12.78 ± 0.36	26.58 ± 0.80	14.43 ± 0.47	29.72 ± 0.88
106	Moana, Moana Rape	13.40 ± 0.47	27.05 ± 1.07	15.18 ± 0.61	29.80 ± 1.11
110	N01D-1330	15.08 ± 0.48	33.82 ± 1.10	16.51 ± 0.63	33.08 ± 1.14
111	N02D-1952	13.91 ± 0.36	31.68 ± 0.80	15.28 ± 0.47	32.26 ± 0.88
115	Ningyou 7	13.56 ± 0.48	31.33 ± 1.10	14.27 ± 0.63	30.51 ± 1.14
125	POH 285, Bolko	13.31 ± 0.36	27.95 ± 0.80	14.50 ± 0.47	29.25 ± 0.88
130	Quinta	13.99 ± 0.34	29.89 ± 0.76	16.64 ± 0.45	30.14 ± 0.84
131	Rafal	15.15 ± 0.46	33.39 ± 1.06	17.37 ± 0.61	32.66 ± 1.09
133	Ramses	12.57 ± 0.36	26.21 ± 0.80	14.28 ± 0.47	28.79 ± 0.88
140	Sarepta	13.76 ± 0.47	30.00 ± 1.07	16.82 ± 0.61	31.49 ± 1.11
141	Sensation NZ	14.42 ± 0.29	31.74 ± 0.64	16.23 ± 0.38	33.11 ± 0.72
143	Shannon x Winner	13.20 ± 0.34	28.74 ± 0.76	14.58 ± 0.45	29.87 ± 0.84
150	Slapska Slapy	15.34 ± 0.47	33.49 ± 1.07	16.47 ± 0.61	33.20 ± 1.11
151	Slovenska Krajova	13.82 ± 0.36	29.78 ± 0.80	14.54 ± 0.47	31.16 ± 0.88
155	Stellar	13.51 ± 0.36	31.23 ± 0.80	15.75 ± 0.47	31.59 ± 0.88
164	Taisetsu	14.09 ± 0.36	31.85 ± 0.80	15.50 ± 0.47	32.80 ± 0.88
166	Tapidor	13.40 ± 0.34	28.65 ± 0.76	15.70 ± 0.45	29.89 ± 0.84
168	Temple	14.96 ± 0.46	34.40 ± 1.06	17.77 ± 0.61	32.96 ± 1.09
169	Tequila x Aragon	14.81 ± 0.47	31.68 ± 1.07	18.04 ± 0.61	32.13 ± 1.11
177	Victor	13.31 ± 0.47	27.23 ± 1.07	14.81 ± 0.61	30.98 ± 1.11
184	York	13.87 ± 0.36	28.70 ± 0.80	16.16 ± 0.47	31.95 ± 0.88
185	Yudal	12.93 ± 0.47	25.51 ± 1.07	15.14 ± 0.61	30.09 ± 1.11

Appendix Table 4 – Summary of reducing sugar yields released for each cultivar under various process conditions, controlling for differences between years and sites.

IBTI Code	Cultivar Name	Fermentation inhibitor yields (% original material)											
		Pretreated at 185°C											
		Acetic			Formic			2FA			HMF		
1	Aberdeenshire Prize	0.80	±	0.02	0.79	±	0.07	0.015	±	0.001	0.007	±	0.001
2	Abukuma Natane	0.80	±	0.02	0.91	±	0.05	0.012	±	0.001	0.007	±	0.001
7	Apex-93_5 X Ginyou_3 line	0.77	±	0.02	0.68	±	0.07	0.010	±	0.001	0.003	±	0.001
11	Bienvenu	0.72	±	0.02	0.63	±	0.07	0.006	±	0.001	0.005	±	0.001
15	Brauner Schnittkohl	0.84	±	0.02	0.79	±	0.07	0.007	±	0.001	0.005	±	0.001
16	Bravour	0.81	±	0.02	0.86	±	0.07	0.011	±	0.001	0.005	±	0.001
23	Canard	0.83	±	0.02	0.82	±	0.04	0.010	±	0.001	0.005	±	0.001
24	Canberra x Courage	0.81	±	0.02	0.92	±	0.04	0.011	±	0.001	0.007	±	0.001
37	Columbus X Nickel	0.87	±	0.02	1.00	±	0.05	0.009	±	0.001	0.006	±	0.001
40	Couve nabica	0.82	±	0.02	0.81	±	0.07	0.009	±	0.001	0.006	±	0.001
42	Darmor	0.79	±	0.02	0.72	±	0.07	0.009	±	0.001	0.006	±	0.001
43	Devon Champion	0.82	±	0.02	0.77	±	0.07	0.010	±	0.001	0.006	±	0.001
47	Doral	0.82	±	0.02	0.79	±	0.07	0.013	±	0.001	0.007	±	0.001
48	Drakkar	0.78	±	0.02	0.85	±	0.07	0.013	±	0.001	0.007	±	0.001
49	Duplo	0.77	±	0.02	0.73	±	0.05	0.011	±	0.001	0.006	±	0.001
50	Dwarf Essex	0.73	±	0.02	0.59	±	0.05	0.007	±	0.001	0.005	±	0.001
52	English Giant	0.72	±	0.02	0.58	±	0.04	0.007	±	0.001	0.006	±	0.001
53	Erglu	0.75	±	0.02	0.71	±	0.07	0.009	±	0.001	0.006	±	0.001
62	Fortin Family	0.78	±	0.02	0.80	±	0.07	0.012	±	0.001	0.006	±	0.001
65	Groene Groninger Snijmoes	0.86	±	0.02	0.82	±	0.07	0.007	±	0.001	0.005	±	0.001
70	Hansen x Gaspard	0.81	±	0.02	0.77	±	0.05	0.007	±	0.001	0.005	±	0.001
73	Huguenot	0.78	±	0.02	0.73	±	0.05	0.010	±	0.001	0.005	±	0.001
78	Jaune a Collet Vert	0.75	±	0.02	0.71	±	0.07	0.009	±	0.001	0.006	±	0.001
80	Judzae	0.80	±	0.02	0.77	±	0.05	0.009	±	0.001	0.006	±	0.001
86	Lembkes Malchower	0.80	±	0.02	0.73	±	0.07	0.007	±	0.001	0.005	±	0.001
91	Licrown x Express	0.77	±	0.02	0.78	±	0.05	0.009	±	0.001	0.006	±	0.001
99	Madrigal x Recital	0.81	±	0.02	0.77	±	0.05	0.007	±	0.001	0.005	±	0.001
100	Major	0.75	±	0.02	0.67	±	0.05	0.007	±	0.001	0.005	±	0.001
106	Moana, Moana Rape	0.78	±	0.02	0.75	±	0.07	0.008	±	0.001	0.005	±	0.001
110	N01D-1330	0.74	±	0.02	0.58	±	0.07	0.007	±	0.001	0.005	±	0.001
111	N02D-1952	0.73	±	0.02	0.60	±	0.05	0.008	±	0.001	0.005	±	0.001
115	Ningyou 7	0.77	±	0.02	0.74	±	0.07	0.010	±	0.001	0.007	±	0.001
125	POH 285, Bolko	0.78	±	0.02	0.77	±	0.05	0.010	±	0.001	0.007	±	0.001
130	Quinta	0.79	±	0.02	0.83	±	0.05	0.010	±	0.001	0.011	±	0.001
131	Rafal	0.85	±	0.02	1.03	±	0.07	0.009	±	0.001	0.005	±	0.001
133	Ramses	0.75	±	0.02	0.70	±	0.05	0.009	±	0.001	0.009	±	0.001
140	Sarepta	0.75	±	0.02	0.78	±	0.07	0.009	±	0.001	0.007	±	0.001
141	Sensation NZ	0.81	±	0.02	0.78	±	0.04	0.010	±	0.001	0.006	±	0.001
143	Shannon x Winner	0.78	±	0.02	0.73	±	0.05	0.008	±	0.001	0.005	±	0.001
150	Slapska Slapy	0.76	±	0.02	0.69	±	0.07	0.007	±	0.001	0.006	±	0.001
151	Slovenska Krajova	0.77	±	0.02	0.60	±	0.05	0.006	±	0.001	0.004	±	0.001
155	Stellar	0.73	±	0.02	0.70	±	0.05	0.012	±	0.001	0.008	±	0.001
164	Taisetsu	0.84	±	0.02	0.80	±	0.05	0.010	±	0.001	0.006	±	0.001
166	Tapidor	0.84	±	0.02	0.98	±	0.05	0.012	±	0.001	0.006	±	0.001
168	Temple	0.80	±	0.02	0.94	±	0.07	0.011	±	0.001	0.007	±	0.001
169	Tequila x Aragon	0.81	±	0.02	0.84	±	0.07	0.011	±	0.001	0.007	±	0.001
177	Victor	0.74	±	0.02	0.70	±	0.07	0.009	±	0.001	0.006	±	0.001
184	York	0.81	±	0.02	0.81	±	0.05	0.009	±	0.001	0.006	±	0.001
185	Yudal	0.73	±	0.02	0.67	±	0.07	0.009	±	0.001	0.007	±	0.001

Appendix Table 5 – Summary of fermentation inhibitors released for each cultivar when pretreated at 185°C, 10 min, controlling for differences between years and sites.

IBTI Code	Cultivar Name	Fermentation inhibitor yields (% original material)			
		Pretreated at 210°C			
		Acetic	Formic	2FA	HMF
1	Aberdeenshire Prize	1.45 ± 0.08	2.37 ± 0.40	0.2357 ± 0.0134	0.0355 ± 0.0023
2	Abukuma Natane	1.33 ± 0.07	2.71 ± 0.40	0.1597 ± 0.0101	0.0286 ± 0.0023
7	Apex-93_5 X Ginyou_3 line	1.50 ± 0.08	2.90 ± 0.39	0.1496 ± 0.0133	0.0224 ± 0.0022
11	Bienvenu	1.52 ± 0.08	3.02 ± 0.49	0.1681 ± 0.0134	0.0312 ± 0.0029
15	Brauner Schnittkohl	1.27 ± 0.08	3.03 ± 0.40	0.0886 ± 0.0134	0.0207 ± 0.0023
16	Bravour	1.34 ± 0.08	3.04 ± 0.50	0.1357 ± 0.0133	0.0206 ± 0.0030
23	Canard	1.29 ± 0.06	3.06 ± 0.40	0.1234 ± 0.0096	0.0220 ± 0.0023
24	Canberra x Courage	1.36 ± 0.05	3.20 ± 0.49	0.1363 ± 0.0081	0.0276 ± 0.0029
37	Columbus X Nickel	1.44 ± 0.06	3.25 ± 0.38	0.0985 ± 0.0096	0.0236 ± 0.0022
40	Couve nabica	1.24 ± 0.08	3.25 ± 0.49	0.1113 ± 0.0134	0.0274 ± 0.0029
42	Darmor	1.51 ± 0.08	3.25 ± 0.49	0.1652 ± 0.0134	0.0313 ± 0.0029
43	Devon Champion	1.56 ± 0.08	3.29 ± 0.49	0.1897 ± 0.0134	0.0342 ± 0.0029
47	Doral	1.33 ± 0.08	3.33 ± 0.49	0.1424 ± 0.0134	0.0272 ± 0.0029
48	Drakkar	1.28 ± 0.08	3.37 ± 0.40	0.1457 ± 0.0134	0.0291 ± 0.0023
49	Duplo	1.22 ± 0.07	3.40 ± 0.48	0.1302 ± 0.0101	0.0261 ± 0.0029
50	Dwarf Essex	1.39 ± 0.07	3.42 ± 0.38	0.1313 ± 0.0101	0.0285 ± 0.0022
52	English Giant	1.38 ± 0.05	3.43 ± 0.49	0.1124 ± 0.0081	0.0280 ± 0.0029
53	Erglu	1.45 ± 0.08	3.43 ± 0.40	0.1569 ± 0.0134	0.0322 ± 0.0023
62	Fortin Family	1.46 ± 0.08	3.44 ± 0.40	0.1965 ± 0.0134	0.0313 ± 0.0023
65	Groene Groninger Snijmoes	1.35 ± 0.08	3.46 ± 0.40	0.0776 ± 0.0134	0.0209 ± 0.0023
70	Hansen x Gaspard	1.44 ± 0.07	3.48 ± 0.49	0.1210 ± 0.0101	0.0254 ± 0.0029
73	Huguenot	1.30 ± 0.07	3.55 ± 0.33	0.1134 ± 0.0101	0.0232 ± 0.0019
78	Jaune a Collet Vert	1.38 ± 0.08	3.59 ± 0.40	0.1380 ± 0.0134	0.0274 ± 0.0023
80	Judzae	1.20 ± 0.06	3.62 ± 0.33	0.0870 ± 0.0099	0.0187 ± 0.0019
86	Lembkes Malchower	1.34 ± 0.08	3.64 ± 0.49	0.0981 ± 0.0134	0.0224 ± 0.0029
91	Licrown x Express	1.40 ± 0.06	3.65 ± 0.48	0.1177 ± 0.0096	0.0271 ± 0.0029
99	Madrigal x Recital	1.42 ± 0.07	3.72 ± 0.49	0.0979 ± 0.0101	0.0230 ± 0.0029
100	Major	1.32 ± 0.07	3.76 ± 0.50	0.1133 ± 0.0101	0.0235 ± 0.0030
106	Moana, Moana Rape	1.24 ± 0.08	3.79 ± 0.49	0.0970 ± 0.0134	0.0220 ± 0.0029
110	N01D-1330	1.27 ± 0.08	3.86 ± 0.40	0.1191 ± 0.0138	0.0229 ± 0.0023
111	N02D-1952	1.41 ± 0.07	3.86 ± 0.40	0.1751 ± 0.0101	0.0291 ± 0.0023
115	Ningyou 7	1.36 ± 0.08	3.94 ± 0.40	0.1426 ± 0.0138	0.0282 ± 0.0023
125	POH 285, Bolko	1.29 ± 0.07	3.95 ± 0.40	0.1312 ± 0.0101	0.0289 ± 0.0023
130	Quinta	1.35 ± 0.06	4.08 ± 0.49	0.1312 ± 0.0096	0.0380 ± 0.0029
131	Rafal	1.52 ± 0.08	4.10 ± 0.48	0.1556 ± 0.0133	0.0268 ± 0.0029
133	Ramses	1.29 ± 0.07	4.11 ± 0.49	0.1284 ± 0.0101	0.0328 ± 0.0029
140	Sarepta	1.26 ± 0.08	4.11 ± 0.38	0.1184 ± 0.0134	0.0273 ± 0.0022
141	Sensation NZ	1.35 ± 0.05	4.16 ± 0.49	0.1129 ± 0.0081	0.0228 ± 0.0029
143	Shannon x Winner	1.25 ± 0.06	4.22 ± 0.33	0.0914 ± 0.0096	0.0193 ± 0.0019
150	Slapska Slapy	1.33 ± 0.08	4.26 ± 0.38	0.1170 ± 0.0134	0.0246 ± 0.0022
151	Slovenska Krajova	1.23 ± 0.07	4.28 ± 0.40	0.0827 ± 0.0101	0.0188 ± 0.0023
155	Stellar	1.25 ± 0.07	4.35 ± 0.49	0.1789 ± 0.0101	0.0343 ± 0.0029
164	Taisetsu	1.35 ± 0.07	4.36 ± 0.49	0.1327 ± 0.0101	0.0255 ± 0.0029
166	Tapidor	1.38 ± 0.06	4.37 ± 0.49	0.1185 ± 0.0096	0.0261 ± 0.0029
168	Temple	1.22 ± 0.08	4.38 ± 0.49	0.1098 ± 0.0133	0.0224 ± 0.0029
169	Tequila x Aragon	1.42 ± 0.08	4.43 ± 0.38	0.1557 ± 0.0134	0.0295 ± 0.0022
177	Victor	1.28 ± 0.08	4.76 ± 0.49	0.1284 ± 0.0134	0.0264 ± 0.0029
184	York	1.40 ± 0.07	5.02 ± 0.38	0.1233 ± 0.0101	0.0282 ± 0.0022
185	Yudal	1.47 ± 0.08	5.19 ± 0.48	0.1881 ± 0.0134	0.0376 ± 0.0029

Appendix Table 6 – Summary of fermentation inhibitors released for each cultivar when pretreated at 210°C, 10 min, controlling for differences between years and sites.

Glucose 185L

Brassica napus								*Arabidopsis thaliana*		
Marker	Position		-log10P	Effect size*	SNP Alleles			Ortholog AGI	Candidate AGI	Candidate name
	Chr	Pseudo			-	Modal	+			
JCVI_2389:1073	C4	25113320	2.861	8.20%	Y (16)	T (23)	-	AT3G63000	AT3G62830	UDP-GlcA decarboxylase 2
JCVI_162:596	C3	6906458	2.757	9.42%	Y (10)	T (31)	-	AT5G56760	AT5G56590	O-Glycosyl hydrolases Family 17
JCVI_929:491	C6	36810175	2.669	12.64%	T (4)	Y (37)	-	AT1G65720	AT1G65610	Korrigan 2
JCVI_12228:405	C9	60646527	2.661	27.45%	A (3)	R (33)	-	AT5G08130	AT5G08000	Plasmodesmata Callose-Binding 2
JCVI_6088:571	A8	10661406	2.592	24.38%	C (9)	Y (30)	-	AT4G30160	AT4G30170	Peroxidase family protein
JCVI_23516:986	A4	11787316	2.547	23.22%	C (12)	Y (27)	-	AT2G23950	AT2G23900	Pectin lyase-like superfamily protein
JCVI_3121:213	C7	44734046	2.480	8.84%	S (15)	G (25)	-	AT4G22890	AT4G22970	Radially swollen 4
JCVI_20014:160	A6	15944517	2.448	11.80%	C (4)	Y (36)	-	AT5G63030	AT5G63100	SAM-dependent methyltransferases superfamily protein
JCVI_5702:591	A4	3178332	2.428	7.20%	-	A (20)	W (18)	AT3G54760	AT3G54690	Sugar isomerase (SIS) family protein
JCVI_7765:772	C4	54480036	2.421	11.50%	S (5)	G (33)	-	AT2G45810	AT2G45750	SAM-dependent methyltransferases superfamily protein
JCVI_7891:119	A9	29940741	2.401	9.15%	A (8)	R (32)	-	AT3G51840	-	-
JCVI_26311:900	C1	2739833	2.377	22.48%	Y (3)	T (36)	-	AT4G33700	AT4G33810-60	Series of Glycosyl Hydrolases (F-10)
JCVI_14904:653	C5	37173970	2.359	13.77%	C (3)	Y (37)	-	AT3G15570	AT3G15530	SAM-dependent methyltransferases superfamily protein
JCVI_12228:492	A10	12745297	2.354	11.87%	Y (4)	T (34)	-	AT5G08130	AT5G08000	Plasmodesmata Callose-Binding 2
JCVI_18630:1389	C9	54639474	2.343	7.78%	T (15)	Y (24)	-	AT5G19180	AT5G19160	Trichome birefringence-like 11
JCVI_27767:879	A7	10349313	2.329	8.28%	T (9)	Y (23)	-	AT1G22380	AT1G22360-400	UDP-Glucosyl transferase 85A1, 2 & 3
JCVI_2835:408	A5	18400735	2.297	11.77%	C (4)	Y (33)	-	AT3G17100	Multiple	Pectin methylesterase inhibitor (AT3G17130), Peroxidase (AT3G17070)
JCVI_40933:907	C3	19944437	2.222	13.01%	T (3)	W (35)	-	AT4G09570	AT4G09500	UDP-Glycosyltransferase superfamily protein
JCVI_32045:540	A6	22525163	2.210	11.48%	T (4)	Y (32)	-	AT3G26030	AT3G26040	HXXXD-type acyl-transferase family protein
JCVI_26679:485	C9	49404566	2.202	9.78%	-	T (31)	Y (6)	AT5G58560	AT5G58600	Trichome birefringence-like 44
JCVI_15509:575	A6	19122133	2.194	8.88%	T (11)	Y (28)	-	AT5G24710	AT2G23900	Pectin lyase-like superfamily
JCVI_7088:105	A7	12376898	2.192	12.61%	-	R (36)	G (3)	AT2G27385	AT2G27380	Extensin proline-rich 1
JCVI_18492:975	A6	6026224	2.188	4.82%	Y (13)	C (25)	-	AT1G15410	-	-
JCVI_5256:202	A9	26259228	2.149	9.33%	W (7)	A (32)	-	AT1G27400	AT1G27440	Glucuronoxylan glucuronosyltransferase 2 (GUT2, IRX10)
JCVI_26272:1210	A7	22372439	2.126	7.92%	G (10)	S (30)	-	AT3G63510	AT1G71220	UDP-glucose:glycoprotein glucosyltransferase.
JCVI_6330:482	A7	18211956	2.120	8.22%	R (12)	A (29)	-	AT1G74970	AT1G74910	ADP-glucose pyrophosphorylase family protein
JCVI_7310:1212	C5	3898821	2.085	9.28%	T (6)	K (32)	-	AT1G09970	Multiple	Sucrose transporter 4 / Glycosyl hydrolase family 10 protein (xylanase)
JCVI_40080:473	C3	49986700	2.073	13.99%	-	Y (37)	T (2)	AT3G45010	AT5G42080	Radial swelling 9
JCVI_6180:337	A2	25682417	2.067	8.84%	C (6)	S (34)	-	AT5G64850	AT5G64790	O-Glycosyl hydrolases family 17 protein
JCVI_4323:501	C6	42244003	2.057	9.13%	R (6)	A (31)	-	AT1G71480	AT1G71691	GDSL-like Lipase/Acylhydrolase superfamily protein
JCVI_14508:616	C8	26011248	2.055	4.61%	G (14)	R (25)	-	AT3G54360	Multiple	TBL 36 or Glycoside hydrolase family 2 protein
JCVI_18843:1357	A3	28023029	2.046	7.47%	-	M (28)	C (11)	AT4G30810	AT4G30760	Putative glycosyl hydrolase
JCVI_19836:881	A1	407303	2.042	4.68%	C (12)	Y (29)	-	AT2G39460	AT2G39445	Phosphatidylinositol N-acetylglucosaminyltransferase
JCVI_24148:863	C2	37751528	2.039	4.68%	G (11)	R (29)	-	AT5G46800	AT5G46730	Glycine-rich protein family
JCVI_814:463	A9	37666126	2.039	9.24%	Y (12)	T (25)	-	AT1G15950	AT1G15950	Cinnamoyl CoA reductase 1 (CCR1, IRX4)
JCVI_17446:235	C5	11041316	2.014	7.27%	-	A (32)	M (9)	AT1G20810	AT1G20840	Tonoplast monsaccharide transporter 1
JCVI_15662:838	C6	24362786	2.002	12.40%	-	Y (35)	C (3)	AT3G53900	-	-

*Effect size is the quantitative difference between means for each marker, expressed as % of the mean

Appendix Table 7 – Glc 185 Low – SNP markers found in regions of the *B. napus* transcriptome associated with the variation in Glc yields released from various cultivars ($p < 0.01$) after liquid hot water pretreatment at 185 °C for 10 min, followed by hydrolysis using a low (7 FPU/g) cellulase dose.

Reducing sugar 185L											
Brassica napus									*Arabidopsis thaliana*		
Marker	Position		-log10P	Effect size*	SNP Alleles				Ortholog AGI	Candidate AGI	Candidate name
	Chr	Pseudo			-	Modal	+				
JCVI_2389:1073	C4	25113320	3.432	7.97%	Y (16)	T(23)	-		AT3G63000	AT3G62830	UDP-GlcA decarboxylase 2
JCVI_430:165	A8	17643293	2.675	6.19%	A (19)	R (22)	-		AT1G09070	AT1G08990	GlcA substitution of xylan 5
JCVI_41506:267	**A4**	**2405408**	2.616	7.11%	**S (17)**	**C (22)**	**-**		**AT3G56190**	**AT3G56040**	**UDP-glucose pyrophosphorylase 3**
JCVI_1336:469	**C4**	**34725981**	2.604	9.30%	**C (6)**	**Y (35)**	**-**		**AT5G37600**	**AT5G37660**	**Plasmodesmata-located protein 7**
JCVI_7836:240	**C1**	**18148122**	2.421	9.97%	**W (4)**	**A (37)**	**-**		**AT4G14320**	**AT4G14310**	**CW located Transducin/WD40 repeat-like superfamily protein**
JCVI_16118:186	C9	48687099	2.392	7.00%	T (10)	Y (30)	-		AT5G57710	AT5G57655	Xylose isomerase family protein
JCVI_7891:119	**A9**	**29940741**	**2.386**	**7.88%**	**A (8)**	**R (32)**	**-**		**AT3G51840**	-	-
JCVI_1056:545	C3	58549048	2.366	8.34%	C (6)	M (34)	-		AT4G36690	AT4G36770	UDP-Glycosyltransferase superfamily
JCVI_5260:759	**C7**	**35381951**	**2.331**	**5.74%**	**-**	**G (20)**	**R (19)**		**AT3G27820**	**AT3G27810**	**MYB domain protein 21**
JCVI_27767:879	A7	10349313	2.328	7.25%	-	Y (23)	T (9)		AT1G22380	AT1G22360-40(	UDP-Glucosyl transferase 85A1, 2 & 3
JCVI_6088:571	A8	10661406	2.326	6.94%	C (9)	Y (30)	-		AT4G30160	AT4G30170	Peroxidase family protein
JCVI_197:367	A10	6561639	2.317	7.36%	R (8)	A (33)	-		AT5G57660	AT5G57655	Xylose isomerase family protein
JCVI_8724:1237	A3	30077326	2.295	8.74%	-	S (32)	G (5)		AT4G35790	AT4G35670	Pectin lyase-like superfamily protein
JCVI_8788:1017	A1	22635931	2.288	7.13%	-	Y (28)	C (8)		AT3G18750	AT3G18660	GlcA substitution of xylan 1
JCVI_12228:492	A10	12745297	2.268	9.79%	Y (4)	T (34)	-		AT5G08130	AT5G08000	Plasmodesmata Callose-Binding 2
JCVI_15647:325	A6	9069124	2.255	7.31%	-	C (23)	S (12)		AT1G21680	Multiple	Cell-wall localised gene / hydroxyproline-rich glycoprotein family protein
JCVI_11821:822	C3	4803363	2.252	7.78%	A (7)	M (29)	-		AT5G19180	AT5G19160	Trichome birefringence-like 11
JCVI_8150:633	**A3**	**8681688**	**2.222**	**7.37%**	**G (7)**	**S (34)**	**-**		**AT2G37110**	**AT2G37090**	**Irregular xylem 9**
JCVI_5000:440	**C7**	**49488794**	**2.211**	**5.88%**	**G (18)**	**K (23)**	**-**		**AT4G35490**	**AT4G35335**	**Nucleotide-sugar transporter family protein**
JCVI_9955:243	**A1**	**2125954**	**2.195**	**10.37%**	**G (3)**	**R (37)**	**-**		**AT4G33510**	**AT4G33440**	**Pectin lyase-like superfamily protein**
JCVI_12842:125	A1	9738463	2.147	6.47%	-	Y (30)	T (11)		AT5G57815	AT5G58090	O-Glycosyl hydrolases family 17
JCVI_1134:437	A10	1708228	2.126	6.85%	M (9)	C (32)	-		AT1G04940	AT1G04910	O-fucosyltransferase family protein
JCVI_1535:492	C9	24392590	2.125	8.87%	-	Y (37)	T (4)		AT5G47120	Multiple	Low-level β-amylase 1 (AT5G47010) or Peroxidase superfamily (AT5G47000)
JCVI_3121:213	C7	44734046	2.116	6.65%	S (15)	G (25)	-		AT4G22890	AT4G22970	Radially swollen 4
JCVI_26272:1210	A7	22372439	2.114	6.33%	G (10)	S (30)	-		AT3G63510	AT1G71220	UDP-glucose:glycoprotein glucosyltransferase.
JCVI_14588:1822	A4	6317040	2.108	9.31%	A (4)	W (35)	-		AT5G35690	AT5G35660	Glycine-rich protein family
JCVI_6033:726	**C9**	**26508586**	**2.102**	**10.23%**	**T (3)**	**Y (33)**	**-**		**AT3G25805**	**AT3G25790**	**MYB-like transcription factor family protein**
JCVI_17961:328	**A9**	**25881394**	**2.085**	6.61%	**-**	**C (28)**	**Y (9)**		**AT1G28150**	**AT1G28290**	**Arabinogalactan protein 31**
JCVI_6684:60	A3	23188791	2.079	5.71%	-	R (27)	G (14)		AT4G18100	AT4G18030	SAM-dependent methyltransferases superfamily protein
JCVI_17664:893	**A2**	**18739683**	**2.073**	**7.82%**	**T (5)**	**K (36)**	**-**		AT2G01320	AT2G01320	ABC-2 type transporter family protein
JCVI_1657:330	**A5**	**1823923**	**2.049**	**5.86%**	**Y (16)**	**C (25)**	**-**		**AT2G43750**	**AT2G43740**	**Mannose-binding lectin superfamily protein**
JCVI_4441:401	**A8**	**12629468**	**2.044**	**5.53%**	**K (18)**	**G (20)**	**-**		**AT4G38100**	**AT4G38080**	**Hydroxyproline-rich glycoprotein family protein**
JCVI_16678:1065	**C2**	**47529134**	**2.034**	**7.93%**	**A (5)**	**R (36)**	**-**		**AT5G26742**	**AT5G26730**	**Fasciclin-like arabinogalactan family protein**
JCVI_30870:389	C4	54006138	2.026	5.83%	S (14)	G (27)	-		AT2G01250	AT2G44110	Mildew resistance locus O 15
JCVI_27253:237	A1	12906362	2.011	5.99%	M (14)	C (23)	-		AT3G49800	Multiple	Carbohydrate-binding protein / glycosyl hydrolase family protein 43
JCVI_7695:701	A3	14117383	2.003	5.57%	-	K (22)	G (19)		AT4G01370	AT4G01210	Glycosyltransferase family 1 protein

Appendix Table 8 – RS 185 Low – SNP markers found in regions of the *B. napus* transcriptome associated with the variation in RS yields released from various cultivars ($p < 0.01$) after liquid hot water pretreatment at 185 °C for 10 min, followed by hydrolysis using a low (7 FPU/g) cellulase dose.

Glucose 185H

Brassica napus									*Arabidopsis thaliana*		
Marker	**Position**		**-log10P**	**Effect size***	**SNP Alleles**				**Ortholog AGI**	**Candidate AGI**	**Candidate name**
	Chr	**Pseudo**			**-**	**Modal**	**+**				
JCVI_2389:1073	C4	25113320	3.321	14.53%	Y (16)	T (23)	-		AT3G63000	AT3G62830	UDP-GlcA decarboxylase 2
JCVI_1973:910	C9	48568519	3.151	13.41%	R (16)	A (24)	-		AT5G57655	AT5G57655	Xylose isomerase family protein
JCVI_12842:44	A1	9738463	2.569	14.53%	C (10)	Y (28)	-		AT5G57815	AT5G58090	O-Glycosyl hydrolases family 17
JCVI_1056:545	C3	58549048	2.532	16.29%	C (6)	M (34)	-		AT4G36690	AT4G36770	UDP-Glycosyltransferase superfamily
JCVI_37942:1200	C1	12385276	2.496	11.69%	Y (17)	T (22)	-		AT4G24550	AT4G24530	O-fucosyltransferase family protein
JCVI_430:165	A8	17643293	2.423	10.69%	A (19)	R (22)	-		AT1G09070	AT1G08990	GlcA substitution of xylan 5
JCVI_7695:701	A3	14117383	2.413	11.92%	-	K (22)	G (19)		AT4G01370	AT4G01210	Glycosyltransferase family 1 protein
JCVI_17446:235	C5	11041316	2.411	13.19%	-	A (32)	M (9)		AT1G20810	AT1G20840	Tonoplast monosaccharide transporter 1
JCVI_27767:879	A7	10349313	2.304	13.40%		Y (23)	T (9)		AT1G22380	AT1G22360-400	UDP-Glucosyl transferase 85A1, 2 & 3
JCVI_2316:417	C7	43305017	2.278	11.03%	T (13)	K (28)	-		AT4G18730	AT4G1878	Cellulose Synthase 8 (IRX 1)
JCVI_18630:960	C9	54639474	2.260	12.09%	Y (15)	C (24)	-		AT5G19180	AT5G19160	Trichome birefringence-like 11
JCVI_35535:277	C6	6967230	2.253	12.50%	C (11)	Y (25)	-		AT1G50740	-	-
JCVI_8788:1017	A1	22635931	2.245	12.97%	-	Y (28)	C (8)		AT3G18750	AT3G18660	GlcA substitution of xylan 1
JCVI_5840:1390	C9	5629956	2.240	12.09%	C (15)	Y (22)	-		AT5G08570	-	-
JCVI_8724:1237	A3	30077326	2.235	15.86%	-	S (32)	G (5)		AT4G35790	AT4G35670	Pectin lyase-like superfamily protein
JCVI_3365:391	C3	56470210	2.223	12.60%	R (15)	G (24)	-		AT4G38600	AT4G38620	MYB domain protein 4 (acts on C4H)
JCVI_21919:1178	C3	41744982	2.215	15.01%	G (6)	R (31)	-		AT5G61810	AT5G61840	IRX10-Like - Glucuronosyltransferase
JCVI_197:367	A10	6561639	2.202	13.16%	R (8)	A (33)	-		AT5G57660	AT5G57655	Xylose isomerase family protein
JCVI_1535:492	C9	24392590	2.190	16.65%	-	Y (37)	T (4)		AT5G47120	Multiple	Low-level beta-amylase 1 (AT5G47010) or Peroxidase (AT5G47000)
JCVI_16252:1130	C8	10965209	2.188	13.28%	-	T (32)	Y (8)		AT4G14570	AT4G14570	Acylamino acid-releasing enzyme
JCVI_6053:538	C2	5726314	2.172	10.70%	G (17)	K (23)	-		AT5G19855	Multiple	Proline-rich extensin-like family protein / Peroxidase superfamily
JCVI_27253:237	A1	12906362	2.168	11.71%	M (14)	C (23)	-		AT3G49800	Multiple	Carbohydrate-binding (AT3G49790) / GH family 43 (AT3G49880)
JCVI_9082:107	A5	23386533	2.157	16.80%	C (5)	Y (30)	-		AT3G03870	Multiple	O-GH family 17 / hydroxyproline-rich glycoprotein family protein
JCVI_2613 6:231	A9	14141369	2.153	14.52%	-	W (30)	T (6)		AT2G02080	AT2G02061	Nucleotide-diphospho-sugar transferase family protein
JCVI_30885:329	A9	38971969	2.141	11.13%	-	C (31)	Y (10)		AT1G11860	AT1G11820	O-Glycosyl hydrolases family 17 protein
JCVI_5454:527	C9	21256155	2.140	12.74%	-	G (33)	S (8)		AT5G43850	AT5G43770	Proline-rich family protein
JCVI_32045:540	A6	22525163	2.126	18.72%	T (4)	Y (32)	-		AT3G26030	AT3G26040	HXXXD-type acyl-transferase family protein
JCVI_7017:143	A8	11313784	2.111	11.52%	-	Y (29)	T (12)		AT4G27090	AT4G27110	COBRA-like protein 11 precursor
JCVI_12969:772	A10	10101684	2.110	15.75%	S (5)	G (36)	-		AT5G17230	Multiple	Nucleotide-sugar transporter family protein / glycine-proline-rich
JCVI_16451:1305	A5	20562232	2.087	10.79%	-	Y (19)	C (17)		AT3G12360	-	-
JCVI_6684:60	A3	23188791	2.076	10.67%	-	R (27)	G (14)		AT4G18100	AT4G18030	SAM-dependent methyltransferases superfamily
JCVI_929:491	C6	36810175	2.058	17.92%	T (4)	Y (37)	-		AT1G65720	AT1G65610	Korrigan 2
JCVI_2675:104	A3	2630016	2.045	12.25%	R (8)	G (29)	-		AT5G14030	AT5G13980	Glycosyl hydrolase family 38 protein
JCVI_23017:490	A3	4895602	2.042	11.08%	T (15)	Y (22)	-		AT5G58640	AT5G58600	Trichome birefringence-like protein - powdery mildew resistant 5
JCVI_14588:1822	A4	6317040	2.024	16.98%	A (4)	W (35)	-		AT5G35690	AT5G35660	Glycine-rich protein family
JCVI_853:308	A3	19431874	2.024	10.42%	-	A (24)	M (17)		AT3G24170	AT3G24180	Beta-glucosidase, GBA2 type family protein
JCVI_26470:129	C4	33041573	2.015	16.79%	R (4)	G (33)	-		AT4G23260	AT4G13410	Cellulose-synthase like A15
JCVI_12562:343	A9	6935506	2.010	10.26%	M (20)	C (20)	-		AT2G20420	-	-

*Effect size is the quantitative difference between means for each marker, expressed as % of the mean

Appendix Table 9 – Glc 185 High – SNP markers found in regions of the *B. napus* transcriptome associated with the variation in Glc yields released from various cultivars ($p < 0.01$) after liquid hot water pretreatment at 185 °C for 10 min, followed by hydrolysis using a high (36 FPU/g) cellulase dose.

Reducing sugar 185H

Brassica napus								*Arabidopsis thaliana*		
Marker	Position		-log10P	Effect size*	SNP Alleles			Ortholog AGI	Candidate AGI	Candidate name
	Chr	Pseudo			-	Modal	+			
JCVI_19836:881	A1	407303	2.002	8.64%	C (12)	Y (29)	-	AT2G39460	AT2G39445	Phosphatidylinositol N-acetylglucosaminyltransferase
JCVI_12842:44	**A1**	**9738463**	**2.449**	**12.34%**	**-**	**Y (28)**	**C (10)**	**AT5G57815**	**AT5G58090**	**O-Glycosyl hydrolases family 17**
JCVI_27253:237	A1	12906362	2.341	9.96%	M (14)	C (23)	-	AT3G49800	Multiple	Carbohydrate-binding protein / glycosyl hydrolase family protein 43
JCVI_2675:104	A3	2630016	2.439	12.09%	R (8)	G (29)	-	AT5G14030	AT5G13980	Glycosyl hydrolase family 38 protein
JCVI_2881:545	**A3**	**13228402**	**2.196**	**2.19%**	**-**	**A (29)**	**R (12)**	**AT4G04640**	**-**	**-**
JCVI_8724:1237	A3	30077326	2.041	4.01%	-	S (32)	G (5)	AT4G35790	AT4G35670	Pectin lyase-like superfamily protein
JCVI_14588:1822	A4	6317040	2.128	14.33%	A (4)	W (35)	-	AT5G35690	AT5G35660	Glycine-rich protein family
JCVI_5779:448	**A5**	**1545416**	**2.008**	**12.82%**	**A (4)**	**W (37)**	**-**	**AT2G42680**	**AT2G42570**	**Trichome birefringence-like 39**
JCVI_9082:107	A5	23386533	2.168	13.26%	C (5)	Y (30)	-	AT3G03870	Multiple	O-Glycosyl hydrolases family 17 / hydroxyproline-rich glycoprotein
JCVI_15647:325	A6	9069124	2.169	10.48%	-	C (23)	S (12)	AT1G21680	Multiple	Cell-wall localised gene / hydroxyproline-rich glycoprotein family protein
JCVI_32045:540	A6	22525163	2.352	4.21%	T (4)	Y (32)	-	AT3G26030	AT3G26040	HXXXD-type acyl-transferase family protein
JCVI_27767:879	A7	10349313	2.313	11.01%	-	Y (23)	T (9)	AT1G22380	AT1G22360-400	UDP-Glucosyl transferase 85A1, 2 & 3
JCVI_9640:292	**A7**	**20340681**	**2.197**	**14.47%**	**Y (4)**	**T (31)**	**-**	**AT1G66330**	**AT1G66270-80**	**β-glucosidase 21 - 22**
JCVI_7017:143	A8	11313784	2.247	9.89%	-	Y (29)	T (12)	AT4G27090	AT4G27110	COBRA-like protein 11 precursor
JCVI_430:165	A8	17643293	2.674	9.06%	A (19)	R (22)	-	AT1G09070	AT1G08990	GlcA substitution of xylan 5
JCVI_959:621	**A9**	**7016640**	**2.033**	**10.44%**	**-**	**A (35)**	**R (14)**	**AT1G64040**	**AT1G63855**	**Putative methyltransferase family protein**
JCVI_30885:329	A9	38971969	2.105	9.11%	-	C (31)	Y (10)	AT1G11860	AT1G11820	O-Glycosyl hydrolases family 17 protein
JCVI_23828:1610	**A10**	**6859222**	**2.326**	**10.46%**	**-**	**R (26)**	**G (13)**	**AT5G58440**	**AT5G58480**	**O-Glycosyl hydrolases family 17 protein**
JCVI_12969:772	A10	10101684	2.051	13.49%	S (5)	G (36)	-	AT5G17230	Multiple	Nucleotide-sugar transporter family protein / glycine-proline-rich protein
JCVI_37942:1200	C1	12385276	2.594	10.10%	Y (17)	T (22)	-	AT4G24550	AT4G24530	O-fucosyltransferase family protein
JCVI_6053:538	C2	5726314	2.176	8.86%	G (17)	K (23)	-	AT5G19855	Multiple	Proline-rich extensin-like family protein / Peroxidase superfamily protein
JCVI_3199:271	**C2**	**47632073**	**2.560**	**12.94%**	**T (6)**	**K (35)**	**-**	**AT5G26000**	**Multiple**	**βG 37-38 or Core-2/I-branching β-1,6-N-acetylglucosaminyltransferase**
JCVI_8464:784	**C3**	**16534486**	**2.009**	**9.84%**	**T (8)**	**Y (31)**	**-**	**AT2G28840**	**AT2G28760**	**UDP-XYL synthase 6**
JCVI_40981:148	**C3**	**18773313**	**2.106**	**8.44%**	**-**	**G (24)**	**R (17)**	**AT2G21340**	**AT2G21370**	**Sugar Kinase-1**
JCVI_21919:1178	C3	41744982	2.159	12.44%	G (6)	R (31)	-	AT5G61810	AT5G61840	IRX10-Like - Glucuronosyltransferase
JCVI_1056:545	C3	58549048	2.487	13.50%	C (6)	M (34)	-	AT4G36690	AT4G36770	UDP-Glycosyltransferase superfamily
JCVI_27578:2248	C4	6392947	2.602	10.80%	-	C (25)	M (16)	AT2G38770	Multiple	Annexins 3 & 4 / microtubule-associated protein 65-5
JCVI_2389:1073	C4	25113320	3.116	11.49%	Y (16)	T (23)	-	AT3G63000	AT3G62830	UDP-GlcA decarboxylase 2
JCVI_17446:235	C5	11041316	2.177	10.19%	-	A (32)	M (9)	AT1G20810	AT1G20840	Tonoplast monsaccharide transporter 1
JCVI_35535:277	C6	6967230	2.412	11.18%	C (11)	Y (25)	-	AT1G50740	-	-
JCVI_5309:271	C6	9236700	2.385	10.53%	T (9)	Y (32)	-	AT1G53210	AT1G53290	Galactosyltransferase family protein
JCVI_929:491	C6	36810175	2.263	14.88%	T (4)	Y (37)	-	AT1G65720	AT1G65610	Korrigan 2
JCVI_16387:539	**C7**	**31709156**	**2.114**	**12.87%**	**G (5)**	**K (34)**	**-**	**AT2G01820**	**AT2G01850**	**Xyloglucan endotransglucosylase/hydrolase 27**
JCVI_14498:684	**C7**	**46953002**	**2.011**	**12.80%**	**W (5)**	**T (35)**	**-**	**AT4G28740**	**AT4G28780**	**GDSL-like Lipase/Acylhydrolase superfamily protein**
JCVI_8475:133	C7	50748254	2.228	11.13%	C (9)	Y (31)	-	AT2G20930	AT2G20870	Cell wall protein precursor, putative
JCVI_16252:1130	C8	10965209	2.221	11.40%	-	T (32)	Y (8)	AT4G14570	AT4G14570	Acylamino acid-releasing enzyme
JCVI_5840:1390	C9	5629956	2.072	9.95%	C (15)	Y (22)	-	AT5G08570	-	-
JCVI_1973:910	C9	48568519	3.159	11.18%	R (16)	A (24)	-	AT5G57655	AT5G57655	Xylose isomerase family protein
JCVI_18630:960	C9	54639474	2.217	9.75%	Y (15)	C (24)	-	AT5G19180	AT5G19160	Trichome birefringence-like 11

*Effect size is the quantitative difference between means for each marker, expressed as % of the mean

Appendix Table 10 – RS 185 High – SNP markers found in regions of the *B. napus* transcriptome associated with the variation in Glc yields released from various cultivars ($p < 0.01$) after liquid hot water pretreatment at 185 °C for 10 min, followed by hydrolysis using a high (36 FPU/g) cellulase dose.

Glucose 210L												
Brassica napus									*Arabidopsis thaliana*			
Marker	Position			-log10P	Effect size*	SNP Alleles			Ortholog AGI	Candidate AGI	Candidate name
	Chr	Pseudo				-	Modal	+			
JCVI_1973:910	C9	48568519	3.662	10.88%	R (16)	A (24)	-	AT5G57655	AT5G57655	Xylose isomerase family protein
JCVI_8724:1452	A3	30077326	3.240	15.06%	-	S (36)	G (5)	AT4G35790	AT4G35670	Pectin lyase-like superfamily protein
JCVI_14164:217	C9	23878343	3.019	10.02%	T (17)	K (20)	-	AT4G17870	AT4G17880	MYC4 - JAZ-interacting transcription factor
JCVI_690:199	A8	13196172	2.732	9.68%	-	R (28)	A (13)	AT4G14300	AT4G14310	CW located Transducin/WD40 repeat-like superfamily protein
JCVI_32230:734	A7	24389437	2.712	9.02%	Y (18)	T (19)	-	AT1G76790	AT1G76670	Nucleotide-sugar transporter family protein
JCVI_6088:571	A8	10661406	2.588	10.13%	C (9)	Y (30)	-	AT4G30160	AT4G30170	Peroxidase family protein
JCVI_197:367	A10	6561639	2.571	10.98%	R (8)	A (33)	-	AT5G57660	AT5G57655	Xylose isomerase family protein
JCVI_17322:163	C3	56294303	2.552	9.08%	-	T (28)	Y (13)	AT4G14230	-	-
JCVI_2389:1073	C4	25113320	2.455	8.51%	Y (16)	T (23)	-	AT3G63000	AT3G62830	UDP-GlcA decarboxylase 2
JCVI_15537:707	C2	43992872	2.451	13.48%	-	C (33)	S (4)	AT3G27770	AT3G27785	MYB domain protein 118
JCVI_2549:462	C6	28317182	2.435	8.59%	T (15)	K (25)	-	AT3G59920	AT3G59850	Pectin lyase-like superfamily protein
JCVI_12640:687	A3	25585961	2.406	11.55%	G (14)	R (27)	-	AT4G24280	AT4G24250	Mildew resistance locus O13
JCVI_27767:879	A7	10349313	2.404	9.98%	-	Y (23)	T (9)	AT1G22380	AT1G22360-400	UDP-Glucosyl transferase 85A1, 2 & 3
JCVI_11821:822	C3	4803363	2.356	10.68%	A (7)	M (29)	-	AT5G19160	AT5G19180	Trichome birefringence-like 11
JCVI_26470:129	C4	33041573	2.300	13.27%	R (4)	G (33)	-	AT4G23260	AT4G13410	Cellulose-synthase like A15
JCVI_13472:1244	C9	4611297	2.283	7.96%	Y (18)	T (22)	-	AT5G60920	AT5G60950	COBRA-like protein 5 precursor
JCVI_19484:115	A4	6615860	2.244	12.95%	S (4)	C (35)	-	AT5G35970	AT5G35940-950	Mannose-binding lectin superfamily proteins
JCVI_1101:362	C5	24761456	2.234	10.31%	T (8)	Y (25)	-	AT1G32900	Multiple	Granule bound starch synthase 1 / HXXXD-type acyl-transferase family
JCVI_3734:445	A9	2083890	2.226	12.44%	A (6)	W (25)	-	AT5G27380	Multiple	Sugar-porter family protein 1 / SAM-dependent methyltransferases
JCVI_16252:1130	C8	10965209	2.224	9.87%	-	T (32)	Y (8)	AT4G14570	AT4G14570	Acylamino acid-releasing enzyme
JCVI_3521:474	A3	16156585	2.206	12.74%	Y (4)	T (37)	-	AT3G11170	AT3G11030	Trichome birefringence-like 32
JCVI_3121:213	C7	44734046	2.206	9.11%	S (15)	G (35)	-	AT4G22890	AT4G22970	Radially swollen 4
JCVI_3400:226	A4	1574752	2.183	10.13%	K (7)	T (31)	-	AT1G04810	AT1G04800	Glycine-rich protein
JCVI_32920:1259	C7	46984591	2.172	9.32%	Y (10)	T (28)	-	AT4G28880	AT4G28850	Xyloglucan endotransglucosylase/hydrolase 26
JCVI_8788:1017	A1	22635931	2.140	9.37%	-	Y (28)	C (8)	AT3G18750	AT3G18660	GlcA substitution of xylan 1
JCVI_10983:1218	A9	134960	2.131	8.71%	-	C (28)	M (11)	AT4G00370	-	-
JCVI_37797:420	A3	8311101	2.125	7.98%	-	Y (25)	C (14)	AT2G35800	Multiple	Sucrose-6F-P phosphohydrolase/FASCICLIN-like arabinogalactan P16 precurso
JCVI_8333:112	C2	9279131	2.096	7.67%	M (16)	C (25)	-	AT5G58310	AT5G58310	Methyl esterase 18
JCVI_27302:827	C3	39494864	2.088	7.78%	K (13)	T (26)	-	AT5G63840	AT5G63840	Radial swelling 3
JCVI_1880:413	A3	19772119	2.077	7.46%	-	G (20)	K (18)	AT3G25230	AT3G25140	Galacturonosyl transferase 8 - Quasimodo 1
JCVI_19553:697	C1	43952241	2.069	8.51%	-	Y (29)	T (11)	AT3G17170	AT3G17130-52	Group of plant invertase/pectin methylesterase inhibitor superfamily proteins
JCVI_3799:559	A8	16999190	2.064	7.97%	T (14)	W (27)	-	AT1G11840	AT2G01630	O-Glycosyl hydrolases family 17
JCVI_25843:1159	A6	16030424	2.054	11.01%	T (5)	K (33)	-	AT5G63140	Multiple	SAM-dependent methyltransferases SF / Pectin lyase-like
JCVI_16555:614	A6	19949020	2.048	8.03%	-	G (27)	S (12)	AT5G26830	AT5G26810	Pectin lyase-like superfamily protein
JCVI_20362:225	A9	14453765	2.042	10.92%	-	C (35)	M (6)	AT2G03310	AT2G03480	Quasimodo2-Like 2
JCVI_3670:1056	A2	5342360	2.036	10.83%	-	C (30)	M (7)	AT5G58720	-	-
JCVI_4532:234	C1	49190708	2.008	9.15%	S (8)	C (24)	-	AT3G09300	AT3G09260	ER-localised β-glucosidase
JCVI_18812:1052	C1	2968007	2.005	10.82%	R (5)	A (36)	-	AT4G33090	AT4G33110-20	SAM-dependent methyltransferases superfamily proteins

*Effect size is the quantitative difference between means for each marker, expressed as % of the mean

Appendix Table 11 – Glc 210 Low – SNP markers found in regions of the *B. napus* transcriptome associated with the variation in Glc yields released from various cultivars ($p < 0.01$) after liquid hot water pretreatment at 210 °C for 10 min, followed by hydrolysis using a low (7 FPU/g) cellulase dose.

Reducing sugar 210L

Brassica napus								*Arabidopsis thaliana*		
Marker	Position		-log10P	Effect size*	SNP Alleles			Ortholog AGI	Candidate AGI	Candidate name
	Chr	Pseudo			-	Modal	+			
JCVI_1973:910	C9	48568519	3.095	8.84%	R (16)	A (24)	-	AT5G57655	AT5G57655	Xylose isomerase family protein
JCVI_2389:1073	C4	25113320	3.085	9.18%	Y (16)	T (23)	-	AT3G62830	AT3G62830	UDP-GlcA decarboxylase 2
JCVI_1535:492	C9	24392590	2.731	13.14%	-	Y (37)	T (4)	AT5G47120	Multiple	Low-level beta-amylase 1 (AT5G47010) or Peroxidase (AT5G47000)
JCVI_430:165	A8	17643293	2.495	7.38%	A (19)	R (22)	-	AT1G08990	AT1G08990	GlcA substitution of xylan 5
JCVI_16252:1130	C8	10965209	2.457	9.55%	-	T (32)	Y (8)	AT4G14570	AT4G14570	Acylamino acid-releasing enzyme
JCVI_27302:827	C3	39494864	2.450	7.94%	K (13)	T (26)	-	AT5G63840	AT5G63840	Radial swelling 3
JCVI_197:367	A10	6561639	2.434	9.38%	R (8)	A (33)	-	AT5G57660	AT5G57655	Xylose isomerase family protein
JCVI_27767:879	A7	10349313	2.407	9.10%	-	Y (23)	T (9)	AT1G22380	AT1G22360-40(	UDP-Glucosyl transferase 85A1, 2 & 3
JCVI_17664:1518	**C2**	**39674532**	**2.395**	8.94%	**Y (9)**	**C (31)**	**-**	**AT2G01320**	**AT2G01320**	**ABC-2 type transporter family protein**
JCVI_8464:784	**C3**	**16534486**	**2.390**	9.20%	**T (8)**	**Y (31)**	**-**	**AT2G28840**	**AT2G28760**	**UDP-XYL synthase 6**
JCVI_8724:1237	A3	30077326	2.376	11.13%		S (32)	G (5)	AT4G35670	AT4G35670	Pectin lyase-like superfamily protein
JCVI_11821:822	C3	4803363	2.320	9.81%	A (7)	M (29)	-	AT5G19160	AT5G19160	Trichome birefringence-like 11
JCVI_6088:571	A8	10661406	2.312	8.62%	C (9)	Y (30)	-	AT4G30160	AT4G30170	Peroxidase family protein
JCVI_33124:901	**A10**	**12590750**	**2.289**	7.41%	**A (19)**	**R (21)**	**-**	**AT5G08610**	**AT5G08460**	**GDSL-like Lipase/Acylhydrolase superfamily protein**
JCVI_7695:701	A3	14117383	2.270	7.40%	-	K (22)	G (19)	AT4G01370	AT4G01210	Glycosyltransferase family 1 protein
JCVI_13472:1244	C9	4611297	2.267	7.20%	Y (18)	T (22)	-	AT5G60920	AT5G60950	COBRA-like protein 5 precursor
JCVI_3121:213	C7	44734046	2.260	8.56%	S (15)	G (25)	-	AT4G22890	AT4G22970	Radially swollen 4
JCVI_2549:462	C6	28317182	2.251	7.48%	-	T (15)	K (25)	AT3G59920	AT3G59850	Pectin lyase-like superfamily protein
JCVI_26136:231	A9	14141369	2.249	10.01%	-	W (30)	T (6)	AT2G02080	AT2G02061	Nucleotide-diphospho-sugar transferase family protein
JCVI_11801:773	**A7**	**850820**	**2.197**	7.23%	**R (18)**	**A (23)**	**-**	**AT2G18710**	**AT2G18800**	**Xyloglucan endotransglucosylase/hydrolase 21**
JCVI_21919:1178	C3	41744982	2.196	9.93%	G (6)	R (31)	-	AT5G61810	AT5G61840	IRX10-Like - Glucuronosyltransferase
JCVI_27253:237	A1	12906362	2.166	7.86%	M (14)	C (23)	-	AT3G49800	Multiple	Carbohydrate-binding protein / glycosyl hydrolase family protein 43
JCVI_1124:154	**A7**	**20231486**	**2.145**	9.30%	**G (7)**	**S (34)**	**-**	**AT1G65930**	**AT1G65890**	**Acyl activating enzyme 12**
JCVI_1101:362	C5	24761456	2.135	8.98%	T (8)	Y (25)	-	AT1G32900	Multiple	Granule bound starch synthase 1 / HXXXD-type acyl-transferase
JCVI_38129:288	**C4**	**2285002**	**2.110**	7.41%	**-**	**Y (26)**	**T (13)**	**AT2G41790**	**-**	**-**
JCVI_17961:345	**A9**	**25881394**	**2.101**	8.44%	**-**	**C (28)**	**Y (8)**	**AT1G28150**	**AT1G28290**	**Arabinogalactan protein 31**
JCVI_37797:420	A3	8311101	2.086	7.07%	-	Y (25)	C (14)	AT2G35800	Multiple	Sucrose-6F-P phosphohydrolase FP / FASCICLIN-like arabinogalactan P16 precursor
JCVI_690:199	A8	13196172	2.078	7.37%	-	R (28)	A (13)	AT4G14300	AT4G14310	CW located Transducin/WD40 repeat-like superfamily protein
JCVI_22710:708	**A9**	**7207339**	**2.074**	8.43%	**-**	**A (31)**	**R (9)**	**AT1G58030**	**AT1G57980**	**Nucleotide-sugar transporter family protein**
JCVI_4532:234	C1	49190708	2.065	8.79%	S (8)	C (24)	-	AT3G09300	AT3G09260	ER-localised β-glucosidase
JCVI_1513:421	**C9**	**57906528**	**2.064**	6.81%	**S (18)**	**C (22)**	**-**	**AT5G14040**	**AT5G13980**	**Glycosyl hydrolase family 38 protein**
JCVI_27658:125	**C3**	**21235021**	**2.047**	8.33%	**G (10)**	**K (31)**	**-**	**AT4G04210**	**-**	**-**
JCVI_8788:1017	A1	22635931	2.041	8.23%	-	Y (28)	C (8)	AT3G18750	AT3G18660	GlcA substitution of xylan 1
JCVI_11411:748	**C3**	**60047913**	**2.031**	7.71%	**R (13)**	**A (26)**	**-**	**AT4G19006**	**AT4G18990**	**Xyloglucan endotransglucosylase/hydrolase 29**
JCVI_17664:893	A2	18739683	2.022	9.65%	T (5)	K (36)	-	AT2G01320	AT2G01320	ABC-2 type transporter family protein
JCVI_32230:734	A7	24389437	2.020	6.82%	-	Y (19)	T (19)	AT1G76790	AT1G76670	Nucleotide-sugar transporter family protein
JCVI_1880:413	A3	19772119	2.007	6.63%	-	G (20)	K (18)	AT3G25230	AT3G25140	Galacturonosyl transferase 8 - Quasimodo 1
JCVI_10933:779	A10	14332063	2.001	7.57%	-	Y (31)	T (9)	AT5G03430	-	-

*Effect size is the quantitative difference between means for each marker, expressed as % of the mean

Appendix Table 12 – RS 210 low – SNP markers found in regions of the *B. napus* transcriptome associated with the variation in RS yields released from various cultivars ($p < 0.01$) after liquid hot water pretreatment at 210 °C for 10 min, followed by hydrolysis using a low (7 FPU/g) cellulase dose.

Glucose yield 210H

Brassica napus								*Arabidopsis thaliana*		
Marker	**Position**		**-log10P**	**Effect size***	**SNP Alleles**			**Ortholog AGI**	**Candidate AGI**	**Candidate name**
	Chr	**Pseudo**			**-**	**Modal**	**+**			
JCVI_2389:1073	C4	25113320	3.943	5.26%	Y (16)	T (23)	-	AT3G63000	AT3G62830	UDP-GlcA decarboxylase 2
JCVI_430:165	A8	17643293	3.717	4.84%	A (19)	R (22)	-	AT1G09070	AT1G08990	GlcA substitution of xylan 5
JCVI_16118:186	C9	48687099	3.072	4.87%	T (10)	Y (30)	-	AT5G57710	AT5G57655	Xylose isomerase family protein
JCVI_17664:893	A2	18739683	2.967	6.35%	T (5)	K (36)	-	AT2G01320	AT2G01320	ABC-2 type transporter family protein
JCVI_14508:616	C8	26011248	2.788	4.21%	G (14)	R (25)	-	AT3G54360	Multiple	Trichrome Birefringence-like 36 or Glycoside hydrolase family 2 protein
JCVI_7375:649	A7	17971794	2.649	4.22%	C (18)	Y (24)	-	AT1G75950	AT1G75940	Similar to β-glucosidase 4 in *B. napus*
JCVI_1134:437	A10	1708228	2.604	4.62%	M (9)	C (32)	-	AT1G04940	AT1G04910	O-fucosyltransferase family protein
JCVI_12171:150	C5	37469056	2.553	3.76%	-	S (24)	C (17)	AT3G15353	AT3G15370	Alpha-Expansin Gene family protein
JCVI_8244:130	A1	1731387	2.464	4.80%	C (7)	S (34)	-	AT4G34620	AT4G34480	O-Glycosyl hydrolases family 17 protein
JCVI_40373:376	C1	11850002	2.453	6.08%	T (4)	W (35)	-	AT4G23940	Multiple	Cellulose Synthase-like G1, 2 and 3 (AT4G23990-4010)
JCVI_16555:1036	C7	38397626	2.390	3.92%	M (14)	C (25)	-	AT5G26830	AT5G26810	Pectin lyase-like superfamily protein
JCVI_1877:1331	A5	2190323	2.221	3.63%	R (15)	A (26)	-	AT2G44760	AT2G44660	ALG8 glycosyltransferase family - transfers hexose groups
JCVI_8475:133	C7	50748254	2.212	4.63%	C (9)	Y (31)	-	AT2G20930	AT2G20870	Cell wall protein precursor, putative
JCVI_1056:545	C3	58549048	2.191	4.75%	C (6)	M (34)	-	AT4G36690	AT4G36770	UDP-Glycosyltransferase superfamily
JCVI_18281:742	A3	18046669	2.184	3.92%	-	C (23)	Y (18)	AT3G18780	Multiple	Polyol/monosaccharide transporter 5/Proline-rich extensin-like receptor kinase
JCVI_5309:460	C6	9236700	2.178	6.71%	G (3)	R (38)		AT1G53210	AT1G53290	Galactosyltransferase family protein
JCVI_27578:2248	C4	6392947	2.152	3.47%	-	C (25)	M (16)	AT2G38770	Multiple	Annexins 3 & 4 / microtubule-associated protein 65-5
JCVI_965:577	C3	52235721	2.132	3.69%	-	S (30)	C (11)	AT1G23800	AT1G23760	Polygalacturonase 3 - aromatic rich glycoprotein JP630
JCVI_23018:322	C5	41226110	2.100	3.53%	-	W (22)	T (15)	AT3G09970	Multiple	GDSL-like Lipase/Acylhydrolase superfamily / Acetylated interacting1
JCVI_26272:1210	A7	22372439	2.092	3.83%	G (10)	S (30)	-	AT3G63510	AT1G71220	UDP-glucose:glycoprotein glucosyltransferase.
JCVI_28:318	C6	13486549	2.074	3.55%	A (14)	M (23)	-	AT1G55450	AT1G55450	SAM-dependent methyltransferases superfamily protein
JCVI_34809:725	A4	8253766	2.063	4.09%	A (8)	R (33)	-	AT5G39510	AT5G39790	Chloroplast localized protein with starch binding activity
JCVI_7747:305	C1	46425427	2.062	4.73%	Y (7)	T (29)	-	AT3G14067	AT3G14220	GDSL-like Lipase/Acylhydrolase superfamily protein
JCVI_32230:795	A7	24389437	2.060	3.65%	-	R (25)	G (14)	AT1G76790	AT1G76670	Nucleotide-sugar transporter family protein
JCVI_15647:325	A6	9069124	2.057	4.05%	-	C (23)	S (12)	AT1G21680	Multiple	Cell-wall localised gene / hydroxyproline-rich glycoprotein family protein
JCVI_31647:594	C1	3511759	2.052	4.11%	-	Y (27)	T (14)	AT4G32285	AT4G32272	Nucleotide-sugar transporter family protein
JCVI_21873:793	A3	14532397	2.038	4.91%	T (5)	Y (36)	-	AT3G02520	-	-
JCVI_30870:389	C4	54006138	2.023	3.45%	S (14)	G (27)	-	AT2G01250	AT2G44110	Mildew resistance locus O 15
JCVI_14140:459	C2	30027510	2.016	5.43%	T (4)	W (37)	-	AT1G78140	AT1G78140	SAM-dependent methyltransferases superfamily protein
JCVI_7088:97	A7	12376898	2.005	3.57%	-	S (24)	C (15)	AT2G27385	AT2G27380	Extensin proline-rich 1

*Effect size is the quantitative difference between means for each marker, expressed as % of the mean

Appendix Table 13 – Glc 210 High – SNP markers found in regions of the *B. napus* transcriptome associated with the variation in Glc yields released from various cultivars ($p < 0.01$) after liquid hot water pretreatment at 210 °C for 10 min, followed by hydrolysis using a high (36 FPU/g) cellulase dose.

Reducing sugar 210H

Brassica napus								*Arabidopsis thaliana*		
Marker	**Position**		**-log10P**	**Effect size***	**SNP Alleles**			**Ortholog AGI**	**Candidate AGI**	**Candidate name**
	Chr	**Pseudo**			**-**	**Modal**	**+**			
JCVI_8244:130	A1	1731387	2.072	6.03%	C (7)	S (34)	-	AT4G34620	AT4G34480	O-Glycosyl hydrolases family 17 protein
JCVI_19443:522	**A2**	**1419758**	**2.029**	**12.45%**	**-**	**G (37)**	**S (4)**	**AT5G09660**	**AT5G09700**	**Glycosyl hydrolase family 3 protein**
JCVI_17664:893	**A2**	**18739683**	**2.597**	**8.19%**	**T (5)**	**K (36)**	**-**	**AT2G01320**	**AT2G01320**	**ABC-2 type transporter family protein**
JCVI_21873:793	A3	14532397	2.505	7.96%	T (5)	Y (36)	-	AT3G02520	-	-
JCVI_11801:773	**A7**	**850820**	**2.098**	**4.69%**	**R (18)**	**A (23)**	**-**	**AT2G18710**	**AT2G18800**	**Xyloglucan endotransglucosylase/hydrolase 21**
JCVI_4152:507	**A7**	**14839212**	**2.013**	**5.52%**	**C (9)**	**M (26)**	**-**	**AT3G53970**	-	-
JCVI_1124:154	**A7**	**20231486**	**2.203**	**6.42%**	**G (7)**	**S (34)**	**-**	**ATIG65930**	**ATIG65890**	**Acyl activating enzyme 12**
JCVI_4441:401	**A8**	**12629468**	**2.075**	**4.83%**	**K (18)**	**G (20)**	**-**	**AT4G38100**	**AT4G38080**	**Hydroxyproline-rich glycoprotein family protein**
JCVI_430:165	A8	17643293	2.754	5.56%	A (19)	R (22)	-	ATIG09070	ATIG08990	GlcA substitution of xylan 5
JCVI_959:621	**A9**	**7016640**	**2.153**	**5.79%**	**-**	**A (25)**	**R (14)**	**ATIG64040**	**ATIG63910**	**MYB domain protein 103**
JCVI_1134:437	A10	1708228	2.122	5.68%	M (9)	C (32)	-	ATIG04940	ATIG04910	O-fucosyltransferase family protein
JCVI_197:367	A10	6561639	2.460	6.52%	R (8)	A (33)	-	AT5G57660	AT5G57655	Xylose isomerase family protein
JCVI_37942:1200	C1	12385276	2.341	5.21%	-	Y (17)	T (22)	AT4G24550	AT4G24530	O-fucosyltransferase family protein
JCVI_7836:240	**C1**	**18148122**	**2.293**	**8.14%**	**W (4)**	**A (37)**	**-**	**AT4G14320**	**AT4G14310**	**CW located Transducin/WD40 repeat-like superfamily protein**
JCVI_1549:404	**C1**	**36039826**	**2.124**	**5.81%**	**G (8)**	**R (33)**	**-**	**AT3G23820**	**AT3G23820**	**UDP-D-Glucuronate-4-epimerase 6**
JCVI_39979:502	C2	9600332	2.148	9.79%	-	G (32)	R (6)	AT5G57710	AT5G57655	Xylose isomerase family protein
JCVI_17664:1518	**C2**	**39674532**	**2.100**	**5.50%**	**Y (9)**	**C (31)**	**-**	**AT2G01320**	**AT2G01320**	**ABC-2 type transporter family protein**
JCVI_1821:822	C3	4803363	2.261	6.45%	A (7)	M (29)	-	AT5G19180	AT5G19160	Trichome birefringence-like 11
JCVI_8464:784	**C3**	**16534486**	**2.333**	**6.18%**	**T (8)**	**Y (31)**	**-**	**AT2G28840**	**AT2G28760**	**UDP-XYL synthase 6**
JCVI_8062:467	**C3**	**39486264**	**2.425**	**7.67%**	**T (5)**	**Y (33)**	**-**	**AT5G63860**	**AT5G63840**	**Radial swelling 3**
JCVI_21919:1178	C3	41744982	2.409	7.20%	G (6)	R (31)	-	AT5G61810	AT5G61840	IRX10-Like - Glucuronosyltransferase
JCVI_1056:545	C3	58549048	2.108	6.53%	C (6)	M (34)	-	AT4G36690	AT4G36770	UDP-Glycosyltransferase superfamily
JCVI_27578:2248	C4	6392947	2.205	4.97%	M (16)	C (25)	-	AT2G38770	Multiple	Annexins 3 & 4 / microtubule-associated protein 65-5
JCVI_2389:1073	C4	25113320	2.860	5.93%	Y (16)	T (23)	-	AT3G63000	AT3G62830	UDP-GlcA decarboxylase 2
JCVI_1336:469	**C4**	**34725981**	**2.207**	**6.70%**	**C (6)**	**Y (35)**	**-**	**AT5G37600**	**AT5G37660**	**Plasmodesmata-located protein 7**
JCVI_30870:389	C4	54006138	2.112	5.00%	S (14)	G (27)	-	AT2G01250	AT2G44110	Mildew resistance locus O 15
JCVI_1101:362	C5	24761456	2.420	6.90%	T (8)	Y (25)	-	ATIG32900	Multiple	Granule bound starch synthase 1 / HXXXD-type acyl-transferase family protein
JCVI_5309:460	C6	9236700	2.891	11.46%	G (3)	R (38)	-	ATIG53210	ATIG53290	Galactosyltransferase family protein
JCVI_28:318	C6	13486549	2.086	5.03%	A (14)	M (23)	-	ATIG55450	ATIG55450	SAM-dependent methyltransferases superfamily protein
JCVI_3399:323	**C6**	**34090389**	**2.000**	**4.62%**	**C (15)**	**Y (26)**	**-**	**ATIG73030**	**Multiple**	**Trichome birefringence-like 3 (ATIG73140) or BGAL17 (ATIG72990)**
JCVI_8475:133	C7	50748254	2.339	6.78%	C (9)	Y (31)	-	AT2G20930	AT2G20870	Cell wall protein precursor, putative
JCVI_16252:1130	C8	10965209	2.186	6.02%	-	T (32)	Y (8)	AT4G14570	AT4G14570	Acylamino acid-releasing enzyme
JCVI_1973:910	C9	48568519	2.984	6.01%	-	R (16)	A (24)	AT5G57655.2	AT5G57655	Xylose isomerase family protein

*Effect size is the quantitative difference between means for each marker, expressed as % of the mean

Appendix Table 14 – RS 210 High – SNP markers found in regions of the *B. napus* transcriptome associated with the variation in RS yields released from various cultivars ($p < 0.01$) after liquid hot water pretreatment at 210 °C for 10 min, followed by hydrolysis using a high (36 FPU/g) cellulase dose.

Formic 185

Brassica napus								*Arabidopsis thaliana*		
Marker	**Position**		**-log10P**	**Effect size***	**SNP Alleles**			**Orthолog AGI**	**Candidate AGI**	**Candidate name**
	Chr	**Pseudo**			**-**	**Modal**	**+**			
JCVI_690:199	A8	13196172	4.070	20.48%	-	R (28)	A (13)	AT4G14300	AT4G14310	CW located Transducin/WD40 repeat-like superfamily protein
JCVI_17322:163	C3	56294303	3.481	18.11%	-	T (28)	Y (13)	AT4G14230	AT4G14310	CW located Transducin/WD40 repeat-like superfamily protein
JCVI_1535:492	C9	24392590	3.072	22.65%	-	Y (37)	T (4)	AT5G47120	Multiple	Low-level beta-amylase 1 (AT5G47010) or Peroxidase (AT5G47000)
JCVI_33410:400	A6	4824152	2.830	16.44%	-	C (30)	S (11)	AT1G12800	AT1G12780	UDP-D-Glc/UDP-D-Gal 4-Epimerase 1
JCVI_23579:1248	C4	50409451	2.715	15.26%	R (20)	G (21)	-	AT1G56070	AT1G56010	NAC domain containing protein 21
JCVI_36229:32	C1	3956907	2.701	17.02%	-	M (29)	C (9)	AT4G31390	AT4G31370	Fasciclin-like arabinogalactan protein 5 precursor
JCVI_1727:391	A9	39866223	2.621	20.59%	-	R (36)	A (5)	AT1G07890	AT1G07890	Ascorbate peroxidase 1
JCVI_15537:707	C2	43992872	2.600	20.77%	-	C (33)	S(4)	AT3G27770	AT3G27785	MYB 118
JCVI_33171:965	A6	25843147	2.559	26.07%	-	M (35)	C (3)	AT4G38250	AT4G38270	Galacturonosyltransferase 3
JCVI_27400:747	A5	3820801	2.545	28.93%	-	R (38)	A (2)	AT2G37710	AT2G37720	Trichome birefringence-like 15
JCVI_3790:218	A2	24353905	2.499	21.18%	-	S (37)	G (4)	AT3G48930	AT5G23940	Permeable Leaves 3 (PEL3), a putative acyl-transferase
JCVI_4122:134	C3	62090678	2.473	21.06%	-	K (35)	T (4)	AT4G12590	AT4G12730	Fasciclin-like arabinogalactan 2
JCVI_32663:657	C4	30898557	2.451	17.62%	-	T (34)	W (6)	AT3G53020	Multiple	D-galactoside/L-rhamnose binding SUEL lectin proteins (AT3G53050,65,75)
JCVI_18808:342	A2	7352464	2.348	12.51%	-	R (24)	G (15)	AT5G53170	Multiple	Arabinogalactan protein 22 (AT5G53250) or pectin methylesterase (AT5G53370)
JCVI_3687:386	A5	7753285	2.327	15.21%	C (20)	S (21)	-	AT3G22630	AT3G22560	Acyl-CoA N-acyltransferases (NAT) superfamily protein
JCVI_41018:1843	A3	25658592	2.304	14.58%	-	W (26)	A (15)	AT4G24620	AT4G24620	Phosphoglucose isomerase 1
JCVI_8724:1237	A3	30077326	2.229	17.49%	-	S (32)	G (5)	AT4G35790	AT4G35670	Pectin lyase-like superfamily protein
JCVI_33410:98	C5	5199859	2.221	13.55%	-	G (29)	R (12)	AT1G12800	AT1G12780	UDP-D-Glc/UDP-D-Gal 4-Epimerase 1
JCVI_3493:636	C4	3501308	2.204	18.91%	-	R (34)	G (4)	AT2G43790	Multiple	UDP-glycosyltransferase 74 F1 & F2 (AT2G43820-40)
JCVI_13548:1330	A2	10948166	2.202	15.05%	-	G (22)	S (17)	AT1G70770	AT1G15950	Cinnamoyl CoA reductase 1
JCVI_31241:1377	C5	37879153	2.178	14.11%	-	T (20)	K (18)	AT1G65930	AT1G65890	Acyl activating enzyme 12
JCVI_866:355	C7	35386139	2.173	12.54%	W (15)	T (26)	-	AT3G27830	AT3G27810	MYB 21
JCVI_14852:377	C1	8542973	2.163	14.66%	-	K (30)	T (11)	AT4G20360	AT4G20270	Barely any meristem 3
JCVI_14456:130	C9	4637409	2.161	19.70%	-	Y (35)	C (3)	AT5G61050	AT5G60950	COBRA-LIKE protein 5 precursor
JCVI_594:246	C7	47157643	2.154	19.34%	-	K (35)	G (3)	AT4G29160	AT4G29130	Hexsokinase 1 (Glc insensitive 2)
JCVI_37397:429	A6	9622427	2.137	16.70%	-	Y (31)	T (6)	AT3G49060	Multiple	Peroxidase 33-4 (AT3G49110-20) or O-acyltransferase (WSD1-like)(AT3G49210)
JCVI_18843:1370	A3	28023029	2.128	13.03%	-	C (27)	M (12)	AT4G30810	AT4G30760	Putative endonuclease or glycosyl hydrolase
JCVI_2306:581	A2	4561625	2.119	17.51%	C (19)	Y (22)	-	AT5G22800	AT5G22740	Cellulose synthase-like A2 (β-mannan synthase)
JCVI_10607:386	C5	19058620	2.101	14.16%	-	R (31)	G (9)	AT1G24764	AT1G24764	Microtubule associated proteins 70-2
JCVI_18178:985	C9	61769295	2.047	13.35%	-	R (26)	A (13)	AT5G05780	AT5G05820	Nucleotide-sugar transporter family protein
JCVI_970:133	A9	7109480	2.035	19.10%	-	S (37)	C (4)	AT1G64355	Multiple	UDP-Glc 4 Epimerase (AT1G64440) OR Glycosyl hydrolase 9C2 (AT1G64390)
JCVI_21659:310	C7	50071115	2.007	17.48%	-	K (30)	G (6)	AT4G37470	Multiple	Arabinogalactan protein 18 (AT4G37450) OR Peroxidases (AT4G37520-30)
JCVI_7947:328	A5	16512750	2.002	15.03%	-	T (35)	W (5)	AT3G20230	AT3G20220	SAUR-like auxin-responsive protein family

*Effect size is the quantitative difference between means for each marker, expressed as % of the mean

Appendix Table 15 – 185 Formic – SNP markers found in regions of the *B. napus* transcriptome associated with the variation in Formic acid released from various cultivars ($p < 0.01$) after liquid hot water pretreatment at 185 °C for 10 min.

Formic 210												
Brassica napus										*Arabidopsis thaliana*		
Marker	Position			-log10P	Effect size*	SNP Alleles				Ortholog AGI	Candidate AGI	Candidate name
	Chr	Pseudo				-	Modal	+				
JCVI_11973:810	C6	9478196		2.479	18.82%	-	C (34)	Y (7)		AT1G79930	AT1G79915	Putative methyltransferase family protein
JCVI_18327:25	C6	29874847		2.465	17.60%	-	W (31)	A (9)		AT3G45140	AT3G45230	Arabinoxylan pectin arabinogalactan protein 1
JCVI_60:101	C3	78127		2.342	9.25%	R (12)	G (29)	-		AT5G01600	AT5G01620	Trichome birefringence-like 35
JCVI_11057:337	C3	66259901		2.263	13.61%	C (19)	Y (20)	-		AT4G30010	AT4G30060	Core-2/1-branching beta-1,6-N-acetylglucosaminyltransferase family protein
JCVI_33856:335	A3	20094135		2.250	15.82%	Y (10)	T (30)	-		AT2G07050	AT2G07050	Cycloartenol synthase 1 (Brassinosteroid)
JCVI_6982:148	A6	22186123		2.230	14.01%	C (15)	Y (25)	-		AT3G27090	Multiple	SAM-dependent methyltransferase (AT3G27180), Actin related 2 (AT3G27000)
JCVI_5639:397	A10	8937438		2.183	8.86%	T (11)	K (29)	-		AT5G20280	AT5G20280	Sucrose-phosphate synthase A1
JCVI_188:30	C5	1699147		2.182	16.40%	-	A (34)	W (7)		AT1G05190	Multiple	Galactosyltransferase family protein / Auxin resistant / Peroxidase FP
JCVI_13151:483	C7	50604853		2.160	8.89%	-	Y (29)	C (11)		AT4G39040	AT4G38990-90	Glycosyl hydrolase 9B16,17 & 18,
JCVI_27274:181	A6	25327780		2.145	9.33%	T (7)	Y (32)	-		AT5G43260	-	-
JCVI_35779:1319	A9	25716747		2.136	14.40%	R (27)	A (11)	-		AT1G28320	AT1G28290	Arabinogalactan protein 31
JCVI_9581:491	C2	4617412		2.110	20.73%	S (5)	G (36)	-		AT5G17710	AT5G17650	Glycine/proline-rich protein
JCVI_11826:418	C5	41165479		2.102	13.45%	-	Y (23)	C (17)		AT3G10230	AT3G10200	SAM-dependent methyltransferases superfamily protein
JCVI_13751:354	C1	42125700		2.096	16.12%	Y (11)	C (25)	-		.	-	-
JCVI_463:255	C2	7408931		2.088	14.97%	G (10)	R (31)	-		AT5G60800	AT5G60810	Root meristem growth factor 1
JCVI_18283:236	A4	19749486		2.066	12.77%	-	W (24)	T (16)		AT2G45770	AT2G45750	SAM-dependent methyltransferases superfamily protein
JCVI_26090:510	A8	11707460		2.047	15.12%	A (19)	W (31)	-		AT4G24550	AT4G24530	O-fucosyltransferase family protein
JCVI_26935:262	A2	22907225		2.026	8.30%	W (15)	A (26)	-		.	-	-
JCVI_12639:378	C6	13263637		2.025	9.00%	T (6)	W (33)	-		AT1G56050	AT1G56100	Plant invertase/pectin methylesterase inhibitor superfamily protein
JCVI_5606:472	C9	3115516		2.017	8.37%	T (17)	Y (23)	-		AT5G27990	-	-
JCVI_11348:376	A6	5789693		2.003	8.22%	R (15)	A (25)	-		AT1G14790	AT1G14720	Xyloglucan endotransglucosylase/hydrolase 28
JCVI_387:458	C9	61422675		2.002	22.57%	-	Y (36)	C (3)		AT5G06340	AT5G06390	Fasciclin-like arabinogalactan protein 17 precursor
JCVI_4590:359	A4	5135317		2.002	8.70%	A (8)	W (32)	-		AT4G14070	AT4G14130	Xyloglucan endotransglycosylase-related protein (XTR7)

*Effect size is the quantitative difference between means for each marker, expressed as % of the mean

Appendix Table 16 – 210 Formic – SNP markers found in regions of the *B. napus* transcriptome associated with the variation in Formic acid released from various cultivars ($p < 0.01$) after liquid hot water pretreatment at 210 °C for 10 min.

Acetic 185

Brassica napus								*Arabidopsis thaliana*		
Marker	**Position**		**-log10P**	**Effect size***	**SNP Alleles**			**Ortholog AGI**	**Candidate AGI**	**Candidate name**
	Chr	**Pseudo**			**-**	**Modal**	**+**			
JCVI_17575:307	A8	12639897	3.243	6.82%	R (16)	G (22)	-	AT2G22480	AT2G22480	Phosphofructokinase 5
JCVI_21635:60	C3	24173153	3.110	6.55%	-	S (33)	C (8)	AT3G06700	AT3G06690	Acyl-CoA oxidases;oxidoreductases
JCVI_21635:208	A3	15363737	3.083	6.66%	-	M (35)	A (5)	AT3G06700	AT3G06690	Acyl-CoA oxidases;oxidoreductases
JCVI_33362:702	C1	48551265	3.052	6.33%	-	G (29)	K (7)	AT3G10770	Multiple	α-L-Arabinofuranosidase 1 (AT3G10740), Pectin ME inhibitor (AT3G10720)
JCVI_29823:897	A9	36085419	2.904	6.12%	Y (16)	C (25)	-	AT2G22125	AT2G22125	Cellulose synthase-interactive protein 1 (POM-POM2)
JCVI_1691:189	A3	29761893	2.758	6.09%	-	M (36)	A (5)	AT4G35000	Multiple	MYB 32 (AT4G34990) or β-Galactosidase 11 (AT4G35010)
JCVI_15537:707	C2	43992872	2.747	6.87%	-	C (33)	S (4)	AT3G27770	AT3G27785	MYB 118
JCVI_3790:218	A2	24353905	2.628	6.94%	-	S (37)	G (4)	AT3G48930	Multiple	ACYL-activating enzyme3(AT3G48990),MYB45(AT3G48920),Pectin lyase-L(AT3G48950)
JCVI_7947:412	A5	16512750	2.593	5.66%	-	K (36)	G (5)	AT3G20230	AT3G20220	SAUR-like auxin-responsive protein family
JCVI_7731:258	A3	27727635	2.545	6.15%	-	K (33)	G (5)	AT4G30310	Multiple	Xyloglucan endotransglucosylase/hydrolase 18, 19, 24 (AT4G30280, 90, 70)
JCVI_26405:493	C8	19594697	2.471	8.07%	-	M (37)	A (2)	AT1G16860	-	-
JCVI_24174:714	A5	9185480	2.437	4.38%	R (19)	G (22)	-	AT1G53430	AT1G53500	Rha biosynthesis 2
JCVI_2305:176	A2	17934710	2.394	4.77%	-	A (28)	W (12)	AT5G47030	Multiple	Low-level beta-amylase 1 (AT5G47010) or Peroxidase (AT5G47000)
JCVI_1808:350	C3	69099660	2.305	5.27%	-	R (30)	G (11)	AT1G50010	AT1G50000	Methyltransferase
JCVI_1535:492	C9	24392590	2.300	6.05%	-	Y (37)	T (4)	AT5G47120	Multiple	Low-level beta-amylase 1 (AT5G47010) or Peroxidase (AT5G47000)
JCVI_29646:408	C8	35279774	2.276	4.52%	-	R (28)	G (13)	AT1G14040	Multiple	[Xyloglucan] Fucosyltransferase FUT6 and 8 (AT1G14080 and 100)
JCVI_1205:232	A5	14620647	2.254	4.29%	-	Y (27)	T (12)	AT3G22845	-	-
JCVI_3196:355	C6	16580721	2.239	4.88%	-	A (28)	M (10)	AT1G51760	AT1G51770	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein
JCVI_28:289	C6	13486549	2.211	5.20%	C (15)	Y (22)		AT1G55450	AT1G55450	SAM-dependent methyltransferases superfamily protein
JCVI_20453:109	A7	21585512	2.207	4.69%	-	A (32)	W (8)	AT1G69523	AT1G69523	SAM-dependent methyltransferases superfamily protein
JCVI_33410:400	A6	4824152	2.195	4.57%	-	C (30)	S (11)	AT1G12800	AT1G12780	UDP-D-Glc/UDP-D-Gal 4-Epimerase 1
JCVI_7765:772	C4	54480036	2.180	7.26%	S (5)	G (33)	-	AT2G45810	AT2G45750	SAM-dependent methyltransferases superfamily protein
JCVI_1951:373	C8	32857768	2.175	5.48%	C (19)	Y (21)	-	AT2G21970	AT2G21890	Cinnamyl alcohol dehydrogenase homolog 3
JCVI_6969:143	C1	22564838	2.169	4.19%	-	R (26)	G (15)	AT4G12730	AT4G12730	Fasciclin-like arabinogalactan 2
JCVI_23807:690	C4	18765140	2.165	5.31%	-	Y (22)	T (17)	AT2G26730	AT2G26810	Putative methyltransferase family protein
JCVI_14456:130	C9	4637409	2.136	6.17%	-	Y (35)	C (3)	AT5G61050	AT5G60950	COBRA-LIKE protein 5 precursor
JCVI_17994:658	A3	10613463	2.105	4.21%	A (20)	T (25)	-	AT2G43950	Multiple	Pectin lyase-like proteins (AT2G43840,60,70,80,90)
JCVI_16855:508	A10	14313294	2.104	4.55%	-	T (26)	Y (13)	AT5G03540	Multiple	GDSL-L acylhydrolases(AT5G03590,610),UDP-GT (AT5G03490),Mediator7(AT5G03500)
JCVI_21625:1197	A8	16372024	2.101	4.80%	-	S (28)	C (11)	AT1G15500	-	-
JCVI_316:281	A6	9716013	2.093	5.54%	-	S (34)	C (5)	AT3G48890	Multiple	Acyl-activating enzyme 3 (AT3G48990),MYB45(AT3G48920),Pectin lyase-L(AT3G48950)
JCVI_808:590	A8	1332225	2.070	4.19%	-	R (29)	G (12)	AT1G52280	AT1G52290	Proline-rich extensin-like receptor kinase 15
JCVI_36464:332	A6	6640125	2.055	4.34%	Y (15)	C (26)	-	AT1G78570	AT1G78570	Rha biosynthesis 1
JCVI_23343:1158	A6	2061238	2.050	4.36%	-	M (27)	C (11)	AT1G50380	Multiple	Plant invertase/pectin methylesterase inhibitor (AT1G50325, 340)
JCVI_36229:32	C1	3956907	2.035	4.49%	-	M (29	C (9)	AT4G31390	AT4G31370	Fasciclin-like arabinogalactan protein 5 precursor
JCVI_4300:221	C4	30622924	2.030	3.88%	R (19)	A (20)	-	AT3G53430	-	-

*Effect size is the quantitative difference between means for each marker, expressed as % of the mean

Appendix Table 17 – 185 Acetic – SNP markers found in regions of the *B. napus* transcriptome associated with the variation in Acetic acid released from various cultivars ($p < 0.01$) after liquid hot water pretreatment at 185 °C for 10 min.

Acetic 210

Brassica napus									*Arabidopsis thaliana*		
Marker	**Position**		**-log10P**	**Effect size***	**SNP Alleles**				**Ortholog AGI**	**Candidate AGI**	**Candidate name**
	Chr	**Pseudo**			**-**	**Modal**		**+**			
JCVI_38459:264	A3	10748085	3.013	8.16%	-	M (23)	C (17)		AT3G60320	AT2G40360	CW located Transducin/WD40 repeat-like superfamily protein
JCVI_430:855	C8	16519130	2.797	6.97%	Y (19)	T (22)	-		AT1G09070	AT1G08990	Glucuronic acid substitution of xylan 5
JCVI_1747:141	A5	1916580	2.594	6.98%	M (16)	A (25)	-		AT2G44120	AT2G44110	Mildew resistant locus O 15
JCVI_12388:669	C6	21699475	2.582	10.31%	-	C (36)	Y (4)		AT4G02570	Multiple	Auxin resistant6(AT4G02570),Mildew resistant LO1(2600),UDP-Xylosyltransferase2(2500)
JCVI_14792:310	C7	39199379	2.519	7.57%	-	K (25)	T (13)		AT5G24930	AT5G24860	Flowering promoting factor 1
JCVI_4049:925	C5	15865450	2.517	7.70%	R (9)	G (31)	-		AT1G22850	AT1G22880	Cellulase 5
JCVI_4189:107	C3	66525453	2.480	7.51%	Y (10)	C (31)	-		AT4G29410	AT4G29360	O-Glycosyl hydrolases family 17 protein
JCVI_1155:311	C8	14000213	2.403	6.10%	-	S (30)	C (11)		AT1G03630	AT1G03520	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein
JCVI_11874:1431	A8	10759399	2.349	7.39%	G (9)	S (32)	-		AT4G29840	AT4G29940	Pathogenesis related homeodomain protein A
JCVI_5929:368	C9	11666794	2.344	7.36%	-	S (22)	C (19)		AT1G63980	Multiple	MYB 103 (AT1G63910), WRKY DNA-binding protein 56 (AT1G64000)
JCVI_10461:54	C6	39782483	2.269	5.92%	-	A (23)	R (17)		AT1G69740	AT1G69730	Wall-associated kinase family protein
JCVI_8550:1769	C8	10965216	2.242	7.11%	-	Y (28)	T (11)		AT4G14570	AT4G14570	Acylamino acid-releasing enzyme
JCVI_10843:987	A4	7834446	2.203	5.44%	-	K (24)	T (15)		AT3G14990	AT3G14960	Galactosyltransferase family protein
JCVI_331:404	A8	17492028	2.198	6.64%	C (13)	Y (26)	-		AT1G09780	AT1G09790	COBRA-Like protein 6 precursor 6
JCVI_20882:881	A6	2221366	2.184	6.79%	-	Y (22)	C (16)		AT1G49600	Multiple	Fucosyltransferase 12(AT1G49710),Carboxyesterase 5(AT1G49660),PRX (AT1G49570)
JCVI_975:917	A9	38970664	2.162	6.70%	-	C (24)	M (17)		AT1G11860	AT1G11820	O-Glycosyl hydrolases family 17 protein
JCVI_31958:465	C9	3617245	2.140	7.93%	-	S (29)	G (8)		AT5G26570	Multiple	Sugar transport protein 13 (AT5G26340),Fasciclin-like arabinogalactan protein(AT5G26730)
JCVI_7731:471	C7	47679404	2.129	6.93%	A (8)	R (32)	-		AT4G30310	Multiple	Xyloglucan endotransglucosylase/hydrolase 18, 19, 24 (AT4G30280, 290, 270)
JCVI_11769:728	A8	14645988	2.125	6.27%	-	T (27)	Y (9)		AT1G26830	Multiple	SAM-dependent MT, GalactosylT 1, MYB 117, Expansin A10 (AT1G26850,810,780,770)
JCVI_19063:1108	A9	4199045	2.122	7.09%	-	W (25)	T (16)		AT5G64860	AT5G64860	Disproportionating enzyme (Altered starch degradation)
JCVI_1792:284	A5	18833872	2.043	7.22%	G (7)	R (34)	-		AT3G16190	AT3G16170	Acyl activating enzyme 13
JCVI_23082:139	C1	11823666	2.039	9.09%	A (4)	W (34)	-		AT4G23890	AT4G23850	Long-chain acyl-CoA synthetase 4
JCVI_20728:195	A6	16507292	2.024	7.91%	-	R (36)	G (5)		AT5G64140	AT5G64120	Peroxidases (AT5G64100, 110, 120)
JCVI_13717:615	C5	11357866	2.019	6.65%	C (8)	Y (29)	-		AT1G21000	AT1G21070	Phosphofructokinase family protein
JCVI_17783:975	C9	54573560	2.018	6.24%	T (19)	K (20)	-		AT5G19290	-	-
JCVI_2233:443	A1	2600574	2.017	5.48%	Y (19)	C (22)	-		AT4G32530	-	-
JCVI_38462:513	C4	23756135	2.015	7.32%	-	R (34)	Y (7)		AT1G79930	AT3G55970	Jasmonate regulated gene 21

*Effect size is the quantitative difference between means for each marker, expressed as % of the mean

Appendix Table 18 – 210 Acetic – SNP markers found in regions of the *B. napus* transcriptome associated with the variation in Acetic acid released from various cultivars ($p < 0.01$) after liquid hot water pretreatment at 210 °C for 10 min.

2FA 185

Brassica napus									*Arabidopsis thaliana*		
Marker	Position		-log10P	Effect size*	SNP Alleles				Ortholog AGI	Candidate AGI	Candidate name
	Chr	Pseudo			-	Modal	+				
JCVI_690:199	A8	13196172	3.836	26.19%	-	R (28)	A (13)		AT4G14300	AT4G14310	CW located Transducin/WD40 repeat-like superfamily protein
JCVI_17322:163	C3	56294303	3.183	22.50%	-	T (28)	Y (13)		AT4G14230	AT4G14310	CW located Transducin/WD40 repeat-like superfamily protein
JCVI_3426:726	C5	1111419	2.751	24.86%	-	T (32)	Y (7)		AT1G03870	AT1G03860	Fasciclin-like arabinogalactan 9
JCVI_915:468	C3	58759468	2.744	22.97%	Y (19)	T (21)	-		AT4G36130	AT4G36220	Ferulic acid 5-hydroxylase 1
JCVI_23579:1248	C4	50409451	2.694	19.72%	R (20)	G (21)	-		AT1G56070	AT1G56010	NAC domain containing protein 21
JCVI_21966:180	A6	22730409	2.445	18.95%	C (13)	S (26)	-		AT3G25690	Multiple	Inflorescence Deficient in Abscission (AT3G25655) / Devil 6 (AT3G25717)
JCVI_1939:193	A6	7082612	2.421	21.28%	-	A (23)	M (14)		AT1G18210	Multiple	Glycosyl hydrolase family 81 (AT1G18310), Laccase 1 (AT1G18140)
JCVI_14757:154	C5	41671574	2.389	22.33%	-	Y (34)	T (7)		AT3G08900	AT3G08900	Reversibly glycosylated polypeptide 3
JCVI_33410:98	C5	5199859	2.379	18.63%	-	G (29)	R (19)		AT1G12800	AT1G12780	UDP-D-Glc/UDP-D-Gal 4-Epimerase 1
JCVI_31647:483	C1	3511759	2.373	7.29%	T (12)	Y (27)	-		AT4G32285	AT4G32272	Nucleotide-sugar transporter family protein
JCVI_24738:1195	C9	5629933	2.372	35.20%	-	C (33)	Y (3)		AT5G08570	-	-
JCVI_11874:1083	A8	10759399	2.325	18.95%	-	C (18)	Y (18)		AT4G29840	AT4G29940	Pathogenesis related homeodomain protein A
JCVI_26440:1691	A9	5185571	2.278	19.23%	R (13)	G (27)	-		AT2G15620	AT2G15560	Putative endonuclease or glycosyl hydrolase
JCVI_12470:294	A5	22256454	2.255	17.69%	Y (20)	T (21)	-		AT3G06980	AT3G07010	Pectin lyase-like superfamily protein
JCVI_33410:400	A6	4824152	2.253	18.27%	-	C (30)	S (11)		AT1G12800	AT1G12780	UDP-D-Glc/UDP-D-Gal 4-Epimerase 1
JCVI_6798:393	A1	6437596	2.227	18.02%	-	K (19)	G (17)		AT4G22000	Multiple	Pectin lyase-like superfamily protein (AT4G22090) or βG-2 (AT4G22100)
JCVI_2728:144	A7	156595	2.220	26.22%	-	G (35)	S (4)		AT5G48930	AT5G48930	Hydroxycinnamoyl-CoA Shikimate/Quinate hydroxycinnamoyl transferase (HCT)
JCVI_34410:543	A9	7050675	2.211	20.84%	-	T (30)	Y (10)		AT1G64190	-	-
JCVI_32920:560	A3	27098703	2.207	17.24%		C (26)	Y (15)		AT4G28880	AT4G28850	Xyloglucan endotransglucosylase/hydrolase 26
JCVI_18767:345	C8	31318237	2.206	23.51%		R (33)	G (5)		AT2G25170	AT1G07260	UDP-glucosyl transferase 71C3
JCVI_4926:148	C6	29974368	2.205	20.21%	G (7)	S (33)	-		AT1G79870	AT1G79915	Putative methyltransferase family protein
JCVI_31357:1009	C5	14147669	2.202	32.96%	W (3)	A (38)	-		AT5G09660	AT5G09730	β-Xylosidase 3
JCVI_12228:617	C9	60646527	2.201	33.19%	-	A (35)	R (3)		AT5G08130	AT5G08000	Plasmodesmata Callose-Binding 2
JCVI_28444:638	C3	24863641	2.195	15.91%	T (19)	Y (22)	-		AT3G09200	AT3G09230	MYB 1
JCVI_2438:100	A3	2173208	2.191	22.21%	-	C (28)	Y (10)		AT5G12040	AT5G11990	Proline-rich family protein
JCVI_17160:499	A9	38255886	2.190	34.57%	-	R (38)	G (3)		AT1G14000	AT1G14020	O-fucosyltransferase family protein
JCVI_9077:659	A1	10514643	2.173	19.74%	C (16)	Y (35)	-		AT4G16490	-	-
JCVI_39792:615	A6	24899136	2.150	22.47%	A (5)	R (36)	-		AT1G54040	AT1G54000-20	GDSL-like Lipase/Acylhydrolase superfamily proteins
JCVI_9947:506	A2	8667979	2.126	25.25%	-	A (27)	R (10)		AT1G66070	-	-
JCVI_1244:202	A8	17493042	2.118	17.54%	-	G (26)	R (15)		AT1G09780	AT1G09790	COBRA-LIKE protein 6 precursor 6
JCVI_1416:719	A7	17297465	2.079	22.72%	-	Y (30)	T (9)		AT1G79230	-	-
JCVI_6450:532	A9	1908341	2.044	19.80%	-	A (29)	W (8)		AT5G49460	Multiple	β-Xylosidase 1 (AT5G49360)
JCVI_1063:577	C3	61681227	2.037	18.08%	W (16)	A (22)	-		AT4G21090	AT4G21060	Arabinogalactan protein galactosyltransferase 2
JCVI_2806:473	C9	53905235	2.029	24.20%		A (36)	R (5)		AT5G20290	Multiple	Sucrose phosphate synthase 1F (AT5G20280)
JCVI_32624:626	C8	984738	2.027	21.49%	A (4)	R (37)	-		AT1G49750	AT1G49710	Fucosyltransferase 12

*Effect size is the quantitative difference between means for each marker, expressed as % of the mean

Appendix Table 19 – 185 2FA – SNP markers found in regions of the *B. napus* transcriptome associated with the variation in 2FA released from various cultivars ($p < 0.01$) after liquid hot water pretreatment at 185 °C for 10 min.

2FA 210

Brassica napus								*Arabidopsis thaliana*		
Marker	Position		-log10P	Effect size*	SNP Alleles			Ortholog AGI	Candidate AGI	Candidate name
	Chr	Pseudo			-	Modal	+			
JCVI_2276:286	C4	44966418	2.812	18.6%	G (17)	K (24)	-	AT2G28190	AT2G28110	Irregular Xylem 7
JCVI_7726:122	A10	12226159	2.720	27.9%	-	R (31)	A (6)	AT5G65270	AT5G10280	MYB 92
JCVI_26090:510	A8	11707460	2.581	45.6%	-	W (31)	A (9)	AT4G24550	AT4G24530	O-fucosyltransferase family protein
JCVI_9783:671	C9	62422544	2.515	27.3%	-	R (28)	G (7)	AT5G04360	AT5G04360	Pullunase 1 (Starch mobilisation)
JCVI_7446:301	C2	33724506	2.490	28.5%	-	Y (32)	C (5)	AT4G02500	AT4G02500	UDP-Xylosyl transferase 2
JCVI_1853:134	C9	56998636	2.469	23.1%	C (7)	S (35)	-	AT5G15780	AT4G17910	transferases, transferring acyl groups
JCVI_99:284	C8	20843738	2.468	18.1%	G (16)	R (25)	-	AT1G19150	AT1G19170	Pectin lyase-like superfamily protein
JCVI_17859:350	A3	20349852	2.338	20.3%	C (9)	Y (30)	-	AT2G14740	AT2G14610	Pathogenesis-related gene 1
JCVI_17181:1264	C6	44656345	2.336	19.3%	-	Y (23)	C (17)	AT1G75710	AT1G75680	Glycosyl hydrolase 9B7
JCVI_2675:421	C3	3169994	2.291	19.3%	-	Y (23)	C (13)	AT5G14030	AT5G13980	Glycosyl hydrolase family 38 protein
JCVI_12916:565	A5	5087978	2.289	19.6%	C (9)	M (32)	-	AT2G35040	AT2G35020	N-Acetylglucsamine-1-P Uridylyltransferase 2
JCVI_33892:595	C3	61274085	2.284	16.9%	-	M (23)	A (16)	AT4G12770	Multiple	Glycine rich protein (AT4G21620), Tubulin 9 (AT4G20890) or UGE in root (AT4G21620)
JCVI_22006:643	C3	23129750	2.273	19.2%	-	C (29)	S (12)	AT3G03960	AT3G03780	Methionine synthase 2
JCVI_34446:631	A10	1055581	2.249	18.0%	Y (19)	C (20)	-	AT1G03590	AT1G03520	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein
JCVI_18922:422	C2	50571586	2.221	17.9%	-	T (20)	Y (18)	AT5G63790	AT5G63800	β-galactosidase 6
JCVI_30541:278	A2	352536	2.215	18.3%	R (18)	A (20)	-	AT5G03040	AT5G03170	FASCICLIN-Llike arabinogalactan-protein 11
JCVI_11437:817	C3	56811248	2.203	18.2%	-	R (25)	A (13)	AT4G39170	-	-
JCVI_1567:347	C6	39878147	2.200	24.3%	-	R (36)	A (5)	AT1G14400	AT1G69810	WRKY DNA-binding protein 36
JCVI_874:330	A9	24250409	2.196	19.2%	G (9)	S (32)	-	AT1G30630	AT1G30620	UDP-D-XYL 4-Epimerase 1
JCVI_2610:386	C4	32654859	2.179	23.7%	-	Y (36)	C (5)	AT4G13930	AT4G13840	HXXXD-type acyl-transferase family protein
JCVI_38459:264	A3	10748085	2.173	16.7%	-	M (23)	C (17)	AT3G60320	AT2G40360	CW located Transducin/WD40 repeat-like superfamily protein
JCVI_3975:425	C9	53689681	2.168	19.1%	-	M (26)	C (13)	AT5G20650	AT5G20680	Trichome birefringence-like 16
JCVI_6227:310	A4	16069923	2.166	17.2%	-	C (19)	Y (16)	AT2G33730	AT2G33670	Mildew Resistance Locus O5
JCVI_1155:575	A8	18579429	2.140	31.0%	-	C (38)	Y (3)	AT1G03630	AT1G03520	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein
JCVI_2215:170	A5	2993989	2.083	29.0%	K (3)	G (36)	-	AT2G39800	AT2G39750	SAM-dependent methyltransferases superfamily protein
JCVI_31610:660	C4	8142974	2.083	18.7%	-	G (27)	R (10)	AT2G36970	AT2G36970	UDP-Glycosyltransferase superfamily protein
JCVI_5573:27	A7	23880107	2.071	16.5%	-	K (21)	G (20)	AT1G75690	AT1G75680	glycosyl hydrolase 9B7
JCVI_6656:281	C1	36465929	2.070	20.7%	T (7)	Y (33)	-	AT3G23600	AT3G24180	β-glucosidase, GBA2 type family protein;
JCVI_4482:803	C3	70256376	2.058	17.8%	-	A (26)	R (15)	AT1G51980	AT3G16470	Jasmonate responsive 1
JCVI_16847:161	C7	33723111	2.042	25.8%	-	K (37)	T (4)	AT3G25690	Multiple	Inflorescence Deficient in Abscission (AT3G25655) / Devil 6 (AT3G25717)
JCVI_12567:887	C8	24628124	2.034	19.3%	-	Y (31)	C (10)	AT3G52180	AT3G52180	Starch-excess 4 / SAM-dependent methyltransferases superfamily (AT3G52210)
JCVI_676:127	A7	20856365	2.032	23.7%	-	R (32)	G (9)	AT1G67740	AT1G67750	Pectate lyase family protein
JCVI_26302:478	A5	19989374	2.031	16.4%	C (14)	M (26)	-	AT3G13750	AT3G13750	β-galactosidase 1
JCVI_131:510	A1	25339969	2.003	17.8%	A (12)	R (22)	-	AT3G14200	AT3G14205	Suppressor of Actin 2
JCVI_131:523	C1	46397022	2.003	17.8%	C (12)	Y (22)	-	AT3G14200	AT3G14205	Suppressor of Actin 2

*Effect size is the quantitative difference between means for each marker, expressed as % of the mean

Appendix Table 20 – 210 2FA – SNP markers found in regions of the *B. napus* transcriptome associated with the variation in 2FA released from various cultivars ($p < 0.01$) after liquid hot water pretreatment at 210 °C for 10 min.

HMF 185

Brassica napus								*Arabidopsis thaliana*		
Marker	Position		-log10P	Effect size*	SNP Alleles			Ortholog AGI	Candidate AGI	Candidate name
	Chr	Pseudo			-	Modal	+			
JCVI_1200:441	C9	56825835	3.057	83.3%	-	R (38)	G (1)	AT5G16030	Multiple	TBL 19, 21 (AT5G15890-900)
JCVI_33922:193	A8	11510806	3.030	82.8%	-	Y (38)	C (1)	AT4G25640	AT4G25620	Hydroxyproline-rich glycoprotein family protein
JCVI_308:263	C8	25089768	2.960	80.8%	-	M (40)	C (1)	AT3G52800	AT3G52840	Beta-galactosidase 2
JCVI_41807:254	A8	17341483	2.945	80.7%	-	Y (39)	T (1)	AT1G10470	Mulitple	UDP-Glycosyltransferase (AT1G10400) OR XG:Xyloglucosyltransferase 33 (AT1G10550)
JCVI_25085:69	C2	1218598	2.939	80.6%	-	K (39)	G (1)	AT5G05170	AT5G05170	Cellulose synthase 3
JCVI_3716:282	A7	12305247	2.933	80.7%	-	C (38)	Y (1)	AT2G27100	AT2G27100	Serrate
JCVI_29426:694	C5	36033682	2.920	82.3%	-	R (33)	G (1)	AT2G18790	AT2G18800	Xyloglucan endotransglucosylase/hydrolase 21
JCVI_17316:294	A1	1299493	2.913	53.0%	-	R (38)	G (3)	AT4G35830	AT4G35783	Rotindiflora-like 6 (DEVIL 17)
JCVI_17316:524	C1	1793417	2.913	53.0%	-	C (38)	Y (3)	AT4G35830	AT4G35783	Rotindiflora-like 6 (DEVIL 17)
JCVI_42176:261	C6	45986992	2.887	80.1%	-	S (38)	G (1)	AT1G78300	AT1G78270	UDP-glucosyl transferase 85A4
JCVI_14040:1428	C3	39402542	2.757	42.9%	-	C (32)	Y (6)	AT5G63910	AT5G63840	Radial swelling 3
JCVI_28444:452	C3	24863641	2.734	35.4%	-	W (32)	T (9)	AT3G09200	AT3G09230	MYB 1
JCVI_20507:170	C3	24930758	2.734	35.4%	-	G (32)	R (9)	AT3G08930	AT3G08900	Reversibly glycosylated polypeptide 3
JCVI_23924:508	A1	11013742	2.672	31.3%	-	A (24)	R (12)	AT4G15780	AT4G15750	Plant invertase/pectin methylesterase inhibitor superfamily protein
JCVI_24582:287	C6	36983160	2.503	33.4%	-	G (33)	R (6)	AT1G66150	Multiple	O-Glycosyl hydrolases family 17 (AT1G66250), β-GLU 21 & 22 (AT1G66270-80)
JCVI_1123:1343	A10	1552565	2.486	39.5%	-	T (30)	Y (6)	AT1G04680	AT1G04680	Pectin lyase-like superfamily protein
JCVI_27338:871	A7	25499900	2.467	53.2%	-	Y (37)	C (2)	AT1G79590	AT1G79680	Walll associated kinase (WAK)-LIKE 10
JCVI_11865:629	A5	23731750	2.446	49.6%	-	T (39)	W (2)	AT3G02520	-	-
JCVI_30131:697	C2	18061274	2.446	52.7%	-	Y (39)	T (2)	AT1G66330	Multiple	O-Glycosyl hydrolases family 17 (AT1G66250), β-Glu 21 & 22 (AT1G66270-80)
JCVI_23161:583	A7	20292260	2.433	32.5%	-	C (35)	Y (6)	AT1G66150	Multiple	O-Glycosyl hydrolases family 17 (AT1G66250), β-Glu 21 & 22 (AT1G66270-80)
JCVI_33789:898	C5	933779	2.396	52.7%	-	S (35)	G (2)	AT1G03350	AT1G03310	Isoamylase 2
JCVI_6272:785	A7	6255804	2.366	18.1%	-	G (35)	R (6)	AT3G24160	Multiple	Pectate lyase (AT3G24130, AT3G24230), L-Gln D-Fru-6-P amidotransferase (AT3G24090)
JCVI_19884:388	A4	10777277	2.338	48.9%	-	K (39)	T (2)	AT4G39960	AT4G40000	SAM-dependent methyltransferases superfamily protein
JCVI_430:780	C8	16519130	2.334	23.9%	-	T (25)	Y (15)	AT1G09070	AT1G08990	GlcA substitution of xylan 5
JCVI_17749:393	C4	1042253	2.282	47.1%	-	S (31)	C (8)	AT2G46930	AT2G46930	Pectinacetylesterase family protein
JCVI_16535:211	A6	21654008	2.222	16.2%	-	Y (22)	C (14)	AT3G28730	-	-
JCVI_24861:784	C3	3767227	2.210	17.2%	-	R (26)	G (12)	AT5G16070	Multiple	Carboxyesterase 17 (AT5G16080) Trichrome birefringence-like 19, 21 (AT5G15890-900)
JCVI_4321:169	A6	17954688	2.115	28.2%	-	M (29)	A (7)	AT2G17980	AT2G17950	Wuschel 1
JCVI_12463:129	A6	4135169	2.113	24.8%	T (9)	Y (32)	-	AT1G10580	Mulitple	UDP-Glycosyltransferase (AT1G10400) OR XG:Xyloglucosyltransferase 33 (AT1G10550)
JCVI_2668:388	A4	14942908	2.097	12.9%	-	G (27)	R (13)	AT2G30970	AT2G30933	Carbohydrate-binding X8 domain
JCVI_8086:219	C4	12347366	2.081	42.5%	-	Y (38)	T (3)	AT2G31610	Mulitple	Plant invertase/pectin methylesterase inhibitors (AT2G31425, 30, 32)
JCVI_22098:200	C1	4152951	2.053	48.9%	-	R (35)	G (2)	AT1G01140	AT1G01220	L-fucokinase/GDP-L-Fucose pyrophosphorylase
JCVI_11084:856	C8	30402343	2.045	14.4%	T (14)	Y (24)	-	AT3G62010	Mulitple	Putative GH(AT3G62050),Pectinacetylesterase(AT3G62060),SAM-transfer(AT3G62000)

*Effect size is the quantitative difference between means for each marker, expressed as % of the mean

Appendix Table 21 – 185 HMF – SNP markers found in regions of the *B. napus* transcriptome associated with the variation in HMF released from various cultivars ($p < 0.01$) after liquid hot water pretreatment at 185 °C for 10 min.

HMF 210

Brassica napus							*Arabidopsis thaliana*		
Marker	**Position**		**-log10P**	**Effect size***	**SNP Alleles**		**Ortholog AGI**	**Candidate AGI**	**Candidate name**
	Chr	**Pseudo**			**-**	**Modal**	**+**			
JCVI_16290:1079	A9	28533654	2.999	22.0%	-	R (29)	A (10)	AT1G22070	AT1G22370	UDP-Glucosyl transferase 85A5
JCVI_9783:671	C9	62422544	2.920	25.9%	-	R (28)	G (7)	AT5G04360	AT5G04360	Pullunase 1 (Starch mobilisation)
JCVI_3521:893	C5	40665410	2.824	34.5%	-	C (38)	M (3)	AT3G11170	AT3G11030	Trichome birefringence-like 32
JCVI_2158:1033	C1	8332264	2.780	21.9%	-	G (32)	R (8)	AT4G37550	Mulitple	Peroxidase (AT4G37520) OR Acyl-CoA N-acyltransferases (AT4G37580)
JCVI_2037:155	C4	17033148	2.744	18.5%	-	T (27)	Y (11)	AT2G28470	AT2G28470	β-Galactosidase 8
JCVI_16847:161	C7	33723111	2.668	28.2%	-	K (37)	T (4)	AT3G25690	Multiple	Inflorescence Deficient in Abscission (AT3G25655) / Devil 6 (AT3G25717)
JCVI_9831:1069	A1	736998	2.650	23.3%	-	W (34)	A (7)	AT4G37270	AT4G37260	MYB 73
JCVI_3459:461	C6	9241042	2.643	18.1%	-	T (30)	Y (11)	ATIG53240	ATIG53210	Na^+/Ca^{2+} Exchanger
JCVI_40373:376	C1	11850002	2.588	26.3%	-	W (35)	T (4)	AT4G23940	AT4G23990	Cellulose Synthase-like G1, 2 and 3
JCVI_28986:175	A9	1262107	2.587	28.3%	-	T (35)	Y (4)	AT4G03400	AT3G07160	Glucan synthase-like 10
JCVI_8934:1229	A8	989317	2.576	16.7%	Y (19)	T (22)	-	ATIG53310	ATIG53210	Na^+/Ca^{2+} Exchanger
JCVI_16762:362	A1	25789453	2.498	17.2%	-	Y (20)	C (16)	AT3G13080	AT3G10740	α-L-Arabionfuranosidase 1
JCVI_1473:242	A2	2865140	2.470	23.0%	-	C (34)	S (7)	AT5G16130	AT5G16170	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein
JCVI_2675:421	C3	3169994	2.437	17.4%	-	Y (27)	C (13)	AT5G14030	AT5G13980	Glycosyl hydrolase family 38 protein
JCVI_36398:1059	A1	22755099	2.436	26.9%	-	T (34)	Y (6)	AT3G18490	AT3G18660	GlcA substitution of xylan 1
JCVI_28444:452	C3	24863641	2.409	20.7%	-	W (32)	T (9)	AT3G09200	AT3G09230	MYB 1
JCVI_20507:170	C3	24930758	2.409	20.7%	-	G (32)	R (9)	AT3G08930	AT3G08900	Reversibly glycosylated polypeptide 3
JCVI_10958:827	C4	7834190	2.395	16.2%	-	R (22)	A (15)	AT2G37460	AT2G37585	GlcA Transferase 14C
JCVI_99:284	C8	20843738	2.379	15.6%	G (16)	R (25)	-	ATIG19150	ATIG19170	Pectin lyase-like superfamily protein
JCVI_25757:176	C5	1585200	2.378	16.0%	-	C (24)	Y (15)	ATIG05010	-	-
JCVI_5470:366	C3	6518442	2.334	15.6%	-	K (26)	T (15)	AT5G57860	-	-
JCVI_8520:313	C3	55986823	2.333	22.4%	-	Y (36)	T (5)	ATIG56070	ATIG56010	NAC Domain containing protein 21
JCVI_8228:923	C4	33292035	2.300	26.9%	-	R (36)	A (5)	AT4G13250	AT4G13210	Pectin lyase-like superfamily protein
JCVI_1522:384	C4	33302316	2.300	26.9%	-	R (36)	A (5)	AT4G13200	AT4G13210	Pectin lyase-like superfamily protein
JCVI_1123:1343	A10	1552565	2.299	24.0%	-	T (30)	Y (6)	ATIG04680	ATIG04680	Pectin lyase-like superfamily protein
JCVI_14287:443	C7	6504085	2.266	30.9%	-	A (35)	M (3)	AT2G17970	-	-
JCVI_11084:856	C8	30402343	2.262	15.6%	T (14)	Y (24)	-	AT3G62010	Mulitple	Pectinacetylesterase, Endonuclease/GH or SAM-transferase (AT3G62060, 50 or 00)
JCVI_41844:543	C4	46009220	2.224	19.1%	T (7)	Y (30)	-	AT2G28840	AT2G28760	UDP-XYL synthase 6
JCVI_31706:429	C7	39438828	2.195	14.1%	A (19)	W (21)	-	AT5G24650	-	-
JCVI_34007:532	C3	61576274	2.168	15.2%	-	A (25)	W (14)	AT5G38830	Multiple	Glycine rich protein (AT4G21620), Tubulin 9 (AT4G20890) or UGE in root (AT4G21620)
JCVI_17904:1676	A6	2641565	2.160	14.8%	T (14)	Y (25)	-	ATIG48650	-	-
JCVI_14576:421	C9	60576792	2.156	32.8%	-	G (39)	R (2)	AT5G08280	-	-
JCVI_20139:902	C9	60580461	2.156	32.8%	-	G (39)	R (2)	AT5G08260	-	-
JCVI_23603:1284	A6	14887823	2.152	18.3%	G (7)	R (31)	-	AT3G49530	-	-
JCVI_2217:1072	C9	5323828	2.147	23.4%		Y (35)	T (4)	AT5G63050	-	-
JCVI_23501:449	A8	7108141	2.132	16.3%	R (19)	G (21)	-	AT4G17330	AT4G17230	Scarecrow-like 13
JCVI_1200:441	C9	56825835	2.128	47.0%	-	R (38)	G (1)	AT5G16030	AT5G15900	Trichome birefringence-like 19
JCVI_8521:1068	A2	11412303	2.112	15.4%	-	W (21)	T (17)	ATIG72010	ATIG72110	O-acyltransferase (WSD1-like) family protein
JCVI_3716:282	A7	12305247	2.090	46.4%	-	C (38)	Y (1)	AT2G27100	AT2G27100	Serrate
JCVI_2668:388	A4	14942908	2.079	15.7%	-	G (27)	R (13)	AT2G30970	AT2G30933	Carbohydrate-binding X8 domain

Appendix Table 22 – 210 HMF – SNP markers found in regions of the *B. napus* transcriptome associated with the variation in HMF released from various cultivars ($p < 0.01$) after liquid hot water pretreatment at 210 °C for 10 min.

HMF 210 Continued...											
JCVI_33922:193	A8	11510806	2.077	46.3%	-	Y (38)	C (1)	AT4G25640	AT4G25620	Hydroxyproline-rich glycoprotein family protein
JCVI_308:263	C8	25089768	2.060	45.8%	-	M (40)	C (1)	AT3G52800	AT3G52840	β-galactosidase 2
JCVI_25085:69	C2	1218598	2.054	45.8%	-	K (39)	G (1)	AT5G05170	AT5G05170	Cellulose synthase 3
JCVI_12969:772	A10	10101684	2.047	21.4%	-	G (36)	S (5)	AT5G17230	Multiple	Nucleotide-sugar transporter family protein / glycine-proline-rich protein
JCVI_13677:477	A7	24556175	2.047	25.2%	-	R (36)	G (3)	AT1G77180	AT1G77210	Sugar transport protein 14
JCVI_6656:281	C1	36465929	2.047	18.0%	T (7)	Y (33)	-	AT3G23600	AT3G24180	Beta-glucosidase, GBA2 type family protein;
JCVI_9752:322	A6	23830131	2.043	17.1%	T (18)	K (20)	-	AT1G03370	AT5G47780	Galacturonosyltransferase 4
JCVI_29426:694	C5	36033682	2.037	46.1%	-	R (33)	G (1)	AT2G18790	AT2G18800	Xyloglucan endotransglucosylase/hydrolase 21
JCVI_41807:254	A8	17341483	2.033	45.4%	-	Y (39)	T (1)	AT1G10470	AT1G10400	UDP-Glycosyltransferase superfamily protein
JCVI_23924:508	A1	11013742	2.031	17.6%	-	A (24)	R (12)	AT4G15780	AT4G15750	Plant invertase/pectin methylesterase inhibitor superfamily protein
JCVI_13297:98	A5	16814795	2.031	22.3%	-	Y (37)	T (4)	AT3G19760	-	-

*Effect size is the quantitative difference between means for each marker, expressed as % of the mean

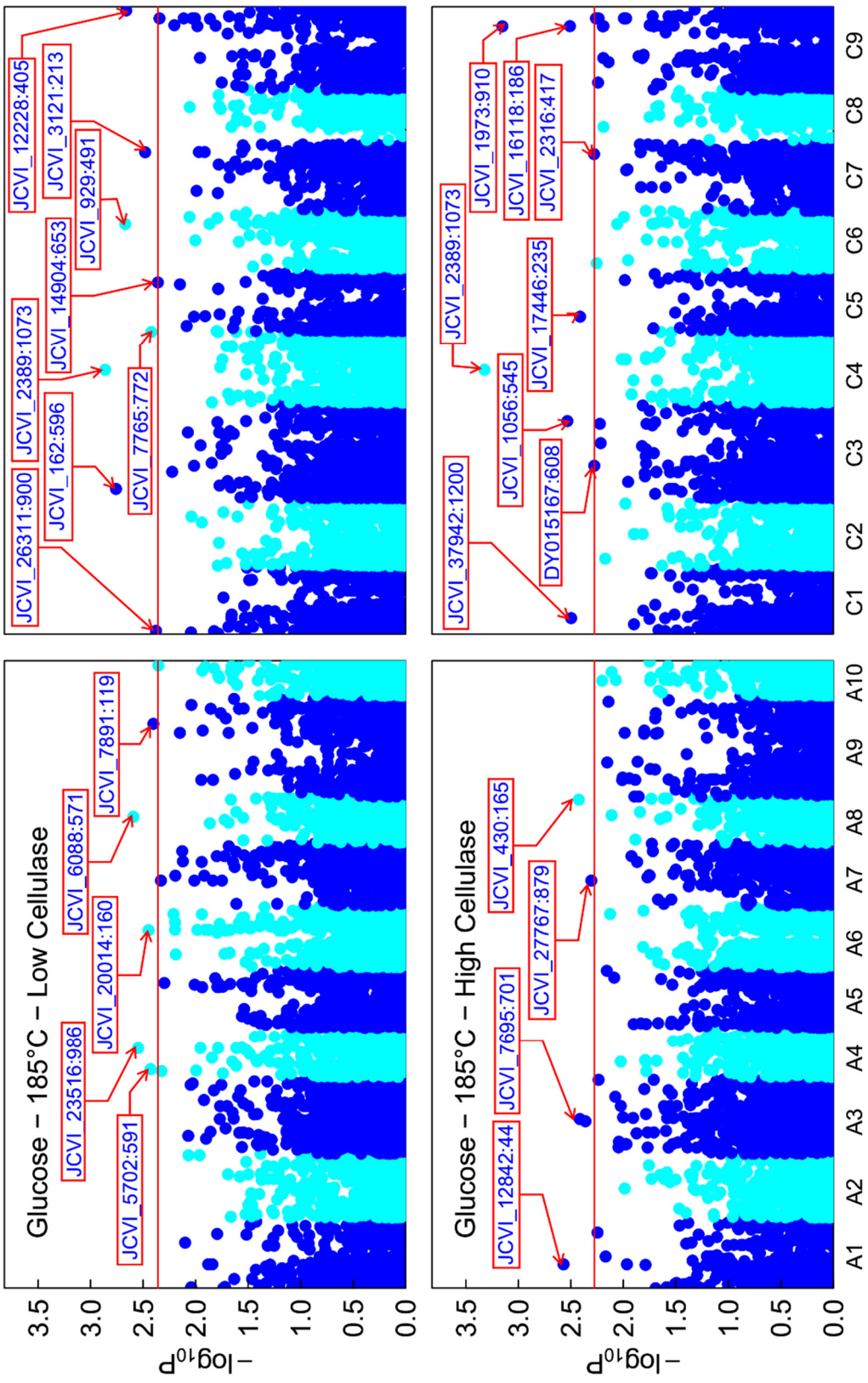
Appendix Table 22 continued – 210 HMF – SNP markers found in regions of the *B. napus* transcriptome associated with the variation in HMF released from various cultivars ($p < 0.01$) after liquid hot water pretreatment at 210 °C for 10 min.

Process Parameter		*Brassica napus*			*Arabidopsis thaliana*		
Product	Process[a]	HemiSNP ID	-log10(P)	Effect size[b]	Analagous AGI	Candidate AGI	Potential candidate name
Glucose	**185L**	**JCVI_29436:531**	**3.22**	**13.67%**	**AT1G29880**	**AT1G29890**	**Reduced wall acetylation 4**
Glucose	185L	JCVI_37516:445	3.03	11.81%	AT5G55400	AT5G54690	Irregular xylem 8
Glucose	185L	JCVI_1757:110	3.00	12.58%	AT4G30600	AT4G31590	Cellulose-synthase-like C5
Glucose	**185L**	**JCVI_2389:1073**	**2.99**	**10.19%**	**AT3G63000**	**AT3G62830**	**UDP-glucuronic acid decarboxylase 2**
Glucose	185L	JCVI_10517:832	2.83	13.82%	AT3G17410	AT3G17365	S-adenosyl-L-methionine-dependent methyltransferase
Glucose	185L	JCVI_5470:94	2.76	11.12%	AT5G57860	-	-
Glucose	185L	JCVI_162:335	2.75	11.34%	AT5G56760	AT5G56590	O-Glycosyl hydrolases Family 17
Glucose	185L	JCVI_30751:564	2.70	13.83%	AT2G23760	AT2G23900	Pectin lyase-like superfamily protein
Glucose	185L	JCVI_2057:431	2.72	12.82%	AT5G47050	AT5G47000 - 46940	Peroxidase superfamily protein with xylosidase activity - series of pectin methylesterase inhibitor proteins
Glucose	**185H**	**JCVI_2480:239**	**3.56**	**22.69%**	**AT3G11510**	**AT3G11570**	**Trichrome birefringence-like 8**
Glucose	**185H**	**JCVI_2389:1068**	**3.38**	**22.71%**	**AT3G63000**	**AT3G62830**	**UDP-glucuronic acid decarboxylase 2**
Glucose	185H	JCVI_12842:44	3.15	26.07%	AT5G57815	AT5G58090	O-Glycosyl hydrolases family 17
Glucose	185H	JCVI_29979:139	3.13	21.17%	AT5G39740	AT5G39700	MYB domain protein 89
Glucose	**185H**	**JCVI_1973:910**	**3.13**	**21.11%**	**AT5G57655**	**AT5G57655**	**Xylose isomerase family protein**
Glucose	185H	JCVI_22675:405	3.00	21.66%	AT1G36320	AT1G36310	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
Glucose	185H	JCVI_31108:249	2.97	20.83%	AT5G13710	AT5G13870	Xyloglucan endotransglucosylase/hydrolase 5
Glucose	185H	JCVI_8081:313	2.99	23.34%	AT5G18380	AT5G18410	PIROG I 121
Glucose	185H	JCVI_24888:728	2.87	22.39%	AT5G47010	Multiple	Low-level beta-amylase 1 (AT5G47010) or Peroxidase (AT5G47000)
Glucose	185H	JCVI_35535:58	2.98	25.41%	AT1G50740	-	-
Glucose	185H	JCVI_22240:62	2.94	18.82%	AT1G64520	AT1G64440	UDP-Glucose 4-Epimerase
Glucose	185H	JCVI_34849:107	2.90	25.41%	AT4G05070	AT4G05070	Wound-responsive family protein
Glucose	185H	JCVI_14190:251	2.84	23.41%	AT1G80380	AT1G80290	Glycosyltransferase family 64 protein
Glucose	185H	JCVI_1056:545	2.83	26.50%	AT4G36690	AT4G36890	Irregular Xylem 14
Glucose	185H	JCVI_3274:161	2.83	22.78%	AT5G55220	AT5G53370	Pectin methylesterase PCR fragment F
Glucose	**210L**	**JCVI_1973:910**	**3.73**	**16.45%**	**AT5G57655**	**AT5G57655**	**Xylose isomerase family protein**
Glucose	210L	JCVI_8724:1452	3.25	21.83%	AT4G35790	Multiple	Pectin lyase-like superfamily protein (AT4G35670) or F5H 1 (AT4G36220)
Glucose	**210L**	**JCVI_18300:537**	**3.23**	**15.31%**	**AT4G17880**	**AT4G17880**	**MYC4 - JAZ-interacting transcription factor**
Glucose	210L	JCVI_18290:149	3.20	14.05%	AT3G23980	AT3G23920	Beta-amylase 7
Glucose	210L	JCVI_15579:449	3.04	17.62%	AT1G29470	AT1G29050	Trichrome birefringence-like 38
Glucose	210L	JCVI_319:589	3.05	14.99%	AT2G05840	AT2G06850	Xyloglucan endotransglucosylase/hydrolase 4
Glucose	210L	JCVI_38468:358	3.05	22.17%	AT5G47110	Unknown	Unknown
Glucose	210L	JCVI_36642:297	2.88	16.61%	AT1G11755	AT1G11730 - 20	Galactosyltransferase family protein OR Starch Synthase 3
Glucose	210L	JCVI_36064:733	2.83	17.37%	AT2G17040	AT2G17040	NAC domain containing protein 36
Glucose	**210H**	**JCVI_2389:1073**	**4.07**	**8.53%**	**AT3G63000**	**AT3G62830**	**UDP-glucuronic acid decarboxylase 2**
Glucose	**210H**	**JCVI_430:165**	**3.86**	**7.87%**	**AT1G09070**	**AT1G08990**	**Glucuronic acid substitution of xylan 5**
Glucose	210H	JCVI_9578:534	3.34	8.53%	AT5G22360	Unknown	Unknown
Glucose	210H	JCVI_31108:249	3.30	7.18%	AT5G13710	AT5G13780	Acyl-CoA N-acyltransferases
Glucose	**210H**	**JCVI_16118:186**	**3.17**	**7.19%**	**AT5G57710**	**AT5G57655**	**Xylose isomerase family protein**
Glucose	210H	JCVI_25265:495	3.06	7.84%	AT3G27690	Unknown	Unknown
Glucose	210H	JCVI_26505:1749	3.06	8.37%	AT5G48385	AT5G48375	Beta glucosidase 39
Glucose	210H	JCVI_17664:893	3.04	10.30%	AT2G01320	AT2G01320	ABC-2 type transporter family protein
Glucose	210H	JCVI_7060:245	3.04	6.89%	AT1G70600	AT1G70630	Reduced arabinose yariv 1
Glucose	210H	JCVI_1445:1032	3.00	8.15%	AT3G20910	AT3G20865	Arabinogalactan protein 40
Glucose	210H	JCVI_14508:616	2.93	6.90%	AT3G54360	Multiple	Trichrome Birefringence-like 36 or Glycoside hydrolase family 2 protein
Glucose	210H	JCVI_21183:247	2.87	8.37%	AT3G11130	AT3G11030- 40	Trichrome birefringence-like 32 or Endo-β-N-acetyglucosaminidase 85B

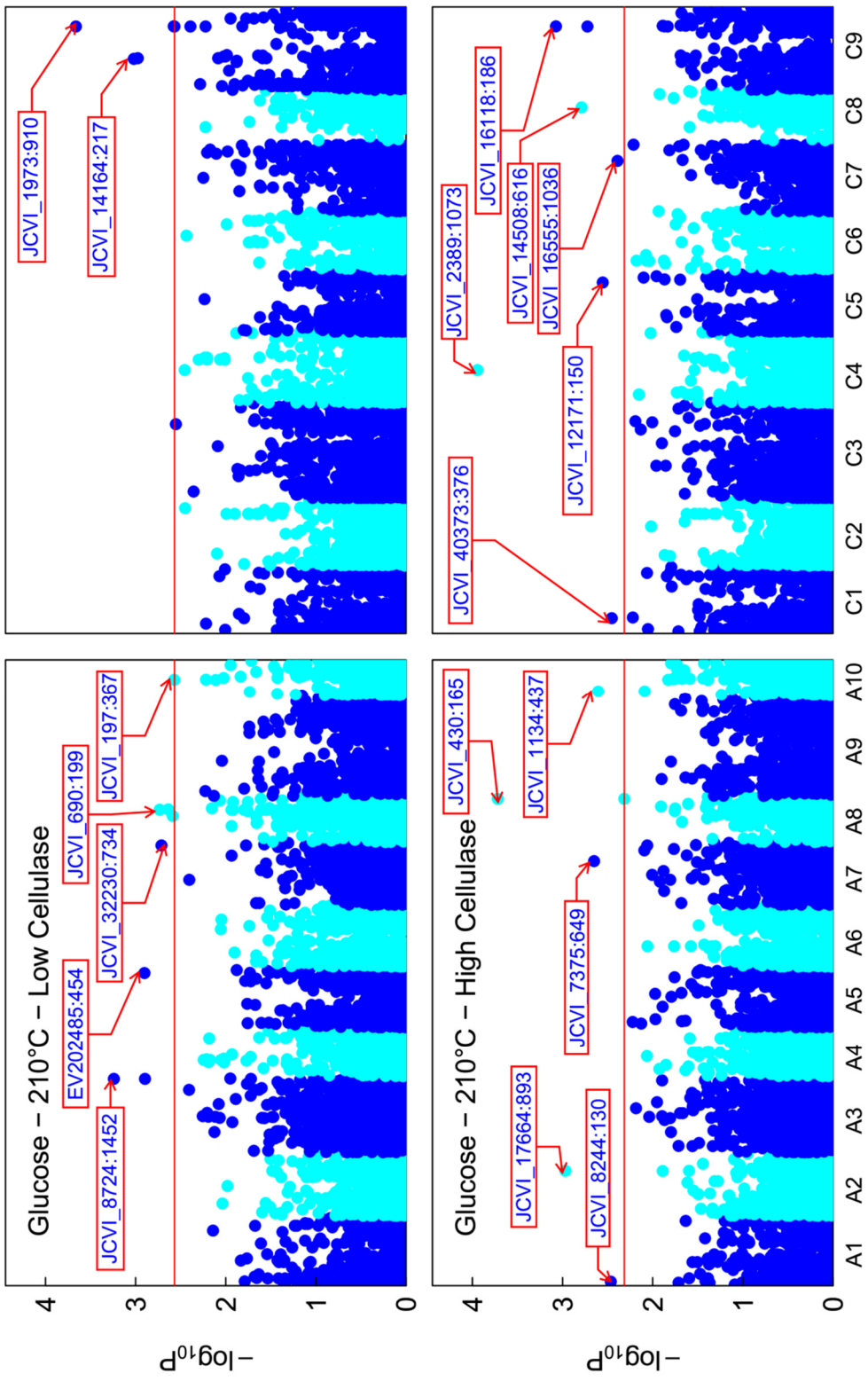
[a] Pretreatment conditions either 185 °C or 210 °C, 10 min (185 or 210) followed by hydrolysis at a high (H) or low (L) cellulase dose.

[b] Effect size is the predicted proportion of the trait range attributed to the marker. These values may be cumulative for markers associated with genes that work in concert.

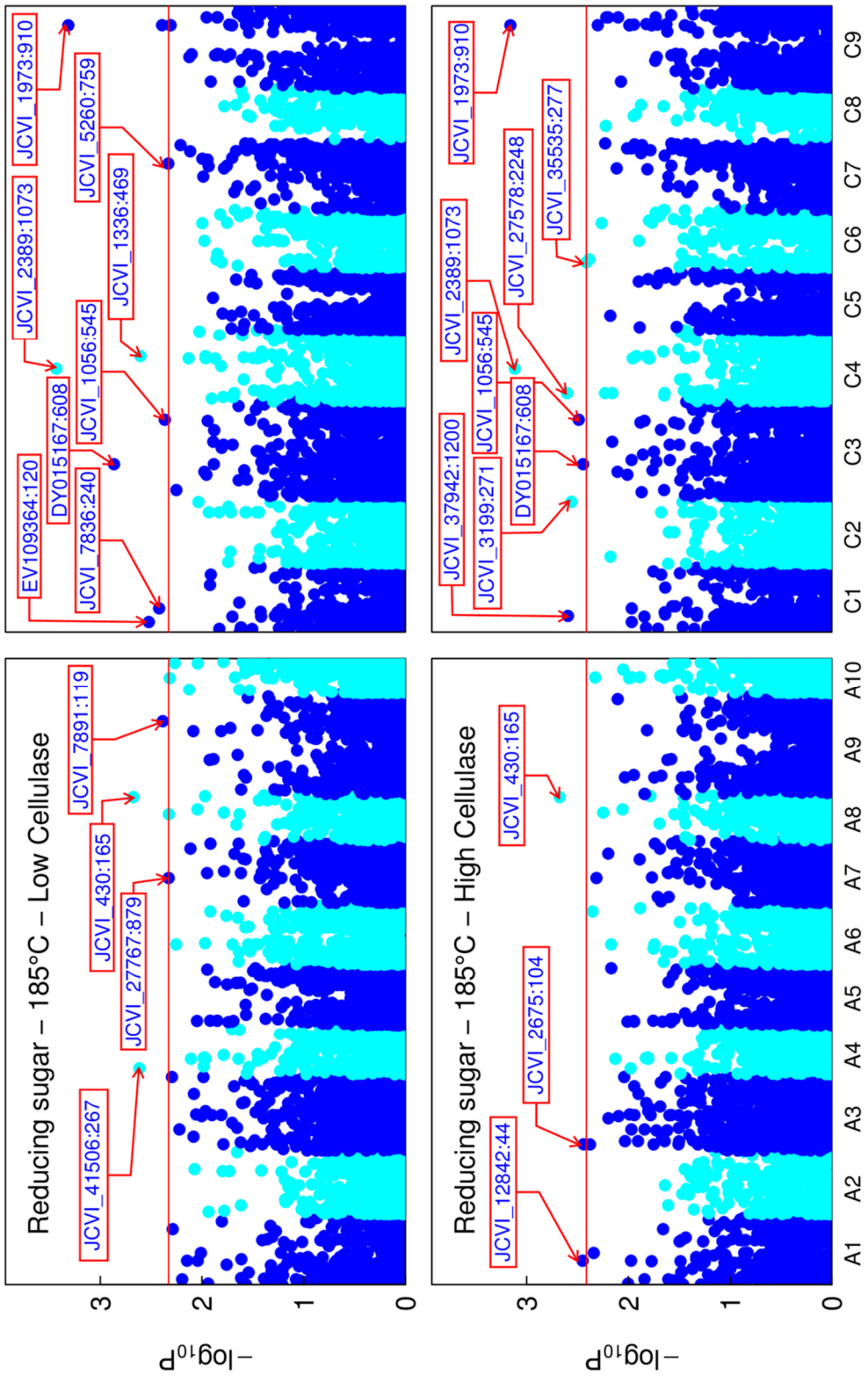
Appendix Table 23 – Candidates initially selected based on non-genome assigned SNP markers – Top 0.1% of markers.



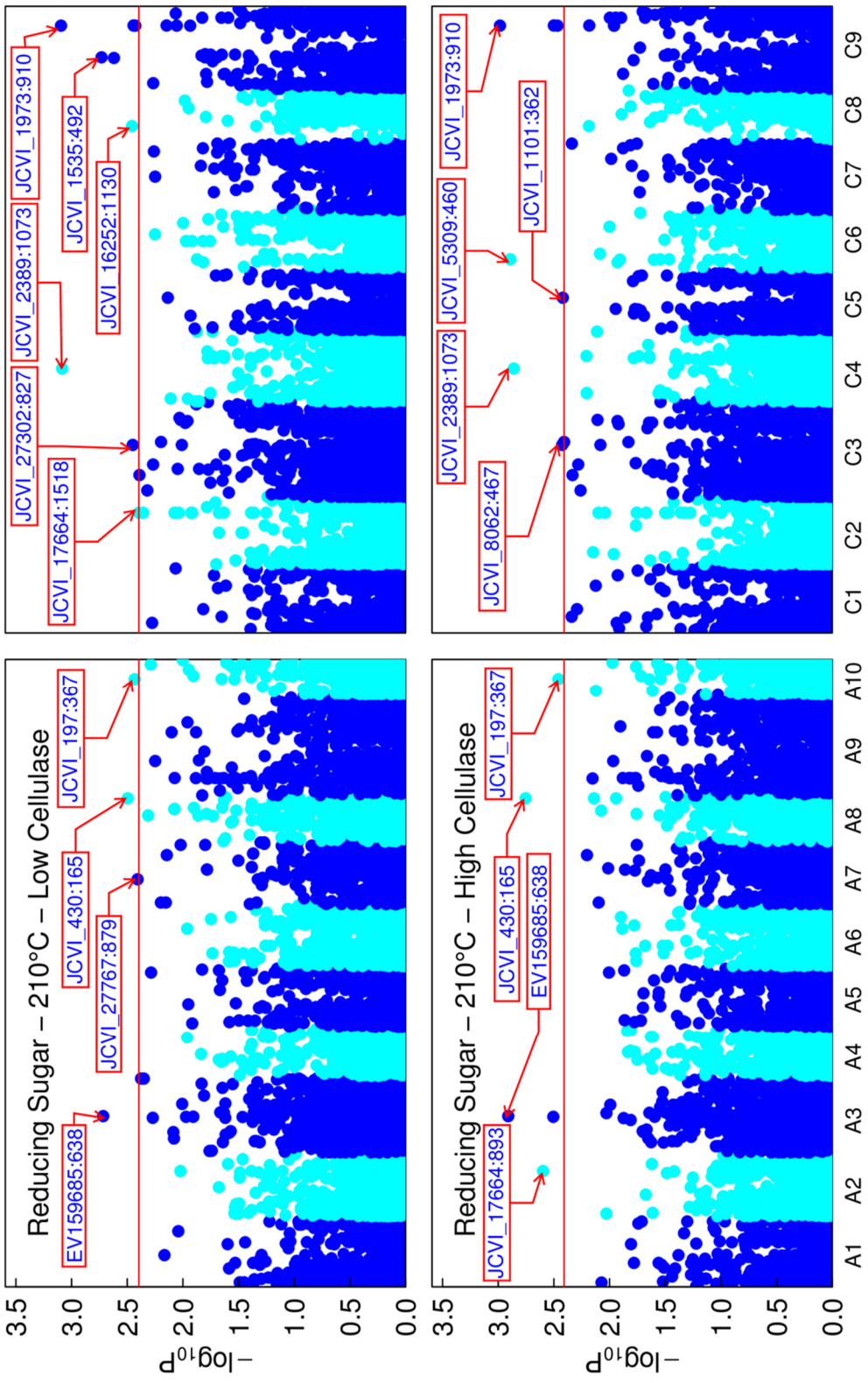
Appendix Figure 1 – Annotated Manhattan plots showing SNP distributions associated with Glucose yields after pretreatment at 185°C and hydrolysis at two cellulase doses, labelling the top 0.1%.



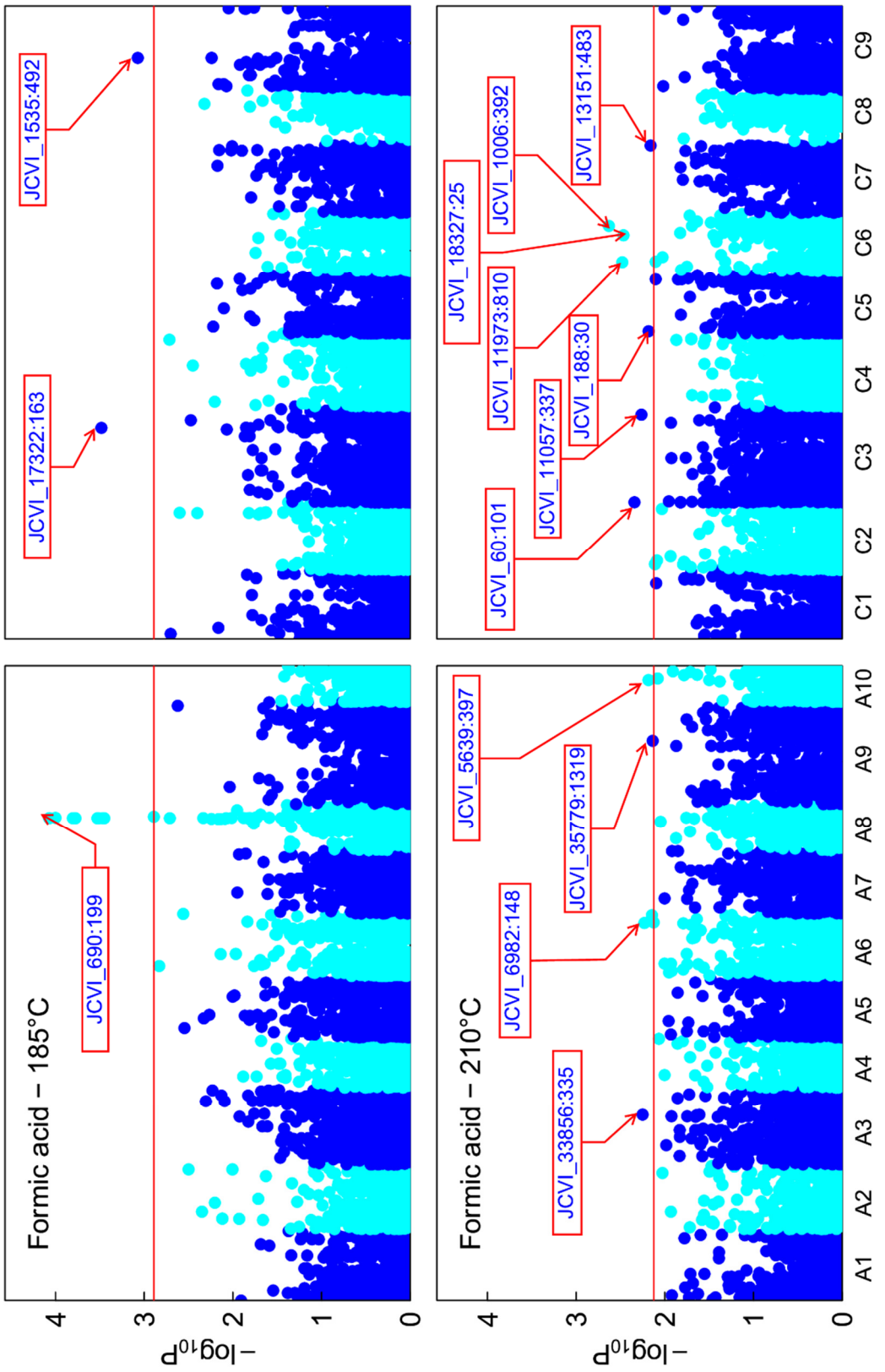
Appendix Figure 2 – Annotated Manhattan plots showing SNP distributions associated with - Glucose yields after pretreatment at 210°C and hydrolysis at two cellulase doses, labelling the top 0.1%.



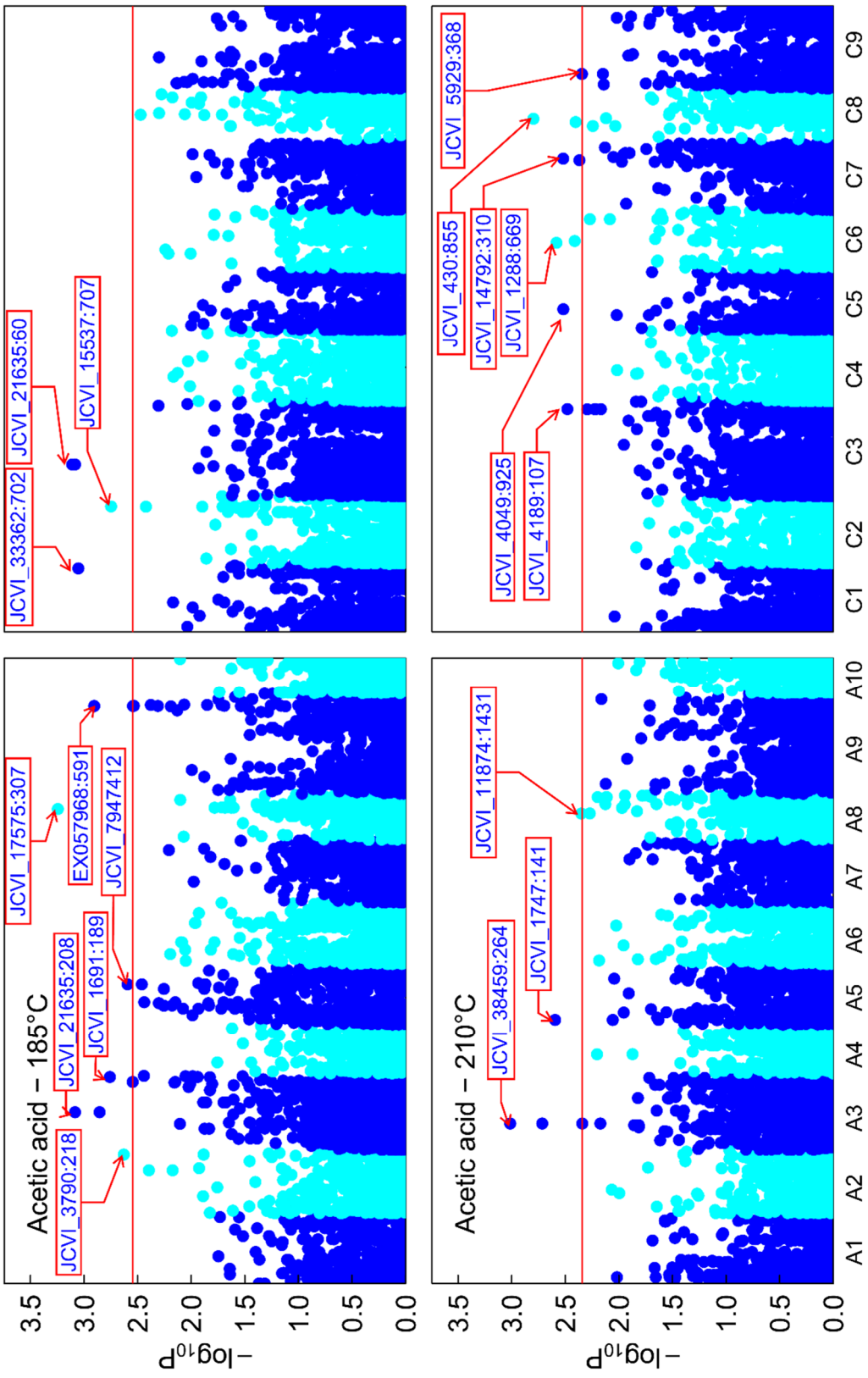
Appendix Figure 3 – Annotated Manhattan plots showing SNP distributions associated with Reducing sugar yields after pretreatment at 185°C and hydrolysis at two cellulase doses, labelling the top 0.1%.



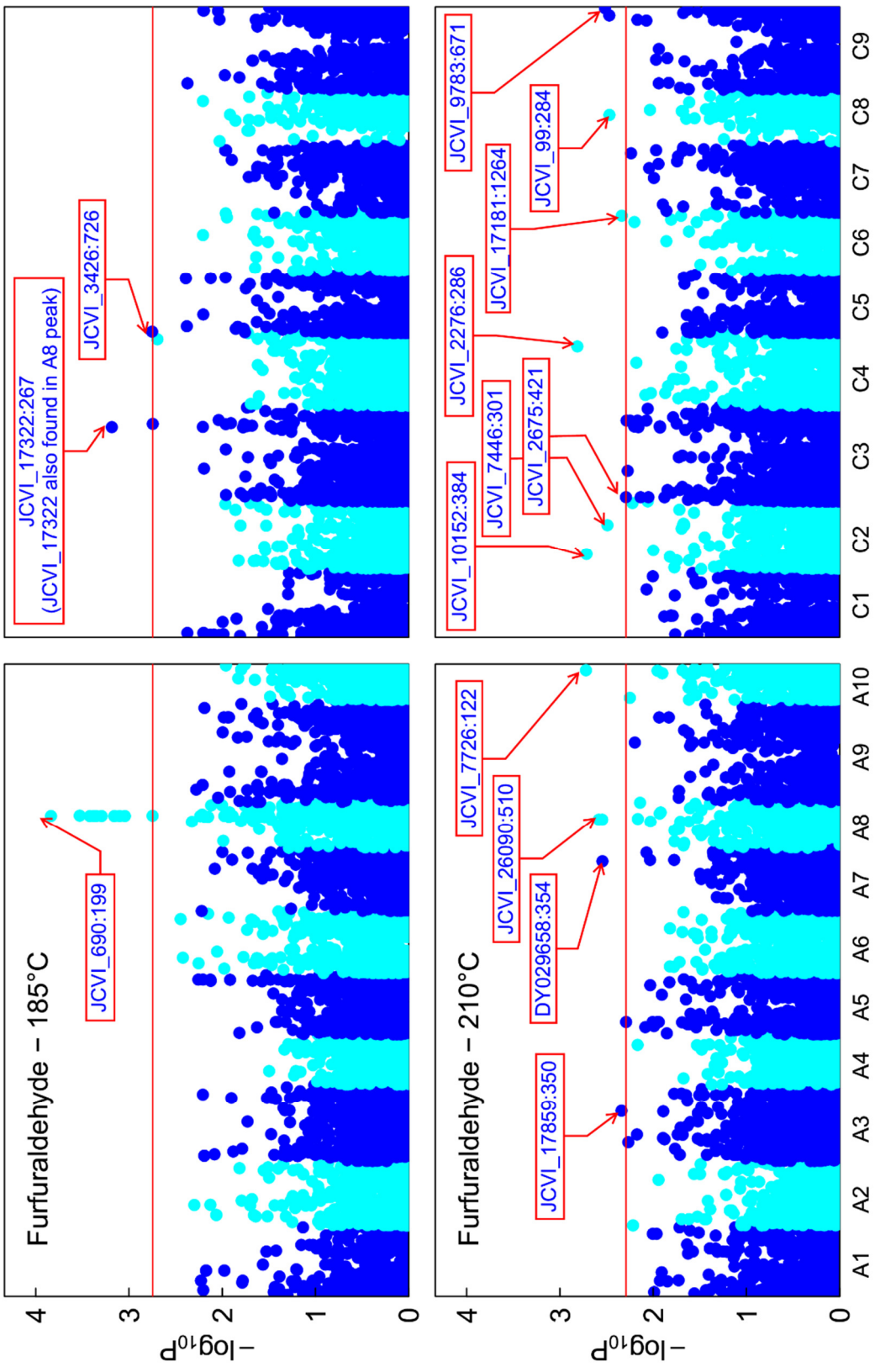
Appendix Figure 4 – Annotated Manhattan plots showing SNP distributions associated with – Reducing sugar yields after pretreatment at 210°C and hydrolysis at two cellulase doses, labelling the top 0.1%.



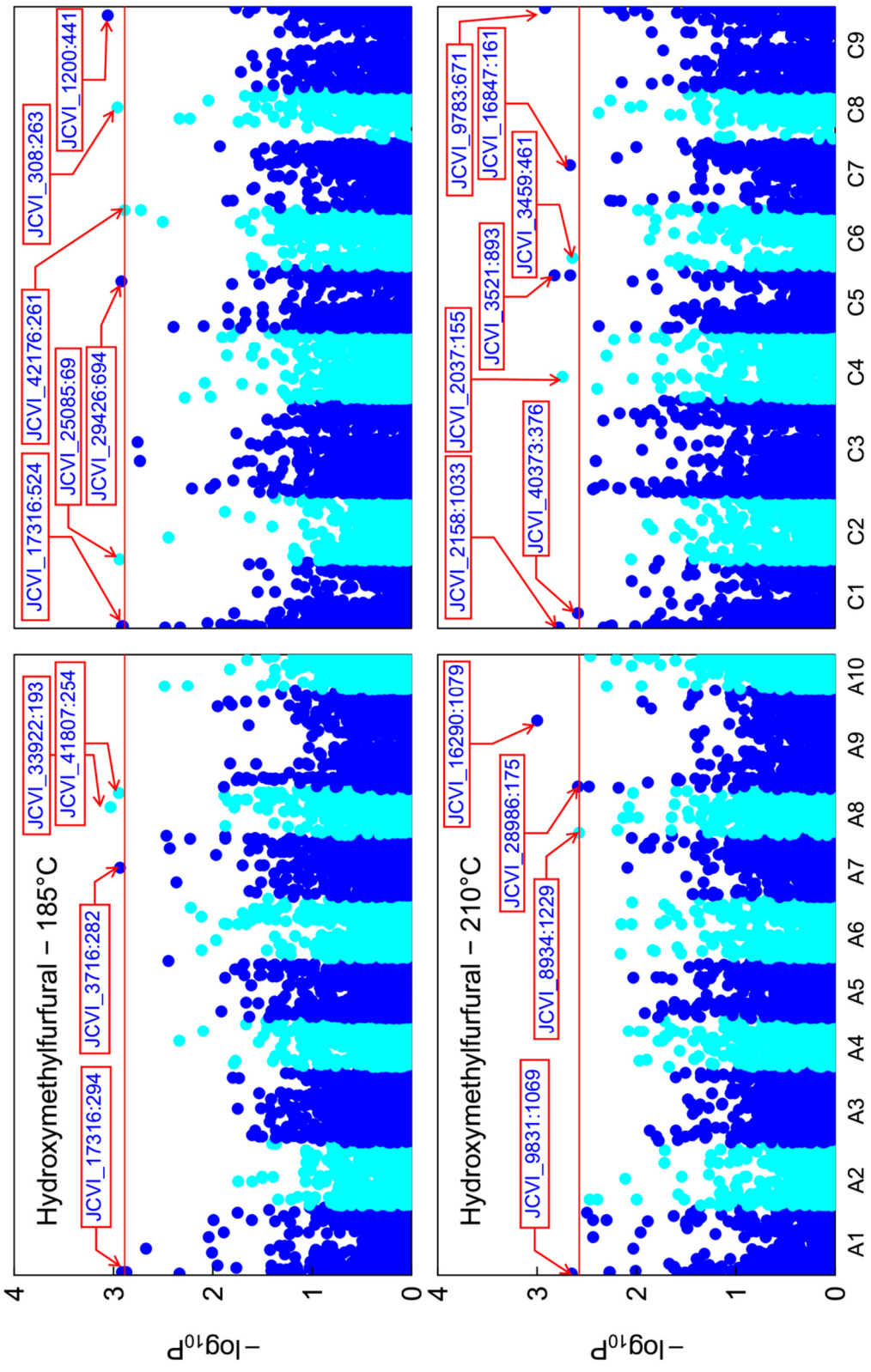
Appendix Figure 5 – Annotated Manhattan plots showing SNP distributions associated with - Formic acid yields after pretreatment at 185°C and 210°C, labelling the top 0.1%.



Appendix Figure 6 – Annotated Manhattan plots showing SNP distributions associated with - Acetic acid yields after pretreatment at 185°C and 210°C, labelling the top 0.1%.



Appendix Figure 7 – Annotated Manhattan plots showing SNP distributions associated with – 2FA yields after pretreatment at 185°C and 210°C, labelling the top 0.1%.



Appendix Figure 8 – Annotated Manhattan plots showing SNP distributions associated with HMF yields after pretreatment at 185°C and 210°C, labelling the top 0.1%.

11 References

Abramson, M., Shoseyov, O., Shani, Z. 2010. Plant cell wall reconstruction toward improved lignocellulosic production and processability. Plant Science, 178, 61-72.

Adani, F., Papa, G., Schievano, A., Cardinale, G., D'Imporzano, G., Tambone, F. 2011. Nanoscale Structure of the Cell Wall Protecting Cellulose from Enzyme Attack. Environmental Science & Technology, 45, 1107-1113.

Advanced Ethanol Council. 2012. Cellulosic biofuels industry progress report 2012-2013.

Akin, D.E. 2008. Plant cell wall aromatics: influence on degradation of biomass. Biofuels, Bioproducts & Biorefining, 2, 288-303.

Albenne, C., Canut, H., Hoffmann, L., Jamet, E. 2014. Plant Cell Wall Proteins: A Large Body of Data, but What about Runaways? Proteomes, 2, 224-242.

Alejandro, S., Lee, Y., Tohge, T., Sudre, D., Osorio, S., Park, J., Bovet, L., Lee, Y., Geldner, N., Fernie, A.R., Martinoia, E. 2012. AtABCG29 Is a Monolignol Transporter Involved in Lignin Biosynthesis. Current Biology, 22, 1207-1212.

Aleklett, K., Höök, M., Jakobsson, K., Lardelli, M., Snowden, S., Söderbergh, B. 2010. The Peak of the Oil Age – Analyzing the world oil production Reference Scenario in World Energy Outlook 2008. Energy Policy, 38, 1398-1414.

Alexander, B.W., Alex H. Gordon, James A. Lomax andrew Chesson. 1987. Composition and rumen degradability of straw from three varieties of oilseed rape before and after alkali, hydrothermal and oxidative treatment. Journal of the Science of Food & Agriculture, 41, 1-15.

Arantes, V., Saddler, J.N. 2010. Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. Biotechnology for Biofuels, 3.

Arantes, V., Saddler, J.N. 2011. Cellulose accessibility limits the effectiveness of minimum cellulase loading on the efficient hydrolysis of pretreated lignocellulosic substrates. Biotechnology for Biofuels, 4.

Arias, T., Beilstein, M.A., Tang, M., McKain, M.R., Pires, J.C. 2014. Diversification times among Brassica (Brassicaceae) crops suggest hybrid formation after 20 million years of divergence. American Journal of Botany, 101, 86-91.

Arvaniti, E., Bjerre, A.B., Schmidt, J.E. 2012. Wet oxidation pretreatment of rape straw for ethanol production. Biomass & Bioenergy, 39, 94-105.

Atmodjo, M.A., Hao, Z., Mohnen, D. 2013. Evolving views of pectin biosynthesis. Annu Rev Plant Biol, 64, 747-79.

Bakker, H., Routier, F., Oelmann, S., Jordi, W., Lommen, A., Gerardy-Schahn, R. and Bosch, D. 2005. Molecular cloning of two Arabidopsis UDP-galactose transporters by complementation of a deficient Chinese hamster ovary cell line. Glycobiology, 15, 193–201.

Bailey, M.J. 1988. A note on the use of dinitrosalicylic acid for determining the products of enzymatic reactions. Applied Microbiology and Biotechnology, 29, 494-496.

Balan, V., Chiaramonti, D., Kumar, S. 2013. Review of US and EU initiatives toward development, demonstration and commercialization of lignocellulosic biofuels. Biofuels Bioproducts & Biorefining-Biofpr, 7, 732-759

Bancroft, I., Morgan, C., Fraser, F., Higgins, J., Wells, R., Clissold, L., Baker, D., Long, Y., Meng, J.L., Wang, X.W., Liu, S.Y., Trick, M. 2011. Dissecting the genome of the polyploid crop oilseed rape by transcriptome sequencing. Nature Biotechnology, 29, 762-128.

Banoub, J. H., Delmas, M. 2003. Structural elucidation of the wheat straw lignin polymer by atmospheric pressure chemical ionization tandem mass spectrometry and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Journal of Mass Spectrometry, 38, 900-903.

Banerjee, G., Car, S., Scott-Craig, J.S., Hodge, D.B., Walton, J.D. 2011. Alkaline peroxide pretreatment of corn stover: effects of biomass, peroxide and enzyme loading and composition on yields of glucose and xylose. Biotechnology for Biofuels, 4.

Banerjee, G., Scott-Craig, J.S., Walton, J.D. 2010. Improving Enzymes for Biomass Conversion: A Basic Research Perspective. Bioenergy Research, 3, 82-92.

Bansal, P., Hall, M., Realff, M.J., Lee, J.H., Bommarius, A.S. 2009. Modeling cellulase kinetics on lignocellulosic substrates. Biotechnology Advances, 27, 833-848.

Bar-Peled, M., O'Neill, M.A. 2011. Plant Nucleotide Sugar Formation, Interconversion and Salvage by Sugar Recycling. Annual Review of Plant Biology, 62, 127-155.

Barber, C., Rösti, J., Rawat, A., Findlay, K., Roberts, K., & Seifert, G. J. 2006. Distinct properties of the five UDP-D-glucose/UDP-D-galactose 4-epimerase isoforms of Arabidopsis thaliana. Journal of Biological Chemistry, 281, 17276-17285.

Baroja-Fernández, E., Muñoz, F.J., Bahaji, A., Almagro, G., Pozueta-Romero, J. 2012a. Reply to Smith et al.: No evidence to challenge the current paradigm on starch and cellulose biosynthesis involving sucrose synthase activity. Proceedings of the National Academy of Sciences, 109, 777.

Baroja-Fernández, E., Muñoz, F.J., Li, J., Bahaji, A., Almagro, G., Montero, M., Etxeberria, E., Hidalgo, M., Sesma, M.T., Pozueta-Romero, J. 2012b. Sucrose synthase activity in the sus1/sus2/sus3/sus4 Arabidopsis mutant is sufficient to support normal cellulose and starch production. Proceedings of the National Academy of Sciences, 109, 321-326.

Barratt, D.H.P., Derbyshire, P., Findlay, K., Pike, M., Wellner, N., Lunn, J., Feil, R., Simpson, C., Maule, A.J., Smith, A.M. 2009. Normal growth of Arabidopsis requires cytosolic invertase but not sucrose synthase. Proceedings of the National Academy of Sciences of the United States of America, 106, 13124-13129.

Barros, S. 2013. Brazil Biofuels Annual Report 2013. Global Agricultural Information Network Report.

Basso, L.C., de Amorim, H.V., de Oliveira, A.J., Lopes, M.L. 2008. Yeast selection for fuel ethanol production in Brazil. Fems Yeast Research, 8, 1155-1163.

Berlin, A., Maximenko, V., Bura, R., Kang, K.-Y., Gilkes, N., Saddler, J. 2006. A rapid microassay to evaluate enzymatic hydrolysis of lignocellulosic substrates. Biotechnology and Bioengineering, 93, 880-886.

Berti, P., Levidow, L. 2014. Fuelling expectations: A policy-promise lock-in of UK biofuel policy. Energy Policy, 66, 135-143.

Bischoff, V., Nita, S., Neumetzler, L., Schindelasch, D., Urbain, A., Eshed, R., Persson, S., Delmer, D., Scheible, W.-R. 2010. TRICHOME BIREFRINGENCE and Its Homolog AT5G01360 Encode Plant-Specific DUF231 Proteins Required for Cellulose Biosynthesis in Arabidopsis. Plant Physiology, 153, 590-602.

Blumenkrantz, N., Asboe-Hansen, G. 1973. New method for quantitative determination of uronic acids. Analytical biochemistry, 54, 484-489.

Blakeney, A. B., Harris, P. J., Henry, R. J., & Stone, B. A. 1983. A simple and rapid preparation of alditol acetates for monosaccharide analysis. Carbohydrate Research, 113, 291-299.

Boerjan, W., Ralph, J., Baucher, M. 2003. Lignin biosynthesis. Annual Review of Plant Biology, 54, 519-546.

Bonawitz, N. D., Chapple, C. 2010. The genetics of lignin biosynthesis: connecting genotype to phenotype. Annual review of genetics, 44, 337-363.

Bonawitz, N.D., Kim, J.I., Tobimatsu, Y., Ciesielski, P.N. anderson, N.A., Ximenes, E., Maeda, J., Ralph, J., Donohoe, B.S., Ladisch, M., Chapple, C. 2014. Disruption of Mediator rescues the stunted growth of a lignin-deficient Arabidopsis mutant. Nature, advance online publication.

Bonawitz, N.D., Soltau, W.L., Blatchley, M.R., Powers, B.L., Hurlock, A.K., Seals, L.A., Weng, J.K., Stout, J., Chapple, C. 2012. REF4 and RFR1, Subunits of the Transcriptional Coregulatory Complex Mediator, Are Required for Phenylpropanoid Homeostasis in Arabidopsis. Journal of Biological Chemistry, 287 5434-5445.

Bosch, M., Hazen, S.P. 2013. Lignocellulosic feedstocks: research progress and challenges in optimising biomass quality and yield. Frontiers in Plant Science, 4.

Boyes, D.C., Zayed, A.M., Ascenzi, R., McCaskill, A.J., Hoffman, N.E., Davis, K.R., Gorlach, J. 2001. Growth stage-based phenotypic analysis of Arabidopsis: A model for high throughput functional genomics in plants. Plant Cell, 13, 1499-1510.

Bozell, J. J., Petersen, G. R. 2010. Technology development for the production of biobased products from biorefinery carbohydrates—the US Department of Energy's "top 10" revisited. Green Chemistry, 12, 539-554.

Brachi, B., Morris, G.P., Borevitz, J.O. 2011. Genome-wide association studies in plants: the missing heritability is in the field. Genome Biol, 12, 232.

Braidwood, L., Breuer, C., Sugimoto, K. 2014. Mybody is a cage: mechanisms and modulation of plant cell growth. New Phytologist, 201, 388-402.

Brandt, A.R., Millard-Ball, A., Ganser, M., Gorelick, S.M. 2013. Peak Oil Demand: The Role of Fuel Efficiency and Alternative Fuels in a Global Oil Production Decline. Environmental Science & Technology, 47, 8031-8041.

Brett, C.T., Waldron, K.W. 1998. Physiology and biochemistry of plant cell walls. Second ed. Springer.

Bringmann, M., Li, E.Y., Sampathkumar, A., Kocabek, T., Hauser, M.T., Persson, S. 2012. POM-POM2/CELLULOSE SYNTHASE INTERACTING1 Is Essential for the Functional Association of Cellulose Synthase and Microtubules in Arabidopsis. Plant Cell, 24, 163-177.

Bromley, J.R., Busse-Wicher, M., Tryfona, T., Mortimer, J.C., Zhang, Z.N., Brown, D.M., Dupree, P. 2013. GUX1 and GUX2 glucuronyltransferases decorate distinct domains of glucuronoxylan with different substitution patterns. Plant Journal, 74, 423-434.

Burn, J.E., Hocart, C.H., Birch, R.J., Cork, A.C., Williamson, R.E. 2002a. Functional analysis of the cellulose synthase genes CesA1, CesA2 and CesA3 in Arabidopsis. Plant Physiol, 129, 797-807.

Burn, J.E., Hurley, U.A., Birch, R.J., Arioli, T., Cork, A., Williamson, R.E. 2002b. The cellulose-deficient Arabidopsis mutant rsw3 is defective in a gene encoding a putative glucosidase

II, an enzyme processing N-glycans during ER quality control. Plant Journal, 32 949-960.

Burton, R.A., Fincher, G.B. 2014. Plant cell wall engineering: applications in biofuel production and improved human health. Current Opinion in Biotechnology, 26, 79-84.

Burton, R.A., Gidley, M. J., Fincher, G. B. 2010. Heterogeneity in the chemistry, structure and function of plant cell walls. Nature Chemical Biology, 6, 724-732.

Busse-Wicher, M., Gomes, T.C.F., Tryfona, T., Nikolovski, N., Stott, K., Grantham, N.J., Bolam, D.N., Skaf, M.S., Dupree, P. 2014. The pattern of xylan acetylation suggests xylan may interact with cellulose microfibrils as a two-fold helical screw in the secondary plant cell wall of Arabidopsis thaliana. The Plant Journal, 79, 492–506.

Caffall, K.H., Mohnen, D. 2009. The structure, function and biosynthesis of plant cell wall pectic polysaccharides. Carbohydrate Research, 344, 1879-1900.

Caffall, K.H., Pattathil, S., Phillips, S.E., Hahn, M.G., Mohnen, D. 2009. Arabidopsis thaliana T-DNA Mutants Implicate GAUT Genes in the Biosynthesis of Pectin and Xylan in Cell Walls and Seed Testa. Molecular Plant, 2, 1000-1014.

Cano-Delgado, A., Penfield, S., Smith, C., Catley, M., Bevan, M. 2003. Reduced cellulose synthesis invokes lignification and defense responses in Arabidopsis thaliana. Plant Journal, 34, 351-362.

Carpita, N.C. 2011. Update on Mechanisms of Plant Cell Wall Biosynthesis: How Plants Make Cellulose and Other (1 - > 4)-beta-D-Glycans. Plant Physiology, 155, 171-184.

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

Carpita, N.C., Gibeaut, D.M. 1993. Structural Models of Primary-Cell Walls in Flowering Plants - Consistency of Molecular-Structure with the Physical-Properties of the Walls during Growth. Plant Journal, 3, 1-30.

Castro, E., Díaz, M.J., Cara, C., Ruiz, E., Romero, I., Moya, M. 2011. Dilute acid pretreatment of rapeseed straw for fermentable sugar generation. Biores Tech. 102, 1270-6

Chan, E.K.F., Rowe, H.C., Kliebenstein, D.J. 2010. Understanding the evolution of defense metabolites in Arabidopsis thaliana using genome-wide association mapping. Genetics, 185, 991-1007.

Chanock, S.J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D.J., Thomas, G., Hirschhorn, J.N., Abecasis, G., Altshuler, D., Bailey-Wilson, J.E., Brooks, L.D., Cardon, L.R., Daly, M., Donnelly, P., Fraumeni, J.F., Freimer, N.B., Gerhard, D.S., Gunter, C., Guttmacher, A.E.,

Guyer, M.S., Harris, E.L., Hoh, J., Hoover, R., Kong, C.A., Merikangas, K.R., Morton, C.C., Palmer, L.J., Phimister, E.G., Rice, J.P., Roberts, J., Rotimi, C., Tucker, M.A., Vogan, K.J., Wacholder, S., Wijsman, E.M., Winn, D.M., Collins, F.S., Rep, N.-N.W.G. 2007. Replicating genotype-phenotype associations. Nature, 447, 655-660.

Chatterjee, M., Berbezy, P., Vyas, D., Coates, S., Barsby, T. 2005. Reduced expression of a protein homologous to glycogenin leads to reduction of starch content in Arabidopsis leaves. Plant Science, 168, 501-509.

Chen, J., Engle, N.L., Gunter, L.E., Jawdy, S., Tschaplinski, T.J., Tuskan, G.A. 2014. Key gene regulating cell wall biosynthesis and recalcitrance in poplus, Gene Y, US Patent 20,140,182,013.

Chen, L.M., Carpita, N.C., Reiter, W.D., Wilson, R.H., Jeffries, C., McCann, M.C. 1998. A rapid method to screen for cell-wall mutants using discriminant analysis of Fourier transform infrared spectra. Plant Journal, 16, 385-392.

Cheung, F., Trick, M., Drou, N., Lim, Y.P., Park, J.Y., Kwon, S.J., Kim, J.A., Scott, R., Pires, J.C., Paterson, A.H., Town, C., Bancroft, I. 2009. Comparative Analysis between Homoeologous Genome Segments of Brassica napus and Its Progenitor Species Reveals Extensive Sequence-Level Divergence. Plant Cell, 21, 1912-1928.

Chong, S.L., Virkki, L., Maaheimo, H., Juvonen, M., Derba-Maceluch, M., Koutaniemi, S., Roach, M., Sundberg, B., Tuomainen, P., Mellerowicz, E.J., Tenkanen, M. 2014. O-Acetylation of glucuronoxylan in Arabidopsis thaliana wild type and its change in xylan biosynthesis mutants. Glycobiology, 24, 494-506.

Chou, Y.-H., Pogorelko, G., Zabotina, O.A. 2012. Xyloglucan Xylosyltransferases XXT1, XXT2 and XXT5 and the Glucan Synthase CSLC4 Form Golgi-Localized Multiprotein Complexes. Plant Physiology, 159, 1355-1366.

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

Clark, G.B., Sessions, A., Eastburn, D.J., Roux, S.J. 2001. Differential expression of members of the annexin multigene family in Arabidopsis. Plant Physiology, 126, 1072-1084.

Cobb, J.N., DeClerck, G., Greenberg, A., Clark, R., McCouch, S. 2013. Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. Theoretical and Applied Genetics, 126, 867-887.

Coleman, H.D., Yan, J., Mansfield, S.D. 2009. Sucrose synthase affects carbon partitioning to increase cellulose production and altered cell wall ultrastructure. Proceedings of the National Academy of Sciences, 106, 13118-13123.

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

Collings, David A., Leigh K. Gebbie, Paul A. Howles, Ursula A. Hurley, Rosemary J. Birch, Ann H. Cork, Charles H. Hocart, Tony Arioli, and Richard E. Williamson. 2008. Arabidopsis dynamin-like protein DRP1A: a null mutant with widespread defects in endocytosis, cellulose synthesis, cytokinesis, and cell expansion. Journal of experimental botany, 59, 361-376.

Collins, S. R., Wellner, N., Bordonado, I. M., Harper, A. L., Miller, C. N., Bancroft, I., & Waldron, K. W. 2014. Variation in the chemical composition of wheat straw: the role of tissue ratio and composition. Biotechnology for Biofuels, 7, 121.

Conde-Mejía, C., Jiménez-Gutiérrez, A., El-Halwagi, M. 2011. A comparison of pretreatment methods for bioethanol production from lignocellulosic materials. Process Safety and Environmental Protection. 90, 189–202.

Copeland, J., Turley, D. 2008. National and regional supply/demand balance for agricultural straw in Great Britain

Cosgrove, D.J. 2005. Growth of the plant cell wall. Nature Reviews Molecular Cell Biology, 6, 850-861.

Cosgrove, D.J., Jarvis, M.C. 2012. Comparative structure and biomechanics of plant primary and secondary cell walls. Front Plant Sci, 3, 204.

Dashtban, M., Maki, M., Leung, K.T., Mao, C.Q., Qin, W.S. 2010. Cellulase activities in biomass conversion: measurement methods and comparison. Critical Reviews in Biotechnology, 30, 302-309.

Davin, L.B., Lewis, N.G. 2000. Dirigent proteins and dirigent sites explain the mystery of specificity of radical precursor coupling in lignan and lignin biosynthesis. Plant Physiology, 123, 453-462.

Davison, B. H., Drescher, S. R., Tuskan, G. A., Davis, M. F., & Nghiem, N. P. 2006. Variation of S/G ratio and lignin content in a Populus family influences the release of xylose by dilute acid hydrolysis. In Twenty-Seventh Symposium on Biotechnology for Fuels and Chemicals. Humana Press. 427-435.

Decker, S., Adney, W., Jennings, E., Vinzant, T., Himmel, M. 2003. Automated filter paper assay for determination of cellulase activity. Applied Biochemistry and Biotechnology, 107, 689-703.

Decker, S., Brunecky, R., Tucker, M., Himmel, M., Selig, M. 2009. High throughput Screening Techniques for Biomass Conversion. BioEnergy Research, 2, 179-192.

Decker, S., Carlile, M., Selig, M., Doeppke, C., Davis, M., Sykes, R., Turner, G., Ziebell, A. 2012. Reducing the Effect of Variable Starch Levels in Biomass Recalcitrance Screening. in: Biomass Conversion, (Ed.) M.E. Himmel, Humana Press, 908, 181-195.

Del Rio, J.C., Marques, G., Rencoret, J., Martinez, A.T., Gutierrez, A. 2007. Occurrence of naturally acetylated lignin units. Journal of Agricultural and Food Chemistry, 55, 5461-5468.

Delatte, T., Trevisan, M., Parker, M.L., Zeeman, S.C. 2005. Arabidopsis mutants Atisa1 and Atisa2 have identical phenotypes and lack the same multimeric isoamylase, which influences the branch point distribution of amylopectin during starch synthesis. Plant Journal, 41, 815-830.

Delmer, D.P., Amor, Y., Audrawis, A., Solomon, M., Potikha, T., Ohana, P., Mayer, R., Benziman, M. 1993. Synthesis of Cellulose and Callose in Higher-Plants. Journal of Cellular Biochemistry, 115, 643–656.

Delourme, R., Falentin, C., Fomeju, B.F., Boillot, M., Lassalle, G. andre, I., Duarte, J., Gauthier, V., Lucante, N., Marty, A., Pauchon, M., Pichon, J.P., Ribiere, N., Trotoux, G., Blanchard, P., Riviere, N., Martinant, J.P., Pauquet, J. 2013. High-density SNP-based genetic map development and linkage disequilibrium assessment in Brassica napus L. BMC Genomics, 14, 120

Demura, T., Ye, Z.-H. 2010. Regulation of plant biomass production. Current Opinion in Plant Biology, 13, 298-303.

Department for Business Innovation and Skills. 2010. The allocation of science and research funding 2011/12 to 2014/15, (Ed.) I.a.S. Department for Business. http://www.bis.gov.uk/assets/biscore/science/docs/a/10-1356-allocation-of-science-and-research-funding-2011-2015.pdf.

Department of Energy & Climate Change. 2013. Renewable energy roadmap update 2013, www.gov.uk/government/uploads/system/uploads/attachment_data/file/255182/UK_Renewable_Energy_Roadmap_-_5_November_-_FINAL_DOCUMENT_FOR_PUBLICATIO___.pdf.

Diaz, M.J., Cara, C., Ruiz, E., Romero, I., Moya, M., Castro, E. 2010. Hydrothermal pre-treatment of rapeseed straw. Bioresource Technology, 101, 2428-2435.

Dick-Perez, M., Zhang, Y.A., Hayes, J., Salazar, A., Zabotina, O.A., Hong, M. 2011. Structure and Interactions of Plant Cell-Wall Polysaccharides by Two- and Three-Dimensional Magic-Angle-Spinning Solid-State NMR. Biochemistry, 50, 989-1000.

Diepenbrock W. 2000. Yield analysis of winter oilseed rape (Brassica napus L.): A review. Field Crop Res. 67, 35-49.

Doering, A., Lathe, R., Persson, S. 2012. An Update on Xylan Synthesis. Molecular Plant, 5, 769-771.

Dolezal, O., Cobbett, C.S. 1991. Arabinose Kinase-Deficient Mutant of Arabidopsis-Thaliana. Plant Physiology, 96, 1255-1260.

Driouich, A., follet-gueye, m.-l., Bernard, s., kousar, s., Chevalier, L., Vicré, M., Lerouxel, O. 2012. Golgi-mediated synthesis and secretion of matrix polysaccharides of the primary cell wall of higher plants. Frontiers in Plant Science, 3, 79.

Dunwell, J.M. 2014. Genetically modified (GM) crops: European and transatlantic divisions. Molecular Plant Pathology, 15, 119-121.

Dupree, P., Miles, G.P. 2009. Modified xylan production, Patent number: WO2009037502 A1.

Ellis, M., Egelund, J., Schultz, C.J., Bacic, A. 2010. Arabinogalactan-proteins: key regulators at the cell surface? Plant Physiology, 153, 403-419.

Elliston, A., Collins, S.R., Wilson, D.R., Roberts, I.N., Waldron, K.W. 2013. High concentrations of cellulosic ethanol achieved by fed batch semi simultaneous saccharification and fermentation of waste-paper. Bioresource technology, 134, 117-126.

Endler, A., Persson, S. 2011. Cellulose Synthases and Synthesis in Arabidopsis. Molecular Plant, 4, 199-211.

European Commision, The. 2012. Analysis of options beyond 20% GHG emission reductions: Member State results. Commision staff working paper.

EuroStat. 2013. Energy, transport and environment indicators (2013 edition). in: Eurostat Pocketbooks, Publications Office of the European Union. Luxembourg.

FAOSTAT. 2014. Available at "http://faostat.fao.org/"

Farrokhi, N., Burton, R.A., Brownfield, L., Hrmova, M., Wilson, S.M., Bacic, A., Fincher, G.B. 2006. Plant cell wall biosynthesis: genetic, biochemical and functional genomics approaches to the identification of key genes. Plant Biotechnology Journal, 4, 145-167.

Favaro, L., Basaglia, M., Casella, S. 2014. Innately robust yeast strains isolated from grape marc have a great potential for lignocellulosic ethanol production. Annals of Microbiology, 1-12.

Fernandez-Calvo, P., Chini, A., Fernandez-Barbero, G., Chico, J.M., Gimenez-Ibanez, S., Geerinck, J., Eeckhout, D., Schweizer, F., Godoy, M., Franco-Zorrilla, J.M., Pauwels, L., Witters, E., Puga, M.I., Paz-Ares, J., Goossens, A., Reymond, P., De Jaeger, G., Solano, R. 2011. The Arabidopsis bHLH Transcription Factors MYC3 and MYC4 Are Targets of JAZ Repressors and Act Additively with MYC2 in the Activation of Jasmonate Responses. Plant Cell, 23, 701-715.

Francocci, F., Bastianelli, E., Lionetti, V., Ferrari, S., De Lorenzo, G., Bellincampi, D., Cervone, F. 2013. Analysis of pectin mutants and natural accessions of Arabidopsis highlights the impact of de-methyl-esterified homogalacturonan on tissue saccharification. Biotechnology for Biofuels, 6, 163.

Frankova, L., Fry, S.C. 2013. Biochemistry and physiological roles of enzymes that cut and paste plant cell-wall polysaccharides. Journal of Experimental Botany, 64, 3519-3550.

Fuentes, S., Pires, N., Ostergaard, L. 2010. A clade in the QUASIMODO2 family evolved with vascular plants and supports a role for cell wall composition in adaptation to environmental changes. Plant Molecular Biology, 73, 605-615.

Gang, D.R., Costa, M.A., Fujita, M., Dinkova-Kostova, A.T., Wang, H.-B., Burlat, V., Martin, W., Sarkanen, S., Davin, L.B., Lewis, N.G. 1999. Regiochemical control of monolignol radical coupling: a new paradigm for lignin and lignan biosynthesis. Chemistry & biology, 6, 143-151.

Gasparatos, A., Stromberg, P., Takeuchi, K. 2013. Sustainability impacts of first-generation biofuels. Animal Frontiers, 3, 12-26.

Gerber, L., Zhang, B., Roach, M., Rende, U., Gorzsás, A., Kumar, M., Burgert, I., Niittylä, T., Sundberg, B. 2014. Deficient sucrose synthase activity in developing wood does not specifically affect cellulose biosynthesis, but causes an overall decrease in cell wall polymers. New Phytologist, 203, 1220-30

Geserick, C., Tenhaken, R. 2013. UDP-sugar pyrophosphorylase is essential for arabinose and xylose recycling and is required during vegetative and reproductive growth in Arabidopsis. Plant Journal, 74, 239-247.

Ghose, T.K. 1987. Measurement of cellulase activities. Pure Appl Chem, 2, 257-268.

Gille, S., de Souza, A., Xiong, G.Y., Benz, M., Cheng, K., Schultink, A., Reca, I.B., Pauly, M. 2011. O-Acetylation of Arabidopsis Hemicellulose Xyloglucan Requires AXY4 or AXY4L, Proteins with a TBL and DUF231 Domain. Plant Cell, 23, 4041-4053.

Gille, S., Pauly, M. 2012a. Decreased polysaccharide o-acetylation. USA Patent number: 61/476, 155.

Gille, S., Pauly, M. 2012b. O-acetylation of Plant Cell Wall Polysaccharides. Frontiers in Plant Science, 3.

Glithero, N.J., Wilson, P., Ramsden, S.J. 2013. Straw use and availability for second generation biofuels in England. Biomass & Bioenergy, 55, 311-321.

Gomez, L., Whitehead, C., Barakate, A., Halpin, C., McQueen-Mason, S. 2010. Automated saccharification assay for determination of digestibility in plant materials. Biotechnology for Biofuels, 3, 23.

Gomez, L.D., Steele-King, C.G., McQueen-Mason, S.J. 2008. Sustainable liquid biofuels from biomass: the writing's on the walls. New Phytologist, 178, 473-485.

Goncalves, C., Rodriguez-Jasso, R.M., Gomes, N., Teixeira, J.A., Belo, I. 2010. Adaptation of dinitrosalicylic acid method to microtiter plates. Analytical Methods, 2, 2046-2048.

Gorshkova, T.A., Kozlova, L.V., Mikshina, P.V. 2013. Spatial structure of plant cell wall polysaccharides and its functional significance. Biochemistry-Moscow, 78, 836-853.

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

Goujon, T., Sibout, R., Eudes, A., MacKay, J., Joulanin, L. 2003. Genes involved in the biosynthesis of lignin precursors in Arabidopsis thaliana. Plant Physiology and Biochemistry, 41, 677-687.

Gressel, J. 2008. Transgenics are imperative for biofuel crops. Plant Science, 174, 246-263.

Grigg, S.P., Canales, C., Hay, A., Tsiantis, M. 2005. SERRATE coordinates shoot meristem function and leaf axial patterning in Arabidopsis. Nature, 437, 1022-1026.

Hall, H., Ellis, B. 2013. Transcriptional programming during cell wall maturation in the expanding Arabidopsis stem. Bmc Plant Biology, 13.

Hamilton, J.D. 2011. Historical oil shocks. National Bureau of Economic Research. NBER Working Paper No. 16790.

Handford, M., Baldwin, T., Goubet, F., Prime, T., Miles, J., Yu, X., Dupree, P. 2003. Localisation and characterisation of cell wall mannan polysaccharides in Arabidopsis thaliana. Planta, 218, 27-36.

Hansen, J., Sato, M., Ruedy, R., Lo, K., Lea, D. W., Medina-Elizade, M. (2006). Global temperature change. Proceedings of the National Academy of Sciences, 103, 14288-14293.

Hansen, M.A., Hidayat, B.J., Mogensen, K.K., Jeppesen, M.D., Jorgensen, B., Johansen, K.S., Thygesen, L.G. 2013. Enzyme affinity to cell types in wheat straw (Triticum aestivum L.) before and after hydrothermal pretreatment. Biotechnol Biofuels, 6, 54.

Harholt, J., Suttangkakul, A., Scheller, H.V. 2010. Biosynthesis of Pectin. Plant Physiology, 153, 384-395.

Harper, A.L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., Hattori, C., Werner, P., Bancroft, I. 2012. Associative transcriptomics of traits in the polyploid crop species Brassica napus. Nature Biotechnology, 30, 798-802.

Hayashi, T., Kaida, R. 2011. Functions of Xyloglucan in Plant Cells. Molecular Plant, 4, 17-24.

Hayward, A., Morgan, J.D., Edwards, D. 2012. SNP discovery and applications in Brassica napus. Journal of Plant Biotechnology, 39, 49-61.

Hendriks, A.T.W.M., Zeeman, G. 2009. Pretreatments to enhance the digestibility of lignocellulosic biomass. Bioresource Technology, 100, 10-18.

Hijazi, M., Roujol, D., Nguyen-Kim, H., del Rocio Cisneros Castillo, L., Saland, E., Jamet, E., Albenne, C. 2014. Arabinogalactan protein 31 (AGP31), a putative network-forming protein in Arabidopsis thaliana cell walls? Annals of Botany. [Epub ahead of print]

Himmel, M.E., Ding, S.Y., Johnson, D.K., Adney, W.S., Nimlos, M.R., Brady, J.W., Foust, T.D. 2007. Biomass recalcitrance: Engineering plants and enzymes for biofuels production. Science, 315, 804-807.

Hemmerlin, A., Tritsch, D., Hartmann, M., Pacaud, K., Hoeffler, J. F., van Dorsselaer, A., Bach, T. J. 2006. A cytosolic Arabidopsis D-xylulose kinase catalyzes the phosphorylation of 1-deoxy-D-xylulose into a precursor of the plastidial isoprenoid pathway. Plant physiology, 142, 441-457.

Höök, M., Tang, X. 2013. Depletion of fossil fuels and anthropogenic climate change—A review. Energy Policy, 52, 797-809.

Horn, S. J., Nguyen, Q. D., Westereng, B., Nilsen, P. J., & Eijsink, V. G. 2011. Screening of steam explosion conditions for glucose production from non-impregnated wheat straw. Biomass and Bioenergy, 35, 4879-4886.

Howarth, R.W., Ingraffea, A., Engelder, T. 2011. Natural gas: Should fracking stop? Nature, 477, 271-275.

Hu, J.G., Arantes, V., Pribowo, A., Saddler, J.N. 2013. The synergistic action of accessory enzymes enhances the hydrolytic potential of a "cellulase mixture" but is highly substrate specific. Biotechnology for Biofuels, 6.

Hu, J.G., Arantes, V., Saddler, J.N. 2011. The enhancement of enzymatic hydrolysis of lignocellulosic substrates by the addition of accessory enzymes such as xylanase: is it an additive or synergistic effect? Biotechnology for Biofuels, 4.

Huang, H.-J., Ramaswamy, S., Tschirner, U., Ramarao, B. 2010. Separation and purification processes for lignocellulose-to-bioethanol production. in: Bioalcohol Production: Biochemical Conversion of Lignocellulosic Biomass, (Ed.) K. Waldron, Woodhead Publishing Ltd., 246-269.

Huang, X., Han, B. 2013. Natural Variations and Genome-Wide Association Studies in Crop Plants. Annual review of plant biology , 65, 531-551.

Hubbert, M. 1956. Nuclear Energy and the fossil fuels. American Petroleum Insititute Drilling Production and Practice, Proceedings of Spring Meeting, San Antonio, Texas, 7-25.

Hughes, T. P., Linares, C., Dakos, V., van de Leemput, I. A., & van Nes, E. H. 2013. Living dangerously on borrowed time during slow, unrecognized regime shifts. Trends in Ecology & Evolution, 28, 149-155.

Hussey, S.G., Mizrachi, E., Creux, N.M., Myburg, A.A. 2013. Navigating the transcriptional roadmap regulating plant secondary cell wall deposition. Frontiers in Plant Science, 4.

Iandolo, D., Piscitelli, A., Sannia, G., Faraco, V. 2011. Enzyme Production by Solid Substrate Fermentation of Pleurotus ostreatus and Trametes versicolor on Tomato Pomace. Applied Biochemistry and Biotechnology, 163, 40-51.

Igarashi, K., Uchihashi, T., Koivula, A., Wada, M., Kimura, S., Okamoto, T., Penttila, M. ando, T., Samejima, M. 2011. Traffic Jams Reduce Hydrolytic Efficiency of Cellulase on Cellulose Surface. Science, 333, 1279-1282.

Iniguez-Luy, F.L., Federico, M.L. 2011. The genetics of Brassica napus. in: Genetics and Genomics of the Brassicaceae, Springer, 291-322.

IPCC. 2014. Summary for Policymakers. in: Climate Change 2014: Impacts, Adaptation and Vulnerability.

Isci, A., Murphy, P.T., Anex, R.P., Moore, K.J. 2008. A Rapid Simultaneous Saccharification and Fermentation (SSF) Technique to Determine Ethanol Yields. Bioenergy Research, 1, 163-169.

Ishizawa, C.I., Jeoh, T., Adney, W.S., Himmel, M.E., Johnson, D.K., Davis, M.F. 2009. Can delignification decrease cellulose digestibility in acid pretreated corn stover? Cellulose, 16, 677-686.

Jacquet, N., Quiévy, N., Vanderghem, C., Janas, S., Blecker, C., Wathelet, B. Paquot, M. 2011. Influence of steam explosion on the thermal stability of cellulose fibres. Polymer Degradation and Stability, 96, 1582-1588.

Jackson, S., Nicolson, S.W. 2002. Xylose as a nectar sugar: from biochemistry to ecology. Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology, 131, 613-620.

Jamet, E., Canut, H., Boudart, G., Pont-Lezica, R.F. 2006. Cell wall proteins: a new insight through proteomics. Trends in Plant Science, 11, 33-39.

Jeong, T.S., Um, B.H., Kim, J.S., Oh, K.K. 2010. Optimizing Dilute-Acid Pretreatment of Rapeseed Straw for Extraction of Hemicellulose. Applied Biochemistry and Biotechnology, 161, 22-33.

Jeong, T.S., Oh, K.K. 2011a. Optimization of fermentable sugar production from rape straw through hydrothermal acid pretreatment. Bioresource Technology, 102, 9261-9266.

Jin, H.L., Cominelli, E., Bailey, P., Parr, A., Mehrtens, F., Jones, J., Tonelli, C., Weisshaar, B., Martin, C. 2000. Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis. Embo Journal, 19, 6150-6161.

Johnston, J.S., Pepper, A.E., Hall, A.E., Chen, Z.J., Hodnett, G., Drabek, J., Lopez, R., Price, H.J. 2005. Evolution of genome size in Brassicaceae. Annals of botany, 95, 229-235.

Jolie, R.P., Duvetter, T., Van Loey, A.M., Hendrickx, M.E. 2010. Pectin methylesterase and its proteinaceous inhibitor: a review. Carbohydrate Research, 345, 2583-2595.

Jones, N. 2013. Troubling milestone for CO2. Nature Geosci, 6, 589-589.

Jonsson, L.J., Alriksson, B., Nilvebrant, N.O. 2013. Bioconversion of lignocellulose: inhibitors and detoxification. Biotechnology for Biofuels, 6, 16.

Kabel, M.A., Bos, G., Zeevalking, J., Voragen, A.G.J., Schols, H.A. 2007. Effect of pretreatment severity on xylan solubility and enzymatic breakdown of the remaining cellulose from wheat straw. Bioresource Technology, 98, 2034-2042.

Kabel, M.A., van der Maarel, M.J.E.C., Klip, G., Voragen, A.G.J., Schols, H.A. 2006. Standard assays do not predict the efficiency of commercial cellulase preparations towards plant materials. Biotechnology and Bioengineering, 93, 56-63.

Kacurakova, M., Belton, P.S., Wilson, R.H., Hirsch, J., Ebringerova, A. 1998. Hydration properties of xylan-type structures: an FTIR study of xylooligosaccharides. Journal of the Science of Food and Agriculture, 77, 38-44.

Kacurakova, M., Capek, P., Sasinkova, V., Wellner, N., Ebringerova, A. 2000. FT-IR study of plant cell wall model compounds: pectic polysaccharides and hemicelluloses. Carbohydrate Polymers, 43, 195-203.

Kacurakova, M., Wellner, N., Ebringerova, A., Hromadkova, Z., Wilson, R.H., Belton, P.S. 1999. Characterisation of xylan-type polysaccharides and associated cell wall components by FT-IR and FT-Raman spectroscopies. Food Hydrocolloids, 13, 35-41.

Kang, J., Park, J., Choi, H., Burla, B., Kretzschmar, T., Lee, Y., Martinoia, E. 2011. Plant ABC transporters. The Arabidopsis book/American Society of Plant Biologists, 9.

Katzen, R., Madson, P., Moon, G. 1999. Ethanol distilation: the fundamentals. in: The Alcohol Textbook, 269-299.

Keeling, R.F. 2008. Atmospheric science - Recording Earth's vital signs. Science, 319, 1771-1772.

Keller, B. 1993. Structural Cell-Wall Proteins. Plant Physiology, 101, 1127-1130.

Kim, I. J., Lee, H. J., Choi, I. G., & Kim, K. H. 2014. Synergistic proteins for the enhanced enzymatic hydrolysis of cellulose by cellulase. Applied microbiology and biotechnology, 1-12.

King, B.C., Donnelly, M.K., Bergstrom, G.C., Walker, L.P., Gibson, D.M. 2009. An optimized microplate assay system for quantitative evaluation of plant cell wall-degrading enzyme activity of fungal culture extracts. Biotechnology and Bioengineering, 102, 1033-1044.

Kintisch, E. 2008. Sowing the seeds for high-energy plants. Science, 320, 478.

Klasson, K.T., Uchimiya, M., Lima, I.M., Boihem, L.L. 2011. Feasibility of Removing Furfurals from Sugar Solutions Using Activated Biochars Made from Agricultural Residues. Bioresources, 6, 3242-3251.

Kleczkowski, L.A., Geisler, M., Ciereszko, I., Johansson, H. 2004. UDP-Glucose pyrophosphorylase. An old protein with new tricks. Plant Physiology, 134, 912-918.

Korte, A., Farlow, A. 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant methods, 9, 29.

Kotake, T., Hojo, S., Tajima, N., Matsuoka, K., Koyama, T., Tsumuraya, Y. 2008. A bifunctional enzyme with L-fucokinase and GDP-L-fucose pyrophosphorylase activities salvages free L-fucose in Arabidopsis. Journal of Biological Chemistry, 283, 8125-8135.

Kotake, T., Takata, R., Verma, R., Takaba, M., Yamaguchi, D., Orita, T., Kaneko, S., Matsuoka, K., Koyama, T., Reiter, W.D., Tsumuraya, Y. 2009. Bifunctional cytosolic UDP-glucose 4-epimerases catalyse the interconversion between UDP-D-xylose and UDP-L-arabinose in plants. Biochemical Journal, 424, 169-177.

Kubacka-Zebalska, M., Kacperska, A. 1999. Low temperature-induced modifications of cell wall content and polysaccharide composition in leaves of winter oilseed rape (Brassica napus L-var. oleifera L.). Plant Science, 148, 59-67.

Kumar, D., Murthy, G. 2011. Impact of pretreatment and downstream processing technologies on economics and energy use in cellulosic ethanol production. Biotechnology for Biofuels, 4, 27.

Kumar, P., Barrett, D.M., Delwiche, M.J., Stroeve, P. 2009. Methods for Pretreatment of Lignocellulosic Biomass for Efficient Hydrolysis and Biofuel Production. Industrial & Engineering Chemistry Research, 48, 3713-3729.

Kumar, R., Hu, F., Sannigrahi, P., Jung, S., Ragauskas, A.J., Wyman, C.E. 2013. Carbohydrate derived-pseudo-lignin can retard cellulose biological conversion. Biotechnology and Bioengineering, 110, 737-753.

Kumar, R., Wyman, C.E. 2009. Cellulase Adsorption and Relationship to Features of Corn Stover Solids Produced by Leading Pretreatments. Biotechnology and Bioengineering, 103, 252-267.

Lambert, C.G., Black, L.J. 2012. Learning from our GWAS mistakes: from experimental design to scientific method. Biostatistics, 13, 195-203.

Lamesch, P., Berardini, T.Z., Li, D.H., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M., Karthikeyan, A.S., Lee, C.H., Nelson, W.D., Ploetz, L., Singh, S., Wensel, A., Huala, E. 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Research, 40, 1202-1210.

Larsen, S.U., Bruun, S., Lindedam, J. 2012. Straw yield and saccharification potential for ethanol in cereal species and wheat cultivars. Biomass & Bioenergy, 45, 239-250.

Laux, T., Mayer, K.F.X., Berger, J., Jurgens, G. 1996. The WUSCHEL gene is required for shoot and floral meristem integrity in Arabidopsis. Development, 122, 87-96.

Lee, C., Teng, Q., Zhong, R., Ye, Z.H. 2011a. The four Arabidopsis reduced wall acetylation genes are expressed in secondary wall-containing cells and required for the acetylation of xylan. Plant Cell Physiol, 52, 1289-301.

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

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

Lee, S., Speight, J.G., Loyalka, S.K. 2007a. Handbook of alternative fuel technologies. crc Press.

Lee, E.J., Matsumura, Y., Soga, K., Hoson, T., Koizumi, N. 2007b. Glycosyl hydrolases of cell wall are induced by sugar starvation in Arabidopsis. Plant and Cell Physiology, 48, 405-413.

Leitch, A.R., Leitch, I.J. 2008. Perspective - Genomic plasticity and the diversity of polyploid plants. Science, 320, 481-483.

Li, L., Ilarslan, H., James, M. G., Myers, A. M., & Wurtele, E. S. 2007. Genome wide co-expression among the starch debranching enzyme genes AtISA1, AtISA2, and AtISA3 in Arabidopsis thaliana. Journal of experimental botany, 58, 3323-3342.

Li, X., Ximenes, E., Kim, Y., Slininger, M., Meilan, R., Ladisch, M., & Chapple, C. 2010. Lignin monomer composition affects Arabidopsis cell-wall degradability after liquid hot water pretreatment. Biotechnol Biofuels, 3, 27.

Li, S., Bashline, L., Lei, L., Gu, Y. 2014. Cellulose Synthesis and Its Regulation. The Arabidopsis Book.

Li, S.D., Lei, L., Somerville, C.R., Gu, Y. 2012. Cellulose synthase interactive protein 1 (CSI1) links microtubules and cellulose synthase complexes. Proceedings of the National Academy of Sciences of the United States of America, 109, 185-190.

Lin, Y., Tanaka, S. 2006. Ethanol fermentation from biomass resources: current state and prospects. Appl Microbiol Biotechnol, 69, 627-642.

Lindedam, J. andersen, S.B., DeMartini, J., Bruun, S., Jorgensen, H., Felby, C., Magid, J., Yang, B., Wyman, C.E. 2012. Cultivar variation and selection potential relevant to the production of cellulosic ethanol from wheat straw. Biomass & Bioenergy, 37, 221-228.

Lindedam, J., Bruun, S., Jørgensen, H., Decker, S.R., Turner, G.B., DeMartini, J.D., Wyman, C.E., Felby, C. 2014. Evaluation of high throughput screening methods in picking up differences between cultivars of lignocellulosic biomass for ethanol production. Biomass and Bioenergy 66, 261-267.

Lindedam, J., Bruun, S., Jorgensen, H., Felby, C., Magid, J. 2010. Cellulosic ethanol: Interactions between cultivar and enzyme loading in wheat straw processing. Biotechnology for Biofuels, 3, 25.

Ling, H., Zhao, J.Y., Zuo, K.J., Qiu, C.X., Yao, H.Y., Qin, J., Sun, X.F., Tang, K.X. 2006. Isolation and expression analysis of a GDSL-like lipase gene from Brassica napus L. Journal of Biochemistry and Molecular Biology, 39, 297-303.

Lins, C. 2010. Renewable Energy Technology Roadmap: 20% RES by 2020. Local Governments and Climate Change: Sustainable Energy Planning and Implementation in Small and Medium Sized Communities, 39, 131-162.

Lionetti, V., Francocci, F., Ferrari, S., Volpi, C., Bellincampi, D., Galletti, R., D'Ovidio, R., De Lorenzo, G., Cervone, F. 2010. Engineering the cell wall by reducing de-methyl-esterified homogalacturonan improves saccharification of plant tissues for bioconversion. Proceedings of the National Academy of Sciences of the United States of America, 107, 616-621.

Liu, C.J., Miao, Y.C., Zhang, K.W. 2011. Sequestration and Transport of Lignin Monomeric Precursors. Molecules, 16, 710-727.

Liu, Y.J., Papasian, C.J., Liu, J.F., Hamilton, J., Deng, H.W. 2008. Is Replication the Gold Standard for Validating Genome-Wide Association Findings? Plos One, 3.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., Henrissat, B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research, 42, 490-495.

Lopez-Casado, G., Urbanowicz, B. R., Damasceno, C., & Rose, J. K. 2008. Plant glycosyl hydrolases and biofuels: a natural marriage. Current opinion in plant biology, 11, 329-337.

Lopez-Linares, J.C., Cara, C., Moya, M., Ruiz, E., Castro, E., Romero, I. 2013a. Fermentable sugar production from rapeseed straw by dilute phosphoric acid pretreatment. Industrial Crops and Products, 50, 525-531.

López-Linares, J.C., Romero, I., Cara, C., Ruiz, E., Castro, E., Moya, M. 2013b. Experimental study on ethanol production from hydrothermal pretreated rapeseed straw by simultaneous saccharification and fermentation. Journal of Chemical Technology & Biotechnology, 89, 104-110.

Lopez Ferreria, N., Margeot, A., Blanquet, S., Berrin, J.-G. 2014. Use of Cellulase from Trichoderma reesei in the twenty-first century - Part I: Current industrial uses and future applications in the production of second ethanol generation. in: Biotechnology

and Biology of Trichoderma, (Eds.) V.K. Gupta, M. Schmoll, A. Herrera-Estrella, R.S. Upadhyay, I. Druzhinina, M. Tuohy, Elsevier Science.

Lorenz, A.J., Anex, R.P., Isci, A., Coors, J.G., de Leon, N., Weimer, P.J. 2009. Forage quality and composition measurements as predictors of ethanol yield from maize (Zea mays L.) stover. Biotechnology for Biofuels, 2.

Lu, Y., Yang, B., Gregg, D., Saddler, J. N., & Mansfield, S. D. 2002. Cellulase adsorption and an evaluation of enzyme recycle during hydrolysis of steam-exploded softwood residues. Applied biochemistry and biotechnology, 98, 641-654.

Lu, X., Zhang, Y., Angelidaki, I. 2009. Optimization of H_2SO_4-catalyzed hydrothermal pretreatment of rapeseed straw for bioconversion to ethanol: Focusing on pretreatment at high solids content. Bioresource Technology, 100, 3048-3053.

Lu, X., Xi, B., Zhang, Y., Angelidaki, I. 2011. Microwave pretreatment of rape straw for bioethanol production: Focus on energy efficiency. Bioresource Technology, 102, 7937-7940.

Lubieniechi, S., Peranantham, T., Levin, B.D. 2013. Recent patents on genetic modification of plants and microbes for biomass conversion to biofuels. Recent patents on DNA & gene sequences, 7, 25-35.

Lunn, J.E., Ashton, A.R., Hatch, M.D., Heldt, H.W. 2000. Purification, molecular cloning and sequence analysis of sucrose-6F-phosphate phosphohydrolase from plants. Proceedings of the National Academy of Sciences, 97, 12914-12919.

Maass, P. 2005. The Breaking Point. in: New York Times. New York, USA. published: 21 Aug 2005.

Mackay, I., Powell, W. 2007. Methods for linkage disequilibrium mapping in crops. Trends in plant science, 12, 57-63.

Maehara, T., Takabatake, K., Kaneko, S. 2013. Expression of Arabidopsis thaliana xylose isomerase gene and its effect on ethanol production in Flammulina velutipes. Fungal Biology, 117, 776-782.

Malagoli, P., Laine, P., Rossato, L., & Ourry, A. 2005. Dynamics of nitrogen uptake and mobilization in field-grown winter oilseed rape (Brassica napus) from stem extension to harvest. II. An 15N-labelling-based simulation model of N partitioning between vegetative and reproductive tissues. Annals of Botany, 95, 1187-1198.

Manabe, Y., Nafisi, M., Verhertbruggen, Y., Orfila, C., Gille, S., Rautengarten, C., Cherk, C., Marcus, S.E., Somerville, S., Pauly, M., Knox, J.P., Sakuragi, Y., Scheller, H.V. 2011. Loss-of-function mutation of REDUCED WALL ACETYLATION2 in Arabidopsis leads to

reduced cell wall acetylation and increased resistance to Botrytis cinerea. Plant Physiol, 155, 1068-78.

Manabe, Y., Verhertbruggen, Y., Gille, S., Harholt, J., Chong, S.L., Pawar, P.M., Mellerowicz, E.J., Tenkanen, M., Cheng, K., Pauly, M., Scheller, H.V. 2013. Reduced wall acetylation proteins play vital and distinct roles in cell wall o-acetylation in Arabidopsis. Plant Physiol, 163, 1107-17.

Mansfield, S.D., Mooney, C., Saddler, J.N. 1999. Substrate and enzyme characteristics that limit cellulose hydrolysis. Biotechnology Progress, 15, 804-816.

Marcus, S. E., Verhertbruggen, Y., Hervé, C., Ordaz-Ortiz, J. J., Farkas, V., Pedersen, H. L., Knox, J. P. 2008. Pectic homogalacturonan masks abundant sets of xyloglucan epitopes in plant cell walls. BMC Plant Biology, 8, 60.

Marschner, H., Rimmington, G. 1996. Mineral nutrition of higher plants, Wiley Online Library.

Massman, J.M., Jung, H.-J.G., Bernardo, R. 2013. Genomewide Selection versus Marker-assisted Recurrent Selection to Improve Grain Yield and Stover-quality Traits for Cellulosic Ethanol in Maize. Crop Sci., 53, 58-66.

Mathew, A.K., Chaney, K., Crook, M., Humphries, A.C. 2011a. Alkaline pre-treatment of oilseed rape straw for bioethanol production: Evaluation of glucose yield and pre-treatment energy consumption. Bioresource Technology, 102, 6547-6553.

Mathew, A.K., Chaney, K., Crook, M., Humphries, A.C. 2011b. Dilute acid pre-treatment of oilseed rape straw for bioethanol production. Renewable Energy, 36, 2424-2432.

Mathew, A.K., Crook, M., Chaney, K., Humphries, A.C. 2013. Comparison of entrapment and biofilm mode of immobilisation for bioethanol production from oilseed rape straw using Saccharomyces cerevisiae cells. Biomass & Bioenergy, 52, 1-7.

Mathew, A.K., Crook, M., Chaney, K., Humphries, A.C. 2014. Continuous bioethanol production from oilseed rape straw hydrosylate using immobilised Saccharomyces cerevisiae cells. Bioresource Technology, 154, 248-253.

Matsuda, F., Yamasaki, M., Hasunuma, T., Ogino, C., Kondo, A. 2011. Variation in Biomass Properties among Rice Diverse Cultivars. Bioscience Biotechnology and Biochemistry, 75, 1603-1605.

McCann, M.C., Carpita, N.C. 2005. Looking for invisible phenotypes in cell wall mutants of Arabidopsis thaliana. Plant Biosystems, 139, 80-83.

McFarlane, H.E., Döring, A., Persson, S. 2014. The Cell Biology of Cellulose Synthesis. Annual Review of Plant Biology, 65, 69-94.

Meakin, P. J., Roberts, J. A. 1990. Dehiscence of Fruit in Oilseed Rape (Brassica nap us L.) II. The role of cell wall degrading enzymes and ethylene. Journal of Experimental Botany, 41, 1003-1011.

Miller, G.L. 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal Chem, 3, 426-428.

Miyazaki, K., Takenouchi, M., Kondo, H., Noro, N., Suzuki, M., Tsuda, S. 2006. Thermal stabilization of Bacillus subtilis family-11 xylanase by directed evolution. Journal of Biological Chemistry, 281, 10236-10242.

Mohnen, D. 2008. Pectin structure and biosynthesis. Current Opinion in Plant Biology, 11, 266-277.

Mohnen, D.A., Biswal, A.K., Hao, Z., Hunt, K.D., Gelineo-Albersheim, I., Kataeva, I., Adams, M.W.W. 2013. Plants with altered cell wall biosynthesis and methods of use, Google Patents.

Mølhøj, M., Jørgensen, B., Ulvskov, P., Borkhardt, B. 2001. Two Arabidopsis thaliana genes,KOR2 and KOR3, which encode membrane-anchored endo-1,4-β-D-glucanases, are differentially expressed in developing leaf trichomes and their support cells, Plant Molecular Biology, 46, 263–275.

Mortimer, J.C., Laohavisit, A., Macpherson, N., Webb, A., Brownlee, C., Battey, N.H., Davies, J.M. 2008. Annexins: multifunctional components of growth and adaptation. Journal of Experimental Botany, 59, 533-544.

Mortimer, J.C., Miles, G.P., Brown, D.M., Zhang, Z.N., Segura, M.P., Weimar, T., Yu, X.L., Seffen, K.A., Stephens, E., Turner, S.R., Dupree, P. 2010. Absence of branches from xylan in Arabidopsis gux mutants reveals potential for simplification of lignocellulosic biomass. Proceedings of the National Academy of Sciences of the United States of America, 107, 17409-17414.

Mosier, N., Wyman, C., Dale, B., Elander, R., Lee, Y.Y., Holtzapple, M., Ladisch, M. 2005. Features of promising technologies for pretreatment of lignocellulosic biomass. Bioresource Technology, 96, 673-686.

Moxley, G., Gaspar, A.R., Higgins, D., Xu, H. 2012. Structural changes of corn stover lignin during acid pretreatment. Journal of Industrial Microbiology & Biotechnology, 39, 1289-1299.

Mullerrober, B., Sonnewald, U., Willmitzer, L. 1992. Inhibition of the Adp-Glucose Pyrophosphorylase in Transgenic Potatoes Leads to Sugar-Storing Tubers and Influences Tuber Formation and Expression of Tuber Storage Protein Genes. Embo Journal, 11, 1229-1238.

Nagaharu, U. 1935. Genome analysis in Brassica with special reference to the experimental formation of B. napus and peculiar mode of fertilization. Jpn J Bot, 7, 389-452.

Navarro, D., Couturier, M., da Silva, G., Berrin, J.-G., Rouau, X., Asther, M., Bignon, C. 2010. Automated assay for screening the enzymatic release of reducing sugars from micronized biomass. Microbial Cell Factories, 9, 1-12.

Nicol, Frédéric, Isabelle His, Alain Jauneau, Samantha Vernhettes, Hervé Canut, and Herman Höfte. 1998. A plasma membrane-bound putative endo-1, 4-□-d-glucanase is required for normal wall assembly and cell elongation in Arabidopsis. The EMBO Journal, 17, 5563-5576.

O'Malley, R.C., Ecker, J.R. 2010. Linking genotype to phenotype using the Arabidopsis unimutant collection. The Plant Journal, 61, 928-940.

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

OEDC. 2012. Inventory of Estimated Budgetary Support and Tax Expenditures for Fossil Fuels 2013. OEDC Publishing.

Oka, T., Nemoto, T., & Jigami, Y. 2007. Functional analysis of Arabidopsis thaliana RHM2/MUM4, a multidomain protein involved in UDP-D-glucose to UDP-L-rhamnose conversion. Journal of Biological Chemistry, 282, 5389-5403.

Olofsson, K., Bertilsson, M., Liden, G. 2008. A short review on SSF - an interesting process option for ethanol production from lignocellulosic feedstocks. Biotechnology for Biofuels, 1, 1-14.

Overend, R.P., Chornet, E. 1987. Fractionation of Lignocellulosics by Steam-Aqueous Pretreatments. Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences, 321, 523-536.

Palmqvist, E., Hahn-Hagerdal, B. 2000. Fermentation of lignocellulosic hydrolysates, II: inhibitors and mechanisms of inhibition. Bioresour Technol, 1, 25 - 33.

Pandey, A., Larroche, C., Ricke, S.C., Dussap, C.-G., Gnansounou, E. Eds. 2011. Biofuels: alternative feedstocks and conversion processes. Academic Press.

Park, Y.B., Cosgrove, D.J. 2012. A Revised Architecture of Primary Cell Walls Based on Biomechanical Changes Induced by Substrate-Specific Endoglucanases. Plant Physiology, 158, 1933-1943.

Parkin, I.A., Gulden, S.M., Sharpe, A.G., Lukens, L., Trick, M., Osborn, T.C., Lydiate, D.J. 2005. Segmental structure of the Brassica napus genome based on comparative analysis with Arabidopsis thaliana. Genetics, 171, 765-781.

Parkin, I.A., Sharpe, A., Lydiate, D. 2003. Patterns of genome duplication within the Brassica napus genome. Genome, 46, 291-303.

Pasam, R.K., Sharma, R., Malosetti, M., van Eeuwijk, F.A., Haseneyer, G., Kilian, B., Graner, A. 2012. Genome-wide association studies for agronomical traits in a world wide spring barley collection. BMC plant biology 12, 16.

Paterson, A.H., Lan, T.-h., Amasino, R., Osborn, T.C., Quiros, C. 2001. Brassica genomics: a complement to and early beneficiary of, the Arabidopsis sequence. Genome Biol, 2, 1339-1347.

Pauly, M., Gille, S., Liu, L.F., Mansoori, N., de Souza, A., Schultink, A., Xiong, G.Y. 2013. Hemicellulose biosynthesis. Planta, 238, 627-642.

Pauly, M., Keegstra, K. 2008. Cell-wall carbohydrates and their modification as a resource for biofuels. The Plant journal : for cell and molecular biology, 54, 559-68.

Pauly, M., Scheller, H.V. 2000. O-Acetylation of plant cell wall polysaccharides: identification and partial characterization of a rhamnogalacturonan O-acetyl-transferase from potato suspension-cultured cells. Planta, 210, 659-667.

Pawar, P.M.-A., Koutaniemi, S., Tenkanen, M., Mellerowicz, E.J. 2013. Acetylation of woody lignocellulose: significance and regulation. Frontiers in Plant Science, 4.

Pedersen, M., Meyer, A.S. 2010. Lignocellulose pretreatment severity - relating pH to biomatrix opening. New Biotechnology, 27, 739-750.

Peng, L.C., Hocart, C.H., Redmond, J.W., Williamson, R.E. 2000. Fractionation of carbohydrates in Arabidopsis root cell walls shows that three radial swelling loci are specifically involved in cellulose production. Planta, 211, 406-414.

Penning BW, Sykes RW, Babcock NC, Dugard CK, Held MA, Klimek JF, Shreve JT, Fowler M, Ziebell A, Davis MF, Decker SR, Turner GB, Mosier NS, Springer NM, Thimmapuram J, Weil CF, McCann MC, Carpita NC. 2014. Genetic Determinants for Enzymatic Digestion of Lignocellulosic Biomass Are Independent of Those for Lignin Abundance in a Maize Recombinant Inbred Population. Plant Physiol. 165:1475-1487. [Epub ahead of print]

Peplow, M. 2014. Cellulosic ethanol fights for life. Nature, 507, 152-153.

Perrin, D.D., Dempsey, B., Serjeant, E.P. 1981. pKa prediction for organic acids and bases. Chapman and Hall.

Petersen, M.Ø., Larsen, J., Thomsen, M.H. 2009. Optimization of hydrothermal pretreatment of wheat straw for production of bioethanol at low water consumption without addition of chemicals. Biomass and Bioenergy, 33, 834-840.

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

Petersson, A., Thomsen, M.H., Hauggaard-Nielsen, H., Thomsen, A.-B. 2007. Potential bioethanol and biogas production using lignocellulosic biomass from winter rye, oilseed rape and faba bean. Biomass and Bioenergy, 31, 812-819.

Peumans, W.J., Vandamme, E.J.M. 1995. Lectins as Plant Defense Proteins. Plant Physiology, 109, 347-352.

Phitsuwan, P., Sakka, K., Ratanakhanokchai, K. 2013. Improvement of lignocellulosic biomass in planta: A review of feedstocks, biomass recalcitrance and strategic manipulation of ideal plants designed for ethanol production and processability. Biomass & Bioenergy, 58, 390-405.

Pronyk, C., Mazza, G. 2012. Fractionation of triticale, wheat, barley, oats, canola and mustard straws for the production of carbohydrates and lignins. Bioresource Technology, 106, 117-124.

Pu, Y., Gao, J., Guo, Y., Liu, T., Zhu, L., Xu, P., Yi, B., Wen, J., Tu, J., Ma, C. 2013. A novel dominant glossy mutation causes suppression of wax biosynthesis pathway and deficiency of cuticular wax in Brassica napus. BMC plant biology, 13, 215.

Rafalski, J.A. 2010. Association genetics in crop improvement. Current Opinion in Plant Biology, 13, 174-180.

Rahikainen, J. L., Moilanen, U., Nurmi-Rantala, S., Lappas, A., Koivula, A., Viikari, L., & Kruus, K. 2013. Effect of temperature on lignin-derived inhibition studied with three structurally different cellobiohydrolases. Bioresource Technology, 146, 118-125.

Rasmussen, H., Sørensen, H.R., Meyer, A.S. 2013. Formation of degradation compounds from lignocellulosic biomass in the biorefinery: Sugar reaction mechanisms. Carbohydrate Research, 385, 45-57.

Rautengarten, C., Ebert, B., Herter, T., Petzold, C.J., Ishii, T., Mukhopadhyay, A., Usadel, B., Scheller, H.V. 2011. The Interconversion of UDP-Arabinopyranose and UDP-Arabinofuranose Is Indispensable for Plant Development in Arabidopsis. Plant Cell, 23, 1373-1390.

Rautengarten, C., Ebert, B., Moreno, I., Temple, H., Herter, T., Link, B., Doñas-Cofré, D., Moreno,A., Saéz-Aguayo, S., Blanco, F., Mortimer, J.C., Schultink, A., Reiter, W-D., Dupree, P., Pauly, M., Heazlewood, J.L., Scheller, H.V. Orellana, A. 2014. The Golgi localized bifunctional UDP-rhamnose/UDP-galactose transporter family of Arabidopsis. PNAS. [published ahead of print].

Rennie, E.A., Hansen, S.F., Baidoo, E.E.K., Hadi, M.Z., Keasling, J.D., Scheller, H.V. 2012. Three Members of the Arabidopsis Glycosyltransferase Family 8 Are Xylan Glucuronosyltransferases. Plant Physiology, 159, 1408-1417.

Rennie, E.A., Scheller, H.V. 2014. Xylan biosynthesis. Current Opinion in Biotechnology, 26, 100-107.

Rocha, G.J.M., Martin, C., da Silva, V.F.N., Gomez, E.O., Goncalves, A.R. 2012. Mass balance of pilot-scale pretreatment of sugarcane bagasse by steam explosion followed by alkaline delignification. Bioresource Technology, 111, 447-452.

Rodriguez-Gacio Mdel, C., Iglesias-Fernandez, R., Carbonero, P., Matilla, A.J. 2012. Softening-up mannan-rich cell walls. J Exp Bot, 63, 3976-88.

Rollin, J.A., Zhu, Z.G., Sathitsuksanoh, N., Zhang, Y.H.P. 2011. 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, 108, 22-30.

Rondeau-Mouro, C., Defer, D., Leboeuf, E., Lahaye, M. 2008. Assessment of cell wall porosity in Arabidopsis thaliana by NMR spectroscopy. International Journal of Biological Macromolecules, 42, 83-92.

Rose, J.K.C., Braam, J., Fry, S.C., Nishitani, K. 2002. The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: Current perspectives and a new unifying nomenclature. Plant and Cell Physiology, 43, 1421-1435.

Rossato, L., Laine, P., & Ourry, A. 2001. Nitrogen storage and remobilization in Brassica napus L. during the growth cycle: nitrogen fluxes within the plant and changes in soluble protein patterns. Journal of Experimental Botany, 52, 1655-1663.

Rosti, J., Barton, C.J., Albrecht, S., Dupree, P., Pauly, M., Findlay, K., Roberts, K., Seifert, G.J. 2007. UDP-glucose 4-epimerase isoforms UGE2 and UGE4 cooperate in providing UDP-galactose for cell wall biosynthesis and growth of Arabidopsis thaliana. Plant Cell, 19, 1565-1579.

Roudier F, Schindelman G, DeSalle R, Benfey PN. 2002. The COBRA family of putative GPI-anchored proteins in Arabidopsis. A new fellowship in expansion. Plant Physiology, 130, 538-48.

Ryden, P., Gautier, A., Wellner, N., Tapp, H.S., Horn, S.J., Eijsink, V.G.H., Waldron, K.W. 2014. Changes in the composition of the main polysaccharide groups of oil seed rape straw following steam explosion and saccharification. Biomass and Bioenergy, 61, 121-130.

Saedler, R., Zimmermann, I., Mutondo, M., Hulskamp, M. 2004. The Arabidopsis KLUNKER gene controls cell shape changes and encodes the AtSRA1 homolog. Plant Molecular Biology, 56, 775-782.

Saeman, J.F. 1945. Kinetics of wood saccharification-hydrolysis of cellulose and decomposition of sugars in dilute acid at high temperature. Industrial & Engineering Chemistry, 37, 43-52.

Salnikov, V.V., Grimson, M.J., Delmer, D.P., Haigler, C.H. 2001. Sucrose synthase localizes to cellulose synthesis sites in tracheary elements. Phytochemistry, 57, 823-833.

Sarkar, P., Bosneaga, E., Auer, M. 2009. Plant cell walls throughout evolution: towards a molecular understanding of their design principles. Journal of Experimental Botany, 60, 3615-3635.

Scheller, H.V., Ulvskov, P. 2010. Hemicelluloses. Annual Review of Plant Biology, 61, 263-289.

Seifert, G. J. 2004. Nucleotide sugar interconversions and cell wall biosynthesis: how to bring the inside to the outside. Current opinion in plant biology, 7, 277-284.

Seifert, G.J., Barber, C., Wells, B., Dolan, L., Roberts, K. 2002. Galactose biosynthesis in Arabidopsis: Genetic evidence for substrate channeling from UDP-D-galactose into cell wall polymers. Current Biology, 12, 1840-1845.

Seifert, G.J., Blaukopf, C. 2010. Irritable Walls: The Plant Extracellular Matrix and Signaling. Plant Physiology, 153, 467-478.

Seifert, G.J., Roberts, K. 2007. The biology of arabinogalactan proteins. Annual Review of Plant Biology, 58, 137-161.

Selig, M.J., Viamajala, S., Decker, S.R., Tucker, M.P., Himmel, M.E., Vinzant, T.B. 2007. Deposition of lignin droplets produced during dilute acid pretreatment of maize stems retards enzymatic hydrolysis of cellulose. Biotechnol Prog, 23, 1333-1339.

Shankar, M., Priyadharshini, R., Gunasekaran, P. 2009. Quantitative digital image analysis of chromogenic assays for high throughput screening of ⍺-amylase mutant libraries. Biotechnology Letters, 31, 1197-1201.

Sharples, S.C., Fry, S.C. 2007. Radioisotope ratios discriminate between competing pathways of cell wall polysaccharide and RNA biosynthesis in living plant cells. Plant Journal, 52, 252-262.

Shi, J., Ebrik, M.A., Yang, B., Garlock, R.J., Balan, V., Dale, B.E., Pallapolu, V.R., Lee, Y.Y., Kim, Y., Mosier, N.S., Ladisch, M.R., Holtzapple, M.T., Falls, M., Sierra-Ramirez, R., Donohoe, B.S., Vinzant, T.B., Elander, R.T., Hames, B., Thomas, S., Warner, R.E., Wyman, C.E. 2011. Application of cellulase and hemicellulase to pure xylan, pure cellulose and switchgrass solids from leading pretreatments. Bioresource Technology, 102, 11080-11088.

Showalter, A.M. 1993. Structure and function of plant cell wall proteins. The Plant Cell, 5, 9.

Sibout, R., Hofte, H. 2012. Plant Cell Biology: The ABC of Monolignol Transport. Current Biology, 22, 533-535.

Simmons, B.A., Loqué, D., Ralph, J. 2010. Advances in modifying lignin for enhanced biofuel production. Current Opinion in Plant Biology, 13, 312-319.

Slabaugh, E., Davis, J. K., Haigler, C. H., Yingling, Y. G., Zimmer, J. 2014. Cellulose synthases: new insights from crystallography and modeling. Trends in plant science, 19, 99-106.

Slavov, G., Allison, G., Bosch, M. 2013. Advances in the genetic dissection of plant cell walls: tools and resources available in Miscanthus. Front Plant Sci, 4, 217.

Sluiter, A. 2004. Determination of structural carbohydrates and lignin in biomass. National Renewable Energy Laboratory (NREL) Analytical Procedures.

Smith A.M., Kruger N.J., Lunn J.E. 2012. Source of sugar nucleotides for starch and cellulose synthesis. Proc Natl Acad Sci USA, 109.

Snowdon, R.J., Luy, F.L.I. 2012. Potential to improve oilseed rape and canola breeding in the genomics era. Plant Breeding, 131, 351-360.

Solecka, D., Zebrowski, J., Kacperska, A. 2008. Are pectins involved in cold acclimation and de-acclimation of winter oil-seed rape plants? Annals of Botany, 101, 521-530.

Somerville, C. 2006. Cellulose synthesis in higher plants. Annu Rev Cell Dev Biol, 22, 53-78.

Somerville, C., Bauer, S., Brininstool, G., Facette, M., Hamann, T., Milne, J. & Youngs, H. 2004. Toward a systems approach to understanding plant cell walls. Science, 306, 2206-2211.

Song, L.T., Laguerre, S., Dumon, C., Bozonnet, S., O'Donohue, M.J. 2010. A high throughput screening system for the evaluation of biomass-hydrolyzing glycoside hydrolases. Bioresource Technology, 101, 8237-8243.

Stamatiou, G., Vidaurre, D. P., Shim, I., Tang, X., Moeder, W., Bonetta, D., & McCourt, P. 2013. Forward Genetic Screening for the Improved Production of Fermentable Sugars from Plant Biomass. PloS one, 8.

Stoddart, H., Watts, J. 2012. Energy potential from UK arable agriculture: Straw – what is it good for?! 20th European Biomass Conference and Exhibition (EU BC&E), Milan, Italy.

Stokes D, Fraser F, Morgan C, O'Neill CM, Dreos R, Magusin A, Szalma S and Bancroft I (2010) An Association Transcriptomics approach to the prediction of hybrid performance. Molecular Breeding 26: 91-106.

Strasser, R., Schoberer, J., Jin, C.S., Glossl, J., Mach, L., Steinkellner, H. 2006. Molecular cloning and characterization of Arabidopsis thaliana Golgi alpha-mannosidase II, a key enzyme in the formation of complex N-glycans in plants. Plant Journal, 45, 789-803.

Strabala, T. J., & MacMillan, C. P. 2013. The Arabidopsis wood model—the case for the inflorescence stem. Plant Science, 210, 193-205.

Sturcova, A., His, I., Apperley, D.C., Sugiyama, J., Jarvis, M.C. 2004. Structural details of crystalline cellulose from higher plants. Biomacromolecules, 5, 1333-1339.

Sun, R. 2010. Cereal straw as a resource for sustainable biomaterials and biofuels: chemistry, extractives, lignins, hemicelluloses and cellulose. Elsevier.

Sutton, P.N., Henry, M.J., Hall, J.L. 1999. Glucose and not sucrose, is transported from wheat to wheat powdery mildew. Planta, 208, 426-430.

▪wi▪tek, K., Lewandowska, M., ▪wi▪tek, M., Bednarski, W., Brzozowski, B. 2014. The improvement of enzymatic hydrolysis efficiency of rape straw and Miscanthus giganteus polysaccharides. Bioresource Technology, 151, 323-331.

Szymanska-Chargot, M., Zdunek, A. 2013. Use of FT-IR Spectra and PCA to the Bulk Characterization of Cell Wall Residues of Fruits and Vegetables Along a Fraction Process. Food Biophysics, 8, 29-42.

TAIR, 2014. www.Arabidopsis.org/ portals/ genAnnotation/ gene_structural_ annotation/ agicomplete.jsp.

Tan, L., Eberhard, S., Pattathil, S., Warder, C., Glushka, J., Yuan, C.H., Hao, Z.Y., Zhu, X., Avci, U., Miller, J.S., Baldwin, D., Pham, C., Orlando, R., Darvill, A., Hahn, M.G., Kieliszewski, M.J., Mohnen, D. 2013. An Arabidopsis Cell Wall Proteoglycan Consists of Pectin and Arabinoxylan Covalently Linked to an Arabinogalactan Protein. Plant Cell, 25, 270-287.

Tang, H., Lyons, E. 2012. Unleashing the genome of Brassica rapa. Frontiers in plant science, 3.

Tardieu, F., Parent, B., Caldeira, C.F., Welcker, C. 2014. Genetic and Physiological Controls of Growth under Water Deficit. Plant Physiology, 164, 1628-1635.

The Brassica rapa Genome Sequencing Project Consortium. 2011. The genome of the mesopolyploid crop species Brassica rapa. Nature genetics, 43, 1035-1039.

Thygesen, A., Vahlgren, L., Frederiksen, J.H., Linnane, W., Thomsen, M.H. 2012. SSF Fermentation of Rape Straw and the Effects of Inhibitory Stress on Yeast Bioethanol, (Ed.) M.A.P. Lima.

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. 2014. Brachypodium Cell Wall Mutant with Enhanced Saccharification Potential Despite Increased Lignin Content. BioEnergy Research, 1-15.

Tofanica, B.M., Cappelletto, E., Gavrilescu, D., Mueller, K. 2011. Properties of Rapeseed (Brassica napus) Stalks Fibers. Journal of Natural Fibers, 8, 241-262.

Torres, A.F., Visser, R.G.F., Trindade, L.M. 2013. Bioethanol from maize cell walls: genes, molecular tools and breeding prospects. GCB Bioenergy.

Turlapati, P.V., Kim, K.W., Davin, L.B., Lewis, N.G. 2011. The laccase multigene family in Arabidopsis thaliana: towards addressing the mystery of their gene function(s). Planta, 233, 439-470.

Underwood, W. 2012. The plant cell wall: a dynamic barrier against pathogen invasion. Frontiers in Plant Science, 3.

United Nations World Population Division. 2013. World Population Prospects: The 2012 Revision, Vol. II.

Valerio, L., De Meyer, M., Penel, C., Dunand, C. 2004. Expression analysis of the Arabidopsis peroxidase multigenic family. Phytochemistry, 65, 1331-1342.

Van Acker, R., Vanholme, R., Storme, V., Mortimer, J., Dupree, P., Boerjan, W. 2013. Lignin biosynthesis perturbations affect secondary cell wall composition and saccharification yield in Arabidopsis thaliana. Biotechnology for Biofuels, 6, 46.

Vandenborre, G., Smagghe, G., Van Damme, E.J.M. 2011. Plant lectins as defense proteins against phytophagous insects. Phytochemistry, 72, 1538-1550.

Várnai, A., Siika-aho, M., & Viikari, L. 2013. Carbohydrate-binding modules (CBMs) revisited: reduced amount of water counterbalances the need for CBMs. Biotechnology for Biofuels, 6, 1-12.

Vidal, B., Dien, B., Ting, K., Singh, V. 2011. Influence of Feedstock Particle Size on Lignocellulose Conversion—A Review. Applied Biochemistry and Biotechnology, 164, 1405-1421.

Vincken, J. P., Schols, H. A., Oomen, R. J., McCann, M. C., Ulvskov, P., Voragen, A. G., & Visser, R. G. 2003. If homogalacturonan were a side chain of rhamnogalacturonan I. Implications for cell wall architecture. Plant Physiology, 132, 1781-1789.

Virk, D.S., Pandit, D.B., Sufian, M. A., Ahmed, F., Siddique, M. A. B., Samad M. A., Rahman, M. M., Islam, M. M., Ortiz-ferrara, G., Joshi K. D. and Witcombe J. R. 2009. REML is an effective analysis for mixed modelling of unballanced on-farm varietal trials. Experimental Agriculture, 45, 77-91.

Vivekanand, V., Ryden, P., Horn, S.J., Tapp, H.S., Wellner, N., Eijsink, V.G.H., Waldron, K.W. 2012. Impact of steam explosion on biogas production from rape straw in relation to changes in chemical composition. Bioresource Technology, 123, 608-615.

Vogel, J.P., Raab, T.K., Somerville, C.R., Somerville, S.C. 2004. Mutations in PMR5 result in powdery mildew resistance and altered cell wall composition. Plant Journal, 40, 968-978.

Waldron, K.W. 2010. Bioalcohol Production: Biochemical Conversion of Lignocellulosic Biomass Woodhead Publishing Ltd.

Wang, S., Yin, Y., Ma, Q., Tang, X., Hao, D., Xu, Y. 2012. Genome-scale identification of cell-wall related genes in Arabidopsis based on co-expression network analysis. BMC Plant Biology, 12, 138.

Wang, Y., Chantreau, M., Sibout, R., Hawkins, S. 2013a. Plant cell wall lignification and monolignol metabolism. Frontiers in Plant Science, 4.

Wang, Y., Mortimer, J.C., Davis, J., Dupree, P., Keegstra, K. 2013b. Identification of an additional protein involved in mannan biosynthesis. The Plant Journal, 73, 105-117.

Wang, Y., Zhang, W-Z., Song, L-F., Zou, J-J., Su, Z. and Wu, W-H. 2008. Transcriptome Analyses Show Changes in Gene Expression to Accompany Pollen Germination and Tube Growth in Arabidopsis. American Society of Plant Biologists, 148, 1201-1211.

Wang, Y.H. 2008. How effective is T-DNA insertional mutagenesis in Arabidopsis? Journal of Biochemical Technology, 1, 11-20.

Wang, Y.H., Acharya, A., Burrell, A.M., Klein, R.R., Klein, P.E., Hasenstein, K.H. 2013c. Mapping and candidate genes associated with saccharification yield in sorghum. Genome, 56, 659-665.

Wang, Y.H., Poudel, D.D., Hasenstein, K.H. 2011. Identification of SSR markers associated with saccharification yield using pool-based genome-wide association mapping in sorghum. Genome, 54, 883-889.

Weigel, D., Mott, R. 2009. The 1001 genomes project for Arabidopsis thaliana. Genome Biol, 10, 107.

Weise, A., Barker, L., Kühn, C., Lalonde, S., Buschmann, H., Frommer, W.B., Ward, J.M., 2000. A new subfamily of sucrose transporters, SUT4, with low affinity/high capacity localized in enucleate sieve elements of plants. Plant Cell, 12, 1345-55

Wen, J.Q., Lease, K.A., Walker, J.C. 2004. DVL, a novel class of small polypeptides: overexpression alters Arabidopsis development. Plant Journal, 37, 668-677.

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

Wi, S.G., Chung, B.Y., Lee, Y.G., Yang, D.J., Bae, H.-J. 2011. Enhanced enzymatic hydrolysis of rapeseed straw by popping pretreatment for bioethanol production. Bioresource Technology, 102, 5788-5793

Wi, S.G., Singh, A.P., Lee, K.H. and Kim, Y.S. 2005. The Pattern of Distribution of Pectin, Peroxidase and Lignin in the Middle Lamella of Secondary Xylem Fibres in Alfalfa (Medicago sativa) Annals of Botany, 95, 863–868

Wiedemeier, Allison MD, Jan E. Judy-March, Charles H. Hocart, Geoffrey O. Wasteneys, Richard E. Williamson, and Tobias I. Baskin. 2002. Mutant alleles of Arabidopsis RADIALLY SWOLLEN 4 and 7 reduce growth anisotropy without altering the transverse orientation of cortical microtubules or cellulose microfibrils. Development 129, 4821-4830.

Wightman, R., Turner, S.R. 2008. The roles of the cytoskeleton during cellulose deposition at the secondary cell wall. Plant Journal, 54, 794-805.

Wu, Z., Hao, H., Tu, Y., Hu, Z., Wei, F., Liu, Y & Wang, L. 2014. Diverse cell wall composition and varied biomass digestibility in wheat straw for bioenergy feedstock. Biomass and Bioenergy. In press.

Wyman, C.E. 2007. What is (and is not) vital to advancing cellulosic ethanol. Trends in Biotechnology, 25, 153-157.

Xiao, C. Anderson, C.T. 2013. Roles of pectin in biomass yield and processing for biofuels. Frontiers in Plant Science, 4.

Xin, Z.G., Mandaokar, A., Chen, J.P., Last, R.L., Browse, J. 2007. Arabidopsis ESK1 encodes a novel regulator of freezing tolerance. Plant Journal, 49, 786-799.

Xiong, G.Y., Cheng, K., Pauly, M. 2013. Xylan O-Acetylation Impacts Xylem Development and Enzymatic Recalcitrance as Indicated by the Arabidopsis Mutant tbl29. Molecular Plant, 6, 1373-1375.

Xu, C., Qin, Y., Li, Y., Ji, Y., Huang, J., Song, H., Xu, J. 2010. Factors influencing cellulosome activity in Consolidated Bioprocessing of cellulosic ethanol. Bioresource Technology, 101, 9560-9569.

Xu, Z.W., Escamilla-Trevino, L.L., Zeng, L.H., Lalgondar, M., Bevan, D.R., Winkel, B.S.J., Mohamed, A., Cheng, C.L., Shih, M.C., Poulton, J.E., Esen, A. 2004. Functional genomic analysis of Arabidopsis thaliana glycoside hydrolase family 1. Plant Molecular Biology, 55, 343-367.

Yang, F., Mitra, P., Zhang, L., Prak, L., Verhertbruggen, Y., Kim, J.S., Sun, L., Zheng, K.J., Tang, K.X., Auer, M., Scheller, H.V., Loque, D. 2013. Engineering secondary cell wall deposition in plants. Plant Biotechnology Journal, 11, 325-335.

Yang, T., Bar-Peled, L., Gebhart, L., Lee, S.G., Bar-Peled, M. 2009. Identification of Galacturonic Acid-1-phosphate Kinase, a New Member of the GHMP Kinase Superfamily in Plants and Comparison with Galactose-1-phosphate Kinase. Journal of Biological Chemistry, 284, 21526-21535.

Yang, Y.W., Lai, K.N., Tai, P.Y., Li, W.H. 1999. Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between Brassica and other angiosperm lineages. Journal of Molecular Evolution, 48, 597-604.

Yapo, B.M. 2011. Pectic substances: From simple pectic polysaccharides to complex pectins—A new hypothetical model. Carbohydrate Polymers, 86, 373-385.

Yokoyama, R., Nishitani, K. 2004. Genomic basis for cell-wall diversity in plants. A comparative approach to gene families in rice and Arabidopsis. Plant and Cell Physiology, 45, 1111-1121.

York, W.S., O'Neill, M.A. 2008. Biochemical control of xylan biosynthesis - which end is up? Current Opinion in Plant Biology, 11, 258-265.

Yu, B., Chen, H.Z. 2010. Effect of the ash on enzymatic hydrolysis of steam-exploded rice straw. Bioresource Technology, 101, 9114-9119.

Yu, J.Y., Zhao, M.X., Wang, X.W., Tong, C.B., Huang, S.M., Tehrim, S., Liu, Y.M., Hua, W., Liu, S.Y. 2013. Bolbase: a comprehensive genomics database for Brassica oleracea. Bmc Genomics, 14.

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

Zabaniotou, A., Ioannidou, O., Skoulou, V. 2008. Rapeseed residues utilization for energy and 2nd generation biofuels. Fuel, 87, 1492-1502.

Zablackis, E., Huang, J., Muller, B., Darvill, A.G., Albersheim, P. 1995. Characterization of the cell-wall polysaccharides of Arabidopsis thaliana leaves. Plant Physiol, 107, 1129-38.

Zabotina, O.A. 2012. Xyloglucan and its biosynthesis. Frontiers in Plant Science, 3.

Zacchi, G., Axelsson, A. 1989. Economic evaluation of preconcentration in production of ethanol from dilute sugar solutions. Biotechnology and Bioengineering, 34, 223-233.

Zeng, M.J., Mosier, N.S., Huang, C.P., Sherman, D.M., Ladisch, M.R. 2007. Microscopic examination of changes of plant cell structure in corn stover due to hot water pretreatment and enzymatic hydrolysis. Biotechnology and Bioengineering, 97, 265-278.

Zhang, Y-H.P. 2008. Reviving the carbohydrate economy via multi-product lignocellulose biorefineries. Journal of Industrial Microbiology & Biotechnology 35, 367-375.

Zhang, J., Pakarinen, A., & Viikari, L. 2013. Synergy between cellulases and pectinases in the hydrolysis of hemp. Bioresource Technology, 129, 302-307.

Zhang, H., Fangel, J.U., Willats, W.G., Selig, M.J., Lindedam, J., Jørgensen, H., Felby, C. 2014. Assessment of leaf/stem ratio in wheat straw feedstock and impact on enzymatic conversion. GCB Bioenergy, 6, 90-96.

Zhao, X., Zhang, L., Liu, D. 2012. Biomass recalcitrance. Part I: the chemical compositions and physical structures affecting the enzymatic hydrolysis of lignocellulose. Biofuels, Bioproducts and Biorefining, 6, 465-482.

Zhou, C., Yin, Y., Dam, P., Xu, Y. 2010. Identification of Novel Proteins Involved in Plant Cell-Wall Synthesis Based on Protein−Protein Interaction Data. Journal of Proteome Research, 9, 5025-5037.

Zhu, C.S., Gore, M., Buckler, E.S., Yu, J.M. 2008. Status and Prospects of Association Mapping in Plants. Plant Genome, 1, 5-20.

Zhu, J. Y., Verrill, S. P., Liu, H., Herian, V. L., Pan, X., Rockwood, D. L. 2011. On polydispersity of plant biomass recalcitrance and its effects on pretreatment optimization for sugar production. Bioenergy Research, 4, 201-210.

Ziolkowski, P.A., Kaczmarek, M., Babula, D., Sadowski, J. 2006. Genome evolution in Arabidopsis/Brassica: conservation and divergence of ancient rearranged segments and their breakpoints. The Plant Journal, 47, 63-74.