Using association mapping to identify Quantitative Trait Loci (QTL) for straw digestibility in rice straw

Duong T. Nguyen

Master by research

University of York

Department of Biology

June 2016

Abstract

There is an urgent need to reduce our dependency on fossil fuels, and to reduce greenhouse gas (GHG) emissions. In recent years, a conversion of lignocellulosic biomass from agricultural wastes to biofuel has been considered a promising solution to provide sustainable liquid fuels without competing with food production. At present, this is limited by the cost effectiveness of the production process, due to the recalcitrance of lignocellulosic biomass to digestion. One way to tackle this problem is to improve the digestibility of the biomass feedstock itself, which would help to decrease chemical and enzyme treatments necessary for converting biomass to fermentable sugars for bioethanol production. With this objective in mind, the aim of this study is to identify genes that affect the digestibility of rice straw. Rice (*Oryza sativa*), the third biggest cereal crop in the world, has a small diploid genome, and well-developed molecular genetics tools to underpin the use of Genome Wide Association Studies (GWAS), a powerful tool for identifying quantitative trait loci. The genotypic data of 151 varieties, containing 328,915 Single Nucleotide Polymorphism (SNPs), was analysed together with the phenotypic data from three different cell wall characterisation assays: (1) saccharification, (2) total lignin content, and (3) silica measurement by X-ray fluorescence. This resulted in the identification of 10 significant QTL for saccharification potential, 3 significant QTL for Silica content, and 8 significant QLTs for total Lignin. The digestibility QTL regions identified include three strong potential candidate genes which determine on digestibility:*OsAT10* (Bartley, et al., 2013); *OsIRX9* (Chiniquy, et al., 2013); and *OsMYB58/63-L* (Noda, et al., 2015). A strong correlation between silicon content and digestibility was also identified. A QTL in chromosome 6 occurred in both GWAS for digestibility and silica, and revealed a contrasting bi-allelic effect between the amounts of sugar released and silica content. The *OsAT10* gene, located in this common QTL region was validated using the straw from overexpression lines, showing that OsAT10 had large effect on both silica content and digestibility.

Table of Contents

List of Figures

List of tables

Acknowledgements

Firstly I would like to sincerely thank my supervisor, Simon McQueen-Mason, who has provided me with brilliant and enthusiastic guidance throughout my MSc. I would also like to thank my TAP committee members, Richard Waites and Frans Maathuis for their extremely helpful guidance and ideas. I also want to send many thanks to the Ministry of Agricultural and Rural Development (MARD) and Ministry of Education and Training (MOET) in Vietnam for funding my MSc, and to the University of York for hosting my research.

I am really grateful all the research materials (rice seeds and rice straws) in the rice population that I received from my colleagues in Vietnam. The ones I would like to specially mention here are Dr. Duong Xuan Tu, head of Plant Biotechnology Department, and Ms. Nguyen Thi Huong, the researcher who has been growing and keeping the population in the Field Crops Research Institute (FCRI). Great thanks to Professor Bartley and her colleagues who provided the rice straw of OsAT10 overexpression lines for confirmation studies of its effect on silica content and digestibility.

Huge thanks to Dr. Leo D. Gomez who is not only such a nice co-supervisor with all idea and advice for my MSc but also the person who cares about my social activity while studying here in UK. Along with that, I cannot help but mention to kind helps from Rachel Hallam, senior technician of CNAP, who has been helping and instructing me through all the hard work of cell wall characterisation. I am deeply grateful to Dr. Andrea Harper for her help and instruction with the Genome Wide Association Studies, which is the key of the success of my research.

I would like to thank all members of the McQueen-Mason group for always being so helpful and friendly in and out of the lab. That is really the most important thing giving me more power to keep my MSc going on smoothly. To be studying overseas, my family at home always show their support, keep in frequent contact and cheer me up to get to the end of my MSc. To conclude about my thankfulness, I have to say that without help, advice, and support from everyone I mentioned above, I couldn't have got this done.

Declaration

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

Chapter 1 -Introduction

1.1 Biofuels

Humanity faces a dilemma: How to escape dependency on the fossil fuels and reduce the greenhouse gases (GHG) emissions without further exacerbating the environmental impact of agriculture. This issue faces us all, not only because GHG emissions that are produced locally have global impacts, but also because we now operate in a globalised economy, where commodities such as biofuels are traded on the world stage. Policy-imposed quantifiable sustainability requirements applied to biofuels will help to ensure that they are fit for purpose in environmental terms. However, such measures alone are useless without the appropriate technological developments required to enable sustainable biofuels to be produced in an economically competitive manner.

1.1.1 Relationship between fossil fuels and biofuels

A biofuel is the product of recent biological processes, such as agriculture and anaerobic digestion, which uses biological materials as feedstock. Currently, bioethanol and biodiesel are the two most common types of biofuel for transport, mainly produced as substitutes for gasoline and diesel fuel respectively. According to the biofuel definition from greenfacts.org, "biofuels are fuels produced directly or indirectly from organic material – biomass – including plant materials and animal waste". In that manner of speaking, fossil fuels somehow could be considered as the source of ancient biofuel, as defined by the geologists, "fossil fuels – coal, petroleum oil, and natural gas – are concentrated organic compounds found in the Earth's crust and formed by natural processes such as anaerobic decomposition of prehistoric biological matter for several million years"(Mann, et al., 2003). 'Bio-refinery' is another term often mentioned associated with biofuel. This term was initially established by NREL in the1990s: "A bio-refinery is a facility that integrates conversion processes and equipment to produce fuels, power, bio-product, and chemicals from biomass" (NREL, 2005).

It is undisputed that fossil fuels make modern life possible – they have been used as sources of energy to generate steam and electricity, and to power transportation systems. To credit the importance of fossil fuel to mankind, the US Energy Information

Administration (EIA) emphasizes that it has brought "one of the most profound social transformation in human history". However, we have to accept the reality that fossil fuels are not everlasting (see Figure 1). According to fossil fuel reserve and consumption data from CIA World Fact-book (Figure 3), our known oil deposits will be gone by 2052, if we increase gas and coal production to fill the energy gap left by oil, then those reserves will only take us as far as 2088. Arguably, the greatest downside of fossil fuel which has been widely realized is that the intensive use of fossil fuels has been resulting high levels of atmospheric pollution, releasing CO_2 and irreversibly modifying the climate on the planet (Figure 2).

Due to the limitations of using fossil fuel mentioned above, biofuels are being considered as the alternative source of liquid transportation fuels, and will need to be produced and used on a wide scale. Using biofuels as substitution for fossil fuels has the potential to mitigate emission of greenhouse gases. Although burning biofuel releases CO_2there is a net balance in this as an equivalent amount is "fixed" by the plants through photosynthesis. In a recent study, the carbon footprint of first generation biofuels and second generation biofuels was reported to reduce the greenhouse gas effect by 78% and 94%, respectively, when compared to the greenhouse gas effect caused by fossil fuels(Highina & Bugaje, 2014).

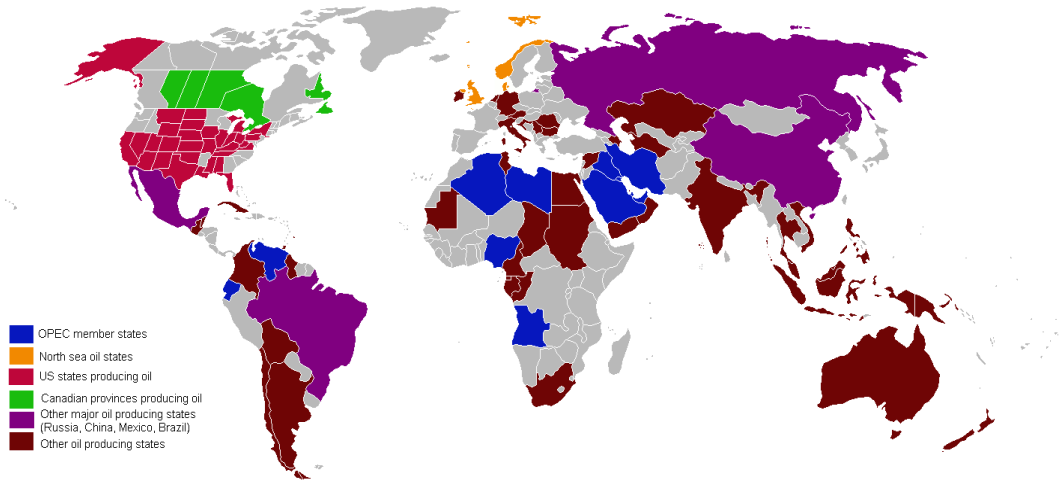


Figure 1: Oil producing countries map. Since oil fields are located only at certain places on earth **(Rosss.W, 2013)**, only a select group of countries are oil-independent; the other countries depend on the oil-production capacities of these countries**.**

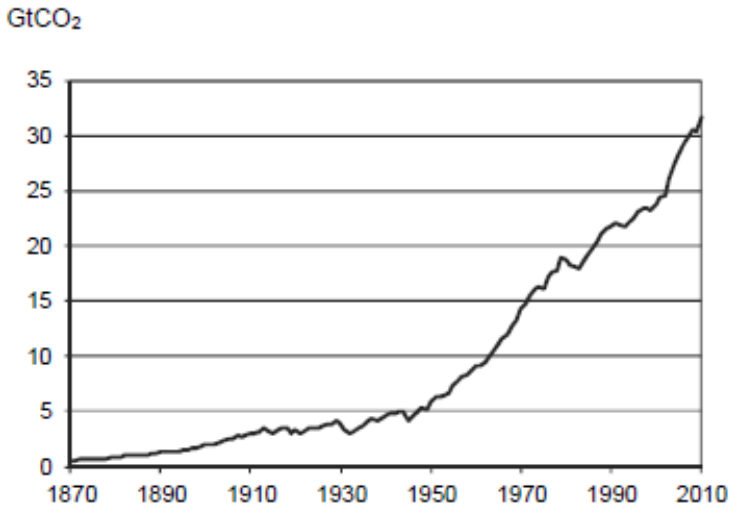


Figure 2: Trend in CO2 emissions from fossil fuel combustion. Source: Carbon Dioxide Information Analysis Centre, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tenn., United States. *Key point: Since 1870, CO2 emissions from fuel combustion have risen exponentially.* Data is taken from the International Energy Agency.

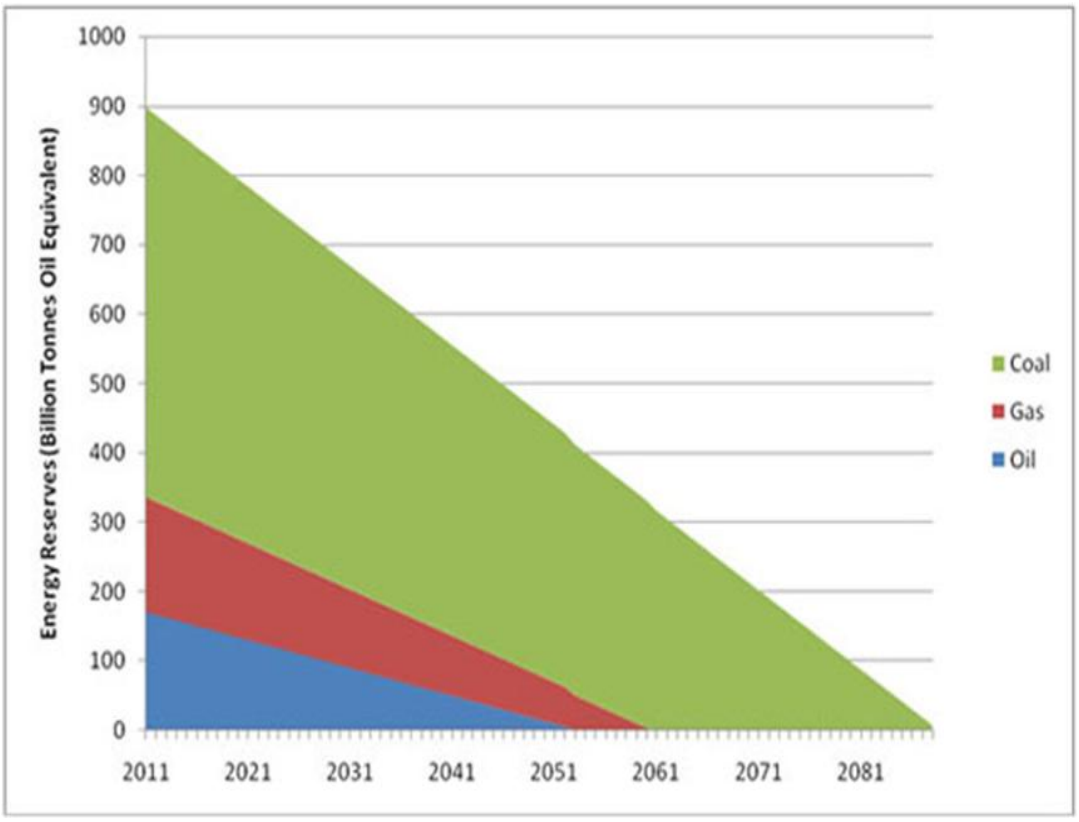


Figure 3: Predicted fossil fuel reserves, taken from CIA World Fact-book

1.1.2 First generation biofuels

First generation biofuels use food crops or animal fats as the main feedstock for production to extract directly starch, sugar, and oil for fermentation, esterification, and distillation process to produce bioethanol and biodiesel (Carroll & Somerville, 2009). The most commonly used crops for first generation of biofuels are corn, wheat, and sugarcane. It has been reported that sugarcane contributes 60% of total global bioethanol production (Demirbas, 2009). In Europe, there are only three crops being used to produce bioethanol commercially, which are wheat (50%), sugar beet (30%), and barley (20%). Remarkably in 2005, based on using these thee feedstocks, EU bioethanol production exceeded 910 million litres: an increase of 73 % on the previous year(STS, 2012). The countries which are considered as the main centres of EU bioethanol production are Spain, Germany, Sweden and France. In 2013, Brazil and the USA accounted for 84% of the bioethanol produced worldwide (Figure 4). In the USA, corn is still the major source of feedstock, used for 90 percent of the ethanol produced there(Goettemoeller & Goettemoeller, 2007). On the other hand, Brazil, the second largest ethanol producer in the world, chiefly uses sugarcane for biofuels production(RFA, 2014). Looking at the cost aspect, although Brazil is not the largest bioethanol producer, the technology developed and improved along the years in Brazil, makes the Brazilian sugarcane bioethanol the cheapest in the world (Govinda R. & Shrestha, 2011)

The contribution and influence of first generation biofuels to the world energy are worth recognizing, and it can satisfy our short term requirements to deal with fuel security. However there are clear limitations as to how much biofuel can be produced, moreover, very large scale production will threaten global food security and biodiversity (Evans, 2008). This is because we are using land and materials that would normally be used for food production and such practice adds pressure to global demand for food commodities and agricultural land. According to a scientific report published recently, biofuels production consume about 2-3% of the global water and land used for agriculture, which could feed about 30% of the malnourished population(Rulli, et al., 2016). This tension between food and fuel security together with the need to reduce greenhouse gases emissions, require us to consider more sustainable ways of producing biofuels

(Gomez, et al., 2008). The use of lignocellulosic plant biomass as a feedstock for biofuel production provides a logical way forward (Carroll & Somerville, 2009).

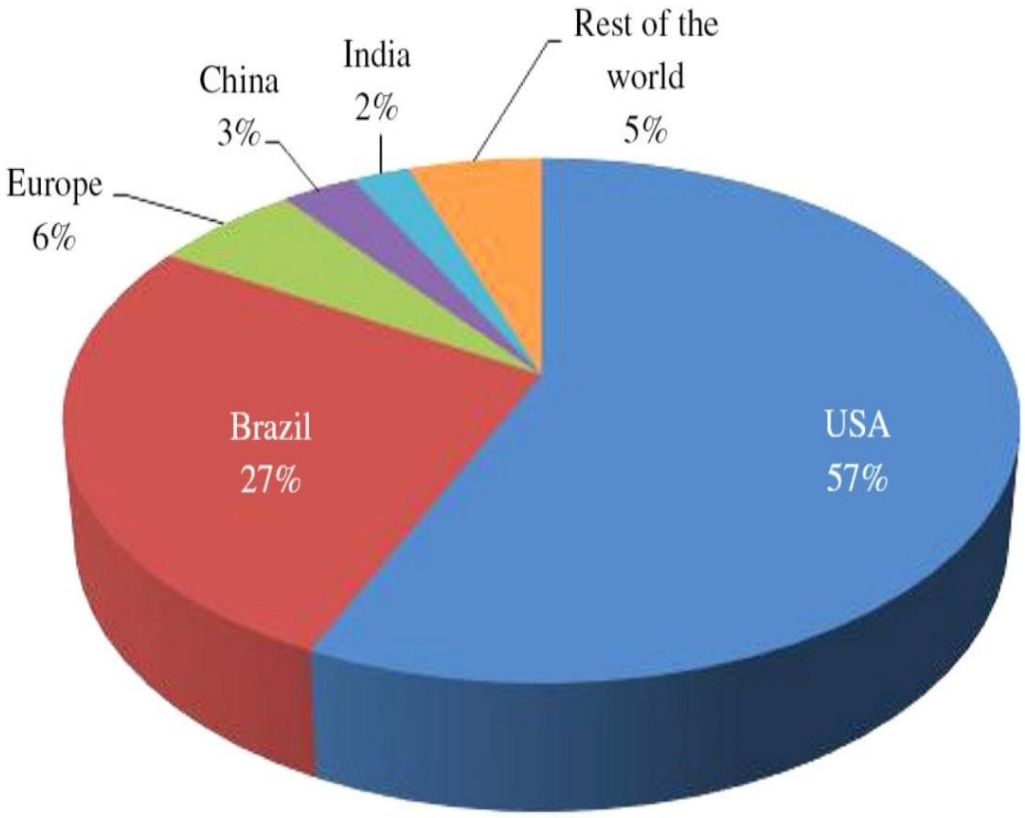


Figure 4: Production of bioethanol per country or region in 2013. Data taken from Renewable Fuel Association **(RFA, 2013)**

1.1.3 Second generation biofuels

In the past few years, academia, industrial companies and governments have shifted their interest to alternative non-food sources of feedstock which have high potential for conversion to biofuels. Biofuels produced from non-food feedstock are referred to as "second generation" biofuels. Overcoming the ethical concerns around the first generation biofuel, second generation biofuels are being intensively researched and developed to be potential and efficient replacements for fossil fuel in the near future. Second generation biofuels can also help to solve almost all the disadvantages mentioned above with sustainable, affordable larger proportion of global fuel supply without threatening food security, and with greater environmental benefits based on its contribution to significant reduction in GHG emissions. Recently, the United States

Department of Energy's Centre for Transportation Research found that cellulosic ethanol has highest reduction of GHG emissions compared to corns-derived bioethanol as illustrated in Figure 5.

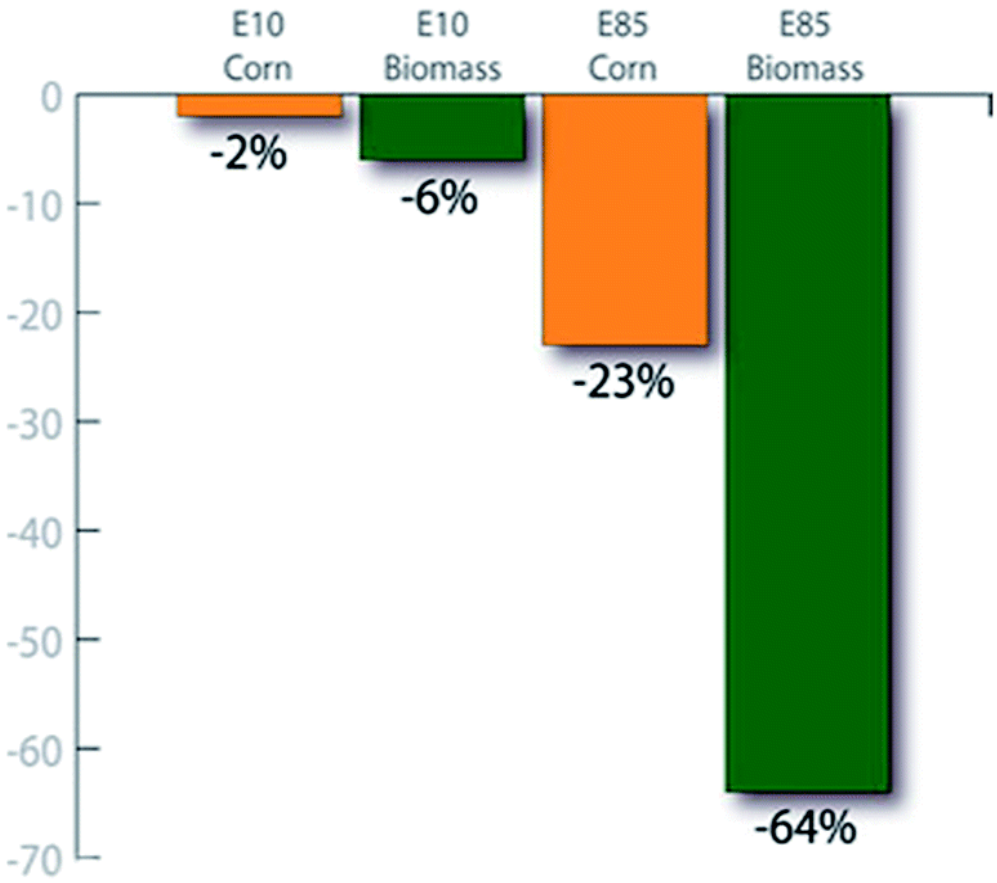


Figure 5: Reduction of greenhouse gas (GHG) emissions from cellulosic bioethanol and corns-derived bioethanol blends **(I. Ceres, 2015)**

Second generation biofuel can be produced from various types of lignocellulosic biomass. A major feedstock could be the cheap and abundant residue available after harvesting the food parts from plants and crops. This lignocellulosic structured-plant biomass also represents one of the most abundant and underutilized biological resource on the planet (Gomez, et al., 2008). Alternatively, some people propose the use of dedicated biomass crops that require low inputs for growth on land unsuitable for growing food crops. However, the main technical barriers to making liquid fuels from lignocellulosic biomass is the high cost of producing sugars for fermentation, which needs to be overcome before their conversion technologies can yield a product that is cost-effective (Neil, 2006).

Table 1showing comparison advantages and disadvantages of the 1st generation biofuels and obvious advantages of 2ndgeneration biofuels. The utilization of biomass for sustainable development are more integral , where all parts of the plant such as leaves, bark, fruits, and seeds can be utilized to useful products (Naik, et al., 2010).

Table 1 Comparison of first, second generation biofuel and petroleum fuel **(Naik, et al., 2010)**

	Petroleum fuels	**First generation biofuel**	**Second generation biofuel**
Feedstocks	Crude petroleum	Sugar cane, corn, milo, wheat. Barley and rice grain, potato, sweet potatoes.	Agricultural waste, food waste, aquatic biomass.
Products	CNG, LPG, diesel, kerosene, petrol, jet fuel.	FAME, ethanol, butanol, etc.	Lignocelulosic ethanol, butanol, FT oil, bio-oil, hydro treating oil
Problems	Depletion of fossil fuel reverses, Environmental pollutions, Economic and ecological problems	Limited sources	Lengthy production process, High cost
Benefits	Low cost of processing	Environmentally friendly, economic and social security	Not competing with food sources, cheap and abundant sources, Environmental friendly,

1.2 Biomass

Biomass refers to the biological substance accumulated in living organism, or recently living organisms. In the context of biofuel production, waste biomass from non-food and non-feed material is often referred to as lignocellulosic biomass (BEC, 2012). Yearly, there are approximately 100 billion tonnes organic dry matter of land biomass, plus 50 billion tonnes aquatic biomass available on earth (Groombridge & Jenkins, 2002). In total of the 150 billion tonnes biomass, there is only 1.25% used partly as feed, industrial raw materials, energy production, and the rest of the biomass is unused or recycled into the soil. Asia is reported as the largest potential producer of biofuel from crop residues and waste crops due to higher biomass availability; Table 2shows the biomass feedstocks and their potential ethanol yields(AFDC, 2015)(Naik, et al., 2010)

Table 2 Biomass feedstocks and their potential ethanol yield **(AFDC, 2015)(Naik, et al., 2010)**

Feedstock	**Potential ethanol yield, (litre per dry tonne of feedstock)**
Corn grain	470
Corn stover	428
Rice straw	416
Cotton gin trash	215
Forest thinnings	309
Hardwood sawdust	382
Bagasse	437

Feedstock	Potential ethanol yield, (litre per dry tonne of feedstock)
Mixed paper	440
Switchgrass[a]	366
***a* Switchgrass alamo whole plant. Source: U.S. Department of Energy Biomass Program, theoretical ethanol yield calculator and biomass feedstock composition and property database.**	

1.2.1 Lignocellulose in plants

Lignocellulose is the largest component of "plant biomass", and this makes up the majority of the cheap and abundant non-food materials available from plants (Gomez, et al., 2008). The potential of lignocellulosic materials has been realized for a while and considered as a renewable source of sugars that could be used to produce biofuel or other final products. However, the sugar polymers that make up plant cell walls are highly recalcitrant, and according to US department of Energy (DOE) – 2015, "biomass recalcitrance still remains the most important factor impeding the development of low-cost biomass processing technology". Hence, before being able to take the full advantage of these resources we have to find a better way to reduce the high cost of lignocellulose conversion, which is crucially dependant and enzymatic deconstruction efficiency of the cell walls (Himmel, et al., 2007). Reducing the recalcitrance of biomass is the basic scientific task that needs to be worked on to improve the rate of enzymatic hydrolysis and fermentation process, and also increase the yield and concentration of fermentable sugars in fermentation medium(US.DOE, 2015).

The modification of biomass recalcitrance requires of understanding the chemical and physical structure of plant cell walls, how they are synthesized, and how they can be deconstructed (Figure 6). Lignocellulosic material can generally be divided into three main components: cellulose (40-50%), hemicellulose (25-30%) and lignin (15-25%) (US.DOE, 2015). In general, cellulose and hemicellulose accounts for up to 70%-80% of total biomass in plants, these two dominant components are strongly associated to

lignin. The hydrophobic nature of lignin makes it highly resistant to physical and biochemical treatments, and hard to be broken down (Edye & Doherty, 2015). There has been a lot of work on understanding the composition and structure of lignocellulosic cell walls, which is informing work on how to improve digestibility (Figure 7)

Efforts to alter the cell wall components based on reverse genetic and molecular technique have resulted in reducing cell wall recalcitrance, and thus improving the digestibility and/or saccharification of lignocellulosic biomass. (Marriott, et al., 2015).

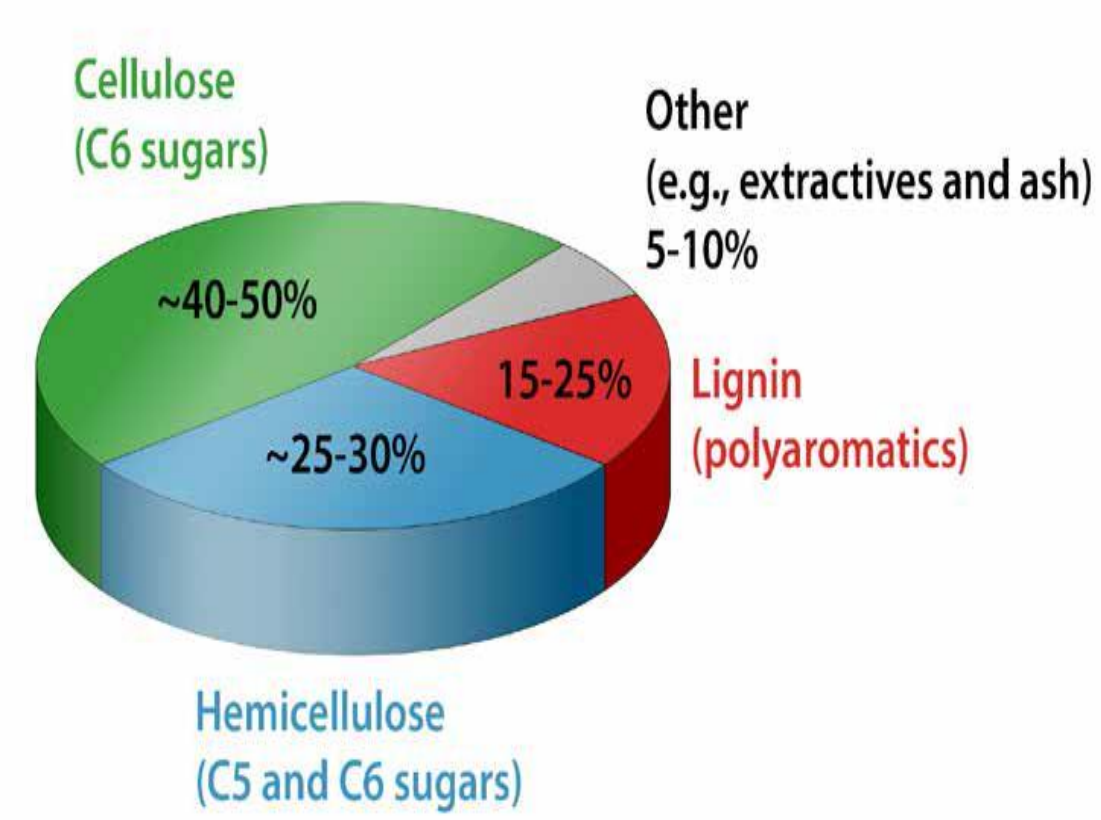


Figure 6: Pie chart representing the approximate distribution of the three primary components of plant cell walls—cellulose, hemicellulose, and lignin**(US.DOE, 2015)**.

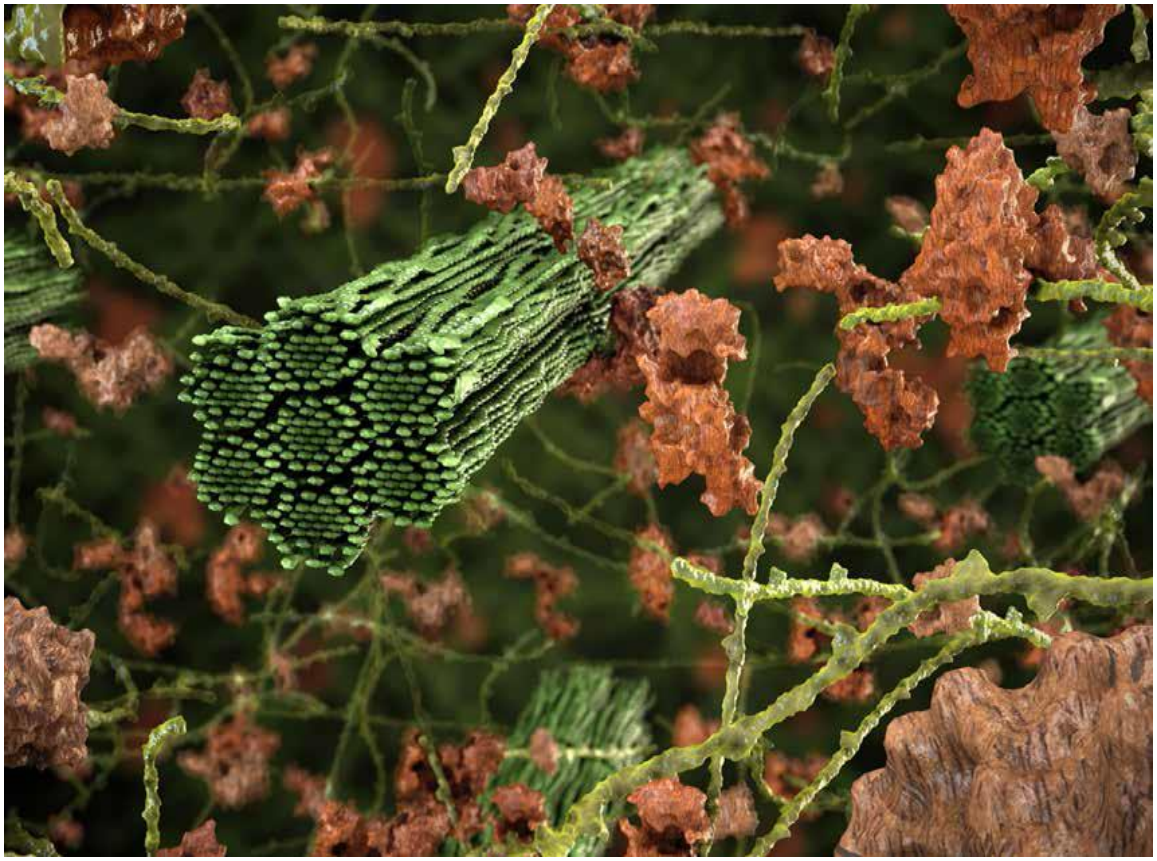


Figure 7: Three-Dimensional Illustration of Lignocellulose Meshwork. Researchers are using computational modelling to gain a molecular-level understanding of the plant cell wall and its major components, including cellulose fibbers (green), lignin molecules (brown wooden texture), and hemicellulose (light green). [Image courtesy Thomas Splett Stoesser, www.scistyle.com, for Oak Ridge National Laboratory]

1.2.1.1 Cellulose

Cellulose is a polysaccharide composed of long chains of at least 500 glucose molecules (Bruce Alberts, et al., 2002). A number of β1,4-linked glucan chains are synthesised in parallel arrays by protein complexes embedded in the plasma membrane to form cellulose micro-fibrils. These polysaccharide chains are bound together by hydrogen bonds to form microfibrils. The microfibrils, in turn, are bundled together to form macro-fibrils (Figure 8).The potential shape of cellulose microfibrils in cross sections of 36-chain (Langan, et al., 2014), 24-chain and 18-chain (Fernandes, et al., 2011) microfibrils and the impact on interactions with other wall components are still a matter of debate(Cosgrove, 2014). Hydrogen bonds make the microfibrils of cellulose strong and resistant to digestion. For an extensive review on recent findings on cellulose

structure and synthesis, see (Cosgrove, 2014). From a biofuel perspective, cellulose is a key target for digestion as it is entirely composed of glucose, which is easy to ferment (Marriott, et al., 2015).

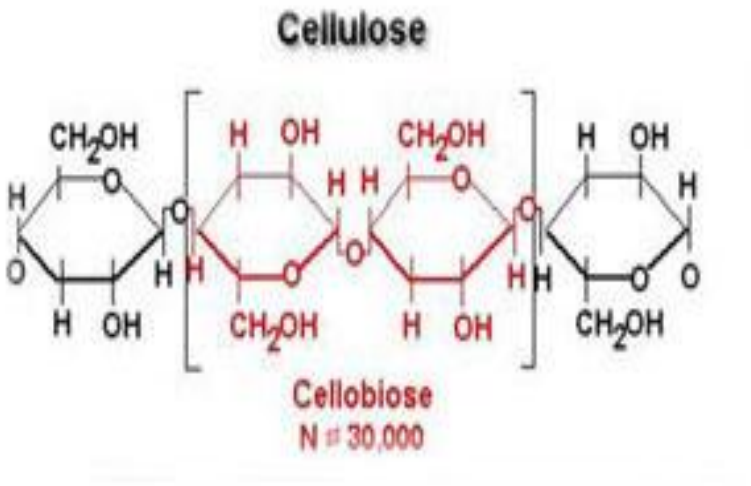


Figure 8: Structure of cellulose in the plant cell wall. A: structure of the β-1,4 glucan chains that make up the cellulose microfibrils, showing the alternating glucose residues rotated 180°**(Marriott, et al., 2015)**.

In higher plants, cellulose is synthesized by plasma membrane-localized rosette cellulose synthase complexes (Li, et al., 2014). The plasma membrane rosettes contain the cellulose synthase catalytic protein subunits (*CESA*) that are encoded by the *CESA* genes (Doblin MS, 2012)(Somerville, 2006).It had been thought that each hexameric rosette is made up by six rosette subunits and that each rosette subunit comprises six *CESA* proteins, enabling the simultaneous synthesis of all the glucan chains in a single micro-fibril to produce, in parallel, 36 glucan chains, providing a total of thirty-six *CESA* proteins per rosette (Delmer, 1999); ;(Somerville, 2006). However, in recent studies on cellulose in celery collenchyma (primary wall) and in spruce wood (secondary wall) with the technical improvement and application of advanced physical methods, combined with modelling, it was proposed that 18–24 glucan chain models was probably closest to the reality, with a slight preference toward a 24-chain model (Fernandes, et al., 2011); (Newman, et al., 2013)(Thomas, et al., 2013). The *CESA* gene family in plant is different in number of genes for each species. For instance, rice has at

least nine (Keegstra & Walton, 2006), maize has at least 12 *CESA* genes (Appenzeller, et al., 2004), barley has at least 8 (Burton, et al., 2004), ten *CESA* genes are present in Arabidopsis, and 18 are the number of *CESA* in poplar (Djerbi, et al., 2005).On the other hand, high similarity is observed between sequence of *CESA* genes in algae and in higher plants (Roberts & Roberts 2004).

Genetic approaches have been applied to discover more genes involved in cellulose biosynthesis, and it three candidates have been discovered: ***(1)*** the *KORRIGAN* gene, which encodes a *b-1,4-glucanase* (Lane, et al., 2001);(Szyjanowicz, et al., 2004); ***(2)****KOBITO1*, a plant-specific gene of unknown function (Gillmor, et al., 2005) but in it has a role in rapidly elongating cells (Pagant, et al., 2002); and ***(3)*** the *COBRA (COB)* gene encodes a glycophosphatidyl inositol (GPI)–anchored plant-specific protein of unknown function (Schindelman, et al., 2001),which might be involved in controlling cellulose microfibril orientation(Roudier, et al., 2005).

1.2.1.2 Hemicelluloses

Hemicelluloses are the non-cellulose cell-wall polysaccharides in plant that have *β-(1→4)*-linked backbones with an equatorial configuration (Scheller & Ulvskov, 2010).They exist in different structural types; and they are sorted into four classes based on the main type of sugar resides present: xylans, xyloglucans, mannans and glucomannans, and *β-(1-3,1-4)-glucans* (also known as mixed linkage glucan (*MLG*)(Ebringerová, et al., 2005). Unlike cellulose that contains only glucose and is unbranched, hemicelluloses can contain many different sugar monomers and are branched polymers consisting of shorter chains – 500–3,000 sugar units as opposed to 7,000–15,000 glucose molecules per polymer(Gibson, 2012).“These linear polymers tend to be insoluble and, because of this, they are usually substituted with other sugar side-chains to prevent the formation of crystalline structure and to increase their overall solubility”(Gomez, et al., 2008). Hemicelluloses can be partially extracted when the biomass is pre-treated with alkaline solution. After that, it can be hydrolysed by enzymes to produce two different types of sugars including hexoses such as glucose, mannose and galactose, and pentoses such as xylose, and (Demirbas, 2009).

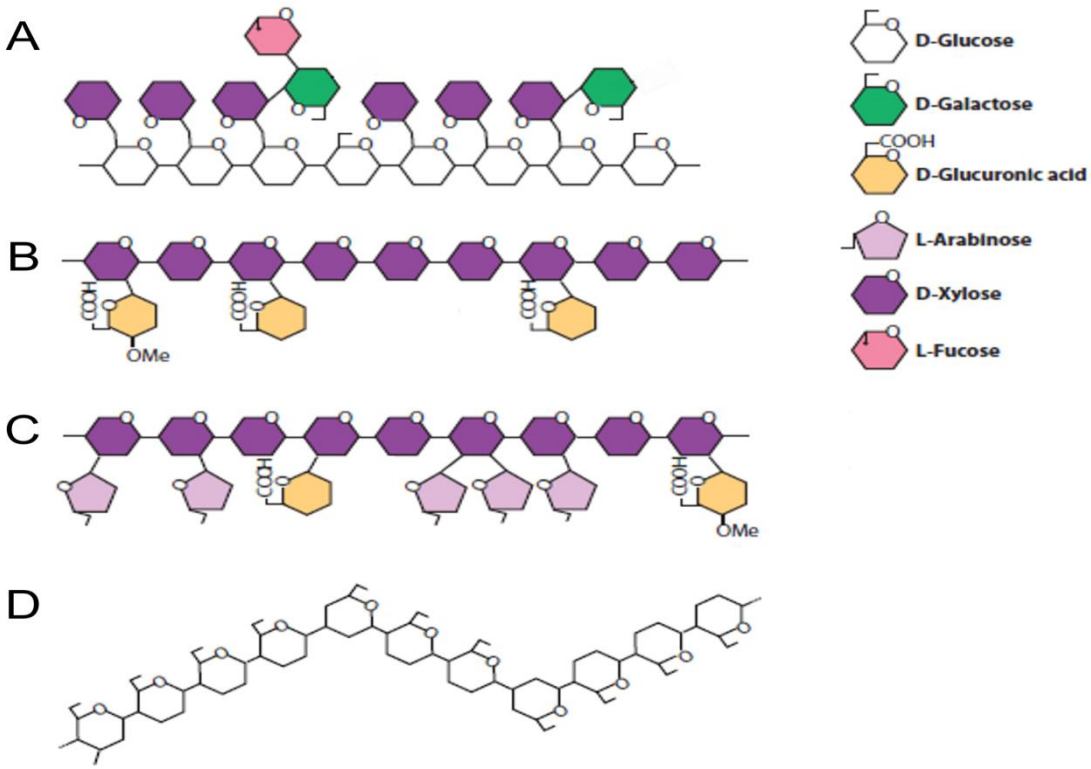


Figure 9: General chemical structure of the most common types of hemicellulose in plant cell walls. A: xyloglucan; B: glucuronoxylan (GX); C: glucuronoarabinoxylans (GAX); D: mixed linkage glucan (MLG). Diagram adapted from Scheller and Ulvskov **(Scheller & Ulvskov, 2010)**

Xylans are the most abundant hemicellulose in nature, found in plant cell walls and some algae (Pradea, 1996). Xylans are polysaccharides comprised of a backbone of *β-(1→4)* linked D-xylose with different side chains depending on species (Hoch, 2007). Xylans make up of 15% to 30% of the total dry biomass in the xylem (Puls, 1997). O-acetyl-4-O-methylglucuronoxylan and arabino-4-O-methylglucuronoxylans are the main xylan components in hardwoods and softwoods (Sixta, 2006).In vegetative part of grasses, heteroxylans, or so-called arabinoxylans, contain glucuronic acid and *4-O*-methyl glucuronosyl residues, hence named as glucoronoarabioxylans (*GAXs*) (Scheller & Ulvskov, 2010).Ester-linked ferulic acid is a crucial component of grass xylans. Ferulate esters linked to O-5 of some of the arabinofuranosyl residues lead to the increase in strength of the cell wall and limit hydrogen-bonding of the xylan to cellulose, decreasing the digestibility of cell walls(Wende & Fry, 1997).*GAX* and lignin are also assumed to be covalently cross-linked through ferulate ester (Grabber, 2005). This cross-link increases the resistance to microorganisms and herbivores, but it also

makes grass cell walls more recalcitrant and hard to be digested by enzyme during hydrolysis process (Buanafina, et al., 2008). It was originally thought that the biosynthesis of the xylan backbone would involve CSL genes due to the structure of the backbone being similar to *β-1,4* glucans. However, there is not yet any research showing that these proteins play any function in xylan biosynthesis. Instead, characterization of a number of xylan-deficient mutants, has revealed that *irx10* (a *GT47* enzyme), *irx14*, *irx9* (two *GT43* members), and their functionally redundant paralogs, *irx10-like (irx10-L), irx9-L,* and *irx14-L*are the ones that carry out this function and are responsible for elongation of the xylan backbone(Cătălin Voiniciuc, et al., 2015). The name of these mutants, *irx,* comes from the observed irregular xylem phenotypes. "Xylem cells are under negative pressure and compromised load-bearing ability is associated with vessel collapse or irregular walls" (Scheller & Ulvskov, 2010). Recently, over-expression of the more closely related rice genes in complemented two well-characterized Arabidopsis *irx* mutants: *irx9* and i*rx14* (Chiniquy, et al., 2013). This demonstrated that *OsIRX9/OsIRX9L*, and *OsIRX14*, have similar functions to the Arabidopsis *IRX9* and *IRX14* genes, respectively, and their role in stem strength (Chiniquy, et al., 2013).

Xyloglucans (XyG) are the major hemicellulose compound in the primary cell wall of all land plant species except for grasses. Secondary walls of xylem seem to contain no xyloglucan(Fry, 1989). Xyloglucans consist of *β-(1→4)* linked D-glucose units with D-xylose residues attached *α- (1→6)* to the glucan chain (Hoch, 2007). Xyloglucan is synthesized in the Golgi trans-cisternae and in the trans-Golgi network (TGN), and transported to the cell membrane by vesicles, where it is expelled and adsorbed on nascent cellulosic micro-fibrils (Moore & Staehelin, 1988). A study applying the approach of genomic comparison identified 862 *XyG*-related genes that included 293 *XTH* sequences, 133 β-galactosidases, 53 β-glucosidases, 24 α-xylosidases, 91 β-*(1→4)*-glucan synthases, 79 α-fucosyltransferases, 108 β-galactosyltransferases and 45 α-xylosyltransferases (*XXTs*) (Del Bem & Vincentz, 2010). It is also thought that there are possibly also hydrolases involved in xyloglucan synthesis but none have yet been identified. It has also been determined that in Arabidopsis members of cellulose synthase like (CSL) genes, which form a superfamily of genes, play a role in the synthesis of backbone of *XyG* (Cocuron, et al., 2007). Although there has been many effort to discover novel genes underlying xyloglucan biosynthesis in recent years

(Zabotina, 2012);(Chou, et al., 2012); (Schultink, et al., 2013), we still don't comprehensively understand the structural organization and regulation of *XyG* and other polysaccharide synthase complexes.

Mannans are a β-(1,4)-linked polysaccharides present not only in land plants but also in many algal species and have been divided in to various types including: pure mannans, galactomannans, glucomannans, and galactoglucomannans. The backbones of manans and galactomannanscontain only mannose, whereas glucomannans and galactoglucomannans have a backbone with combination glucose and mannose residues in a non-repeating pattern, and these backbones may be substituted with α-1,6-linked Gal side chains (Liepman, et al., 2007). It is believed that mannans are functionally diverse, and glucomannans have a structural role (Maeda, et al., 2000). There are evidences that CSLA*s* family involved in synthesising the backbones of mannan or glucomannan in vitro (Dhugga, et al., 2004); (Liepman, et al., 2005).Mutants of nine Arabidopsis *CSLA* genes demonstrate the role of *CSLA2*, *CSLA3* and *CSLA9* in synthesising all detectable glucomannan in Arabidopsis stems, and the role of *CSLA7* in synthesising of glucomannan in embryos. The function of *CSLA* family in encoding glucomannan synthases was consolidated observing the glucomannan deficiency in *CLSA* mutant plants (Goubet, et al., 2009). A gene called "mannan synthesis-related" (MSR), which is also important for mannan biosynthesis, and its lack of function mutant presents a decreased level of mannosyl residues in stem glucomannans of Arabidopsis (Wang, et al., 2013).

(1,3)(1,4)-β-d-Glucan (also known as mixed-linkage glucan or ***MLG***), is a characteristic hemicellulose in primary cell walls of grasses but have not been reported in dicots (Kiemle, et al., 2014).*MLGs* are unbranched and unsubstituted chains of β-glucopyranosyl monomers linked through 30% (1,3) and 70% (1,4) linkages (Albersheim, et al., 2011). Biomass from certain grasses such as rice, Brachypodium distachyon, and sugarcane has a substantial amount of MLG (Souza, et al., 2013).*MLG* biosynthesis is produced by membrane spanning glucan synthase.

1.2.1.3 Lignin

Lignin is the second most abundant organic polymers on Earth, exceeded only by cellulose, accounting for approximately 30% of the organic carbon in the biosphere (Boerjan, et al., 2003). These complex natural polymers are constituted from oxidative coupling of 4-phenylpropanoids and deposited predominantly in the walls of secondarily thickened cells, making them hydrophobic and pose a formidable barrier to wall-degrading enzymes (Vanholme, et al., 2010). Although lignin is "waterproofs the cell wall, enabling transport of water and solutes through the vascular system, and plays a role in protecting plants against pathogens" (Erdtman , 1972), its structural integrity make it extremely recalcitrant and the most important limiting factors to digestion and conversion of plant biomass to pulp or biofuels (Gomez, et al., 2008).

Lignin derives from three major monolignols namely p-coumaryl, coniferyl and sinapyl alcohols, which are characterized by the number of methoxy side groups on the phenolic ring (*p*-coumaryl, zero; coniferyl, one; sinapyl, two); when conjugated into the lignin polymer, the units resulting from the monolignols are called guaiacyl (*G*), syringyl (*S*), and p-hydroxyphenyl (*H*) units, respectively. These units can be bound together in a number of possible patterns (Figure 10).The variation in composition of lignin is seen among taxa, cell types, and different layers of cell wall; and the amount of lignin is influenced by developmental and environmental cues (Boerjan, et al., 2003). Although exceptions exist, dicotyledonous angiosperm (hardwood) lignin consists principally of G and S units and traces of H units, whereas gymnosperm (softwood) lignins are composed mostly of G units with low levels of H units (Boerjan, et al., 2003). Lignin from grasses (monocots) incorporates G and S units at comparable levels and more H units than dicots(Marie, et al., 1998).

Lignin in plants is synthesized in two major stages: monolignol biosynthesis and the subsequent cross-linking of lignin monomers to form polymers, which then connect them to hemicellulose and cellulose(Xu, et al., 2009). Although, current research has mostly focused on monolignol biosynthesis, it has been proposed that laccases and peroxidases are involved in lignin polymerization and cross-linking of the monoligols, but more detailed mechanisms have yet to be unveiled (Liu, et al., 2014).

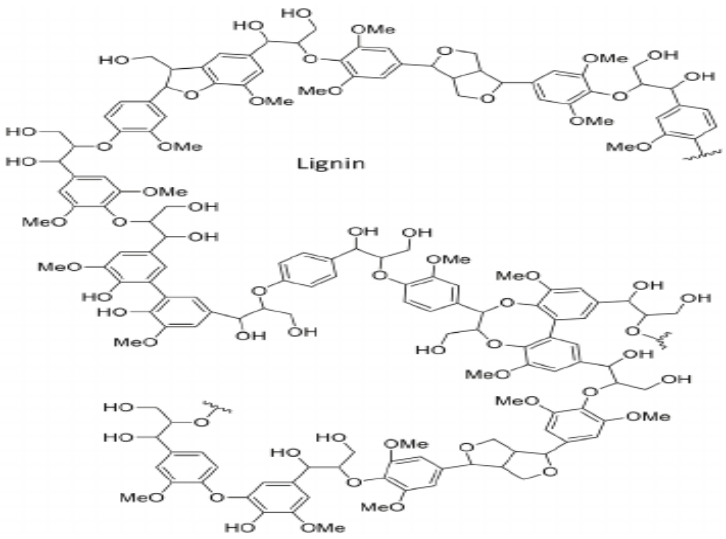


Figure 10: Lignin structure, taken from **(Leisola, et al., 2012)**

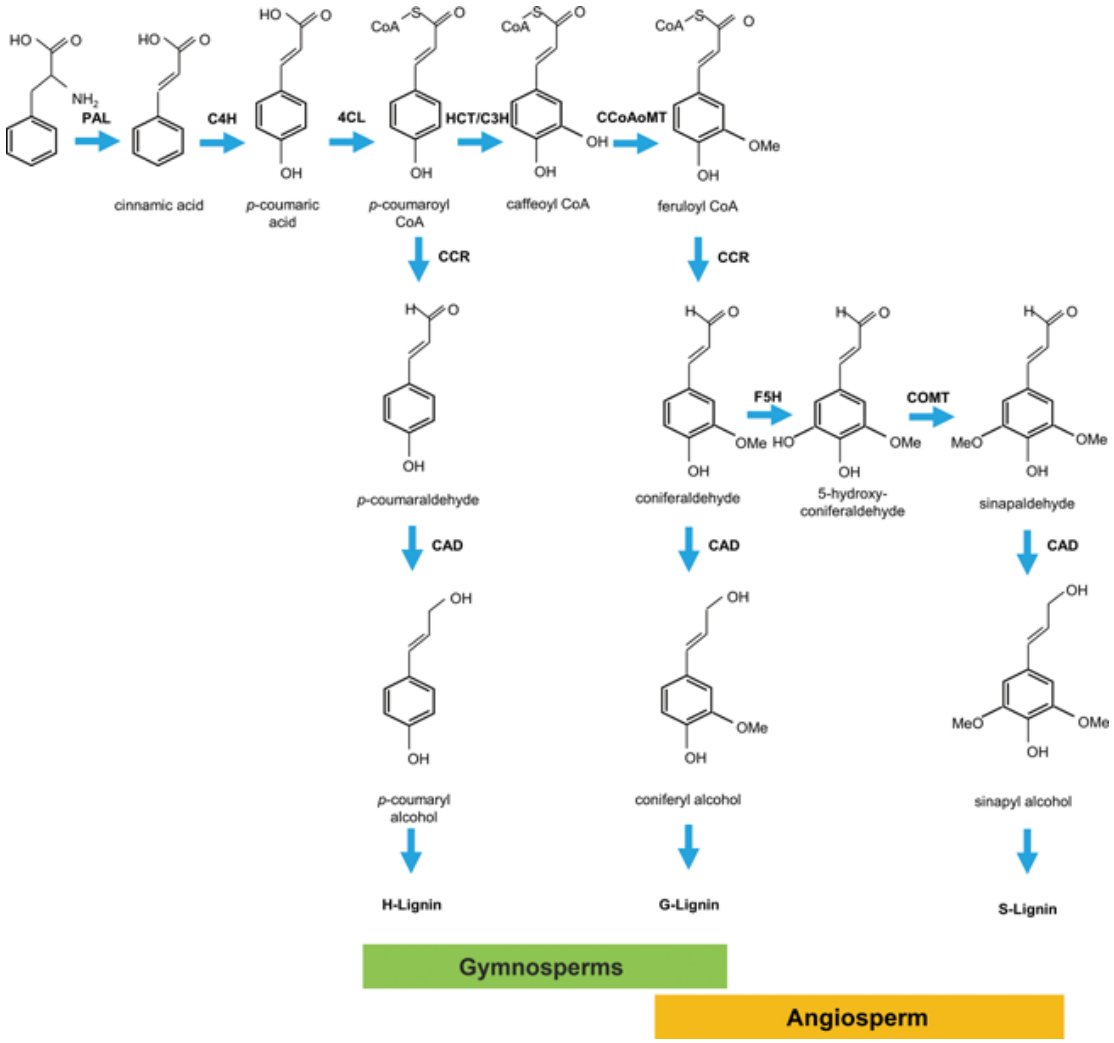


Figure 11: Lignin biosynthesis pathway adapted from Jinmi Yoon et al, 2015

The first step of the phenylpropanoid pathway is the non-oxidative deamination of L-phenylalanine by phenylalanine ammonia-lyase (*PAL*) yielding cinnamic acid (Marie, et al., 1998) (see Figure 11).There are four genes, *PAL1* through *PAL4*, that encode *PAL*in *Arabidopsis* (Raes, et al., 2003).The *pal1/pal2* double mutant is reduced in parallel with an increase in the *S* to G ratio (Rohde, et al., 2004). The plants with the pal1 pal2 pal3 pal4 quadruple mutation have stunted phenotype and lower lignin accumulation(Huang, et al., 2010). In rice, *PAL* genes (*OsPALs*) consist of a large gene family with 11 members (Michigan State University (MSU) rice genome database version 7.0) released from the MSU Rice Genome Annotation Project (MSU, 2009); (Kawahara, et al., 2013). Although the function of *OsPALs* has not been well-confirmed, their involvements in pathogen resistance and their role in phytoalexin biosynthesis in rice leaves has been implied (Park, et al., 2013); (Duan, et al., 2014).

Cinnamic acid 4-hydroxylase (*C4H*) is the second step in the pathway (see Figure 11), in which catalyses hydroxylation at C4 position cinnamic acid to *p*-coumaric acid (*p*-CA).*C4H* is a cytochrome P_{450} monooxygenase, a membrane-bound enzyme involved in a wide range of biosynthetic pathways and usually found in microsomes. In Arabidopsis, *C4H* is encoded at a single locus, and widely expressed in various tissues, particularly in roots and cells undergoing lignification (Bell-Lelong, et al., 1997). Mutants in *C4H* gene in Arabidopsis show pleiotropic phenotypes, including dwarfism, male sterility and the development of swellings at branch junctions (Schilmiller, et al., 2009). Recently, RNAi knock down of the *C4H* gene expression resulted in a significantly reduction of lignin content in transgenic plants and an increase of cellulose content without affecting plant's normal growth. The inhibition of *C4H* gene expression would be an approach to increase the quality of rice straw for bio-refinery purpose(Gengshou, 2013).

4-Coumarate-coenzyme A ligase (*4CL*) is the third enzyme in the phenylpropanoid synthetic pathway, catalysing the formation of CoA thiol esters of 4-coumarate and other hydroxycinnamates in a two-step reaction involving the formation of an adenylate intermediate (Ehlting, et al., 2001). Functional studies of *4CL* in plants have already been conducted in many species such as: in Pinus radiate, severe suppression of *4CL* results in a dwarf phenotype (Wagner, et al., 2009). In switchgrass, silencing of *4CL* (*Panicum virgatum*) leads to reduced lignin content in transgenic biomass and

remarkably increases the efficiency of fermentable sugar release for biofuel production (Xu, et al., 2011). In sorghum (Sorghum bicolor), the brown *midrib2 (bmr2*) mutant lacks function of the most highly expressed 4*CL* in sorghum stems, leaves and roots, both at the seedling stage and in pre-flowering plants. Missense mutations in the two bmr2 alleles show brown coloration in midrib sclerenchyma tissues (Saballos, et al., 2012). In rice, five members of the *4CL* gene family from rice were cloned and analysed (Gu, et al., 2011);(Sun, et al., 2013).*Os4CL2* is associated with flavonoid biosynthesis, whereas the other *4CL* genes (*Os4CL1/3/4/5*) are involved in channelling hydroxycinnamic acid derivatives for lignin synthesis(Gu, et al., 2011);(Sun, et al., 2013). The suppression of *Os4CL3* expression caused significant lignin reduction and shorter plants (Gu, et al., 2011).

The following step in this pathway is mediated by *p*-coumarate 3-hydroxylase (*C3H*), which catalyses the hydroxylation at the C3 position of *pCA* to form CA. Defects in *C3H* cause p-coumarate esters to accumulate rather than p-coumaryl alcohol, and plants display developmental defects such as xylem collapse and a dwarf phenotype (Franke, et al., 2002). In rice, functional *C3H* genes have not yet to be identified.

Hydroxycinnamoyl transferases (*HCTs*) activate the relocation of the cinnamoyl affiliation from hydroxycinnamoyl-CoA to various acyl acceptors such as shikimic acid, quinic acid, hydroxylated acid, and glycerol (Kim, et al., 2012). *HCT* gene repression via RNA silencing in *Arabidopsis* and *N. benthamiana* resulted in a dwarf phenotype and sterility. However, *N. benthamiana* silenced plants are altered soluble phenylpropanoids and lignin content and composition (Hoffmann, et al., 2004). In Rice, four *HCT* homologues were cloned their expression studied in different tissue.

Caffeoyl-CoA O-methyltransferase (*CCoAOMT*) catalyses the methylation at C3 position of the phenolic ring of caffeoyl-CoA into feruloyl-CoA. *CCoAOMT* was initially identified as an enzyme involved in the pathogen defence response of several dicotyledonous species (Pakusch & Matern, 1991); (Kneusel, et al., 1989). Analysis of expression of Arabidopsis *AtCCoAOMT1* utilizing the GUS reporter gene, indicated that this gene is strongly expressed in the vascular tissues of stems and roots(Do, et al., 2007). *CCoAOMT* was functionally studied in maize by RNA interference (RNAi), and this led to the decrease of lignin content and increase in cellulose content in the straw of

transgenic plants (Li, et al., 2013). Three rice *CCoAOMT* genes, *OsCOA1*, *OsCOA20* and *OsCOA26*, were identified and characterised (Zhao, et al., 2004).

Cinnamoyl-CoA reductase (*CCR*) is the first enzyme in the monolignol-specific branch of the lignin biosynthesis pathway, where it catalyses the reduction of hydroxycinnamoyl-CoA thioesters to the corresponding aldehydes (Marie, et al., 1998). Two cinnamoyl-CoA reductase (*CCR*) genes in Arabidopsis, *AtCCR1* and*AtCCR2,* were functionally and structurally characterised. *AtCCR1* is involved in constitutive lignification whereas *AtCCR2* is involved in the biosynthesis of phenolics whose accumulation may lead to pathogen resistance (Lauvergeat, et al., 2001). To confirm the function of *AtCCR1* gene, a down regulation study on homozygous *AtCCR1* plants reported 50% decrease of lignin content and changes in lignin composition and structure, as well as an increase in the enzymatic degradability (Goujon, et al., 2003). Recently, Poplar plants (Populus tremula x Populus alba) down-regulated for cinnamoyl-CoA reductase (*CCR*) were also examined and shown that ethanol yields are 161% higher from trees that are most severely affected (Acker, et al., 2014).

Coniferaldehyde 5-hydroxylase (*F5H*), a cytochrome P450-dependent monooxygenase distinct from *C4H,* has a key function in the formation of S units, catalysing the hydroxylation at the C5 position of ferulic acid, coniferaldehyde, and coniferyl alcohol in the pathways leading to sinapic acid and the syringyl unit of lignin(Marie, et al., 1998). *F5H*protein is rather unstable, low abundance and membrane-bound, therefore it is not easy to target (Chapple, 1998). Rice was the first monocotyledonous plant used to isolate, characterize*F5H*. *OsFSHL* is highly expressed in young leaves, whereas OsFSHL2 is most expressed in mature leaves, both genes are low expressed in the roots and stems.

Caffeic Acid/5- Hydroxyferulic Acid O-Methyltransferase (*COMT*) catalyses the methylation of Caffeic Acid into ferulic acid (*FA*) and 5-hydroxyferulic acid (*5OHFA*) into (sinapic acid) *SA*, using S-adenosyl-L-methionineas methyl group donor (which is converted into S-adenosyl-L-homocysteine) (Marie, et al., 1998).*Brachypodium COMT* gene *BdCOMT4*, were characterized for enzyme activity having significant effect on a broad range of substrates with the highest preference for caffeic acid (Wu, et al., 2013).

Cinnamyl-alcohol dehydrogenase (*CAD*) catalyzes the last step in the biosynthesis of the lignin precursors, converting hydroxyl-cinnamaldehydes into their corresponding alcohols (Marie, et al., 1998)There are nine*CAD*-like genes in Arabidopsis, among them, the *Arabidopsis – AtCAD4 (AtCAD-D*) and *AtCAD5 (AtCAD-C)* are the primary genes involved in lignin biosynthesis in the floral stem of Arabidopsis thaliana by supplying both coniferyl and sinapyl alcohols (Kim, et al., 2004); (Sibouta, et al., 2005). In rice, there are 12 *CADs* in the genome, but few have been studied at functional or expression levels. The proteins encoded by *OsCAD2* and *OsCAD7* are known to function in monolignol biosynthesis(Hirano, et al., 2012). A flexible culm rice mutant (fc1) caused by T-DNA insertion into *OsCAD*7 resulted in reduction of secondary cell wall thickness and decreased mechanical strength of rice plants(Li, et al., 2009). This mutant exhibited an abnormal development phenotype, including late heading time and flexible culm phenotype as well as semi-dwarfism (Li, et al., 2009).On the other hand, among the *OsCAD* genes, *OsCAD2* was the most abundantly expressed in the uppermost internode, followed by more than seven times lower was *OsCAD1*. The other *OsCAD* genes were either not expressed or expressed at very low levels (Hirano, et al., 2012). The GOLD HULL AND INTERNODE2 (*GH2*) gene, which encodes a cinnamyl-alcohol dehydrogenase (*CAD*), was defined as *OsCAD2*. *GH2* acts as a primarily multifunctional *CAD* to synthesize coniferyl and sinapyl alcohol precursors in rice lignin biosynthesis, and the mutant plants displayed the phenotype of reddish-brown pigmentations in the panicle, internode, and basal leaf sheath at the heading stage (Zhang, et al., 2006). In 2014, the idea of combining superior lodging resistance, which can be obtained by introducing thick and stiff culm traits, with low lignin concentrations, which can be obtained using the gh2 variety, was initially developed for improved feed and bio energy production(Ookawa, et al., 2014). The functional studies has been also conducted in maize (Halpin, et al., 1998); in sorghum (Sattler, et al., 2010), and in switchgrass (Saathoff, et al., 2011), and all show that the down regulation of *CAD* expression by RNA-silencing causes a decline in lignification and enhanced saccharification characteristics.

Together, all the studies above show that lignin biosynthesis genes are expressed mainly in vascular tissues at various developmental stages as well as in shoot apical meristem (SAM), epidermis cells, and floral organs. Most mutation in lignin biosynthesis genes causes collapsed xylem element with growth retardation. It is also easy to realize that

mutants of the genes involved in the phenylpropanoid pathway can lead to increased digestibility together with undesirable traits (Yoon, et al., 2015).

1.2.2 Silica in plants

1.2.2.1 Silicon impact in plant growth

Silicon (Si) is one of the most prevalent macro-elements, and it is important for normal growth and development in a number of plant species. In plants, silicon can act biochemically as silicic acid and physically as amorphous silica. “It contributes to cell and plant strength and enables plants to respond adaptively to environmental stresses” (Cooke & Leishman, 2011). A number of studies have shown that Si is able to alleviate both physical stress, including drought, high temperature, UV, water logging, freezing, and chemical stress, including salinity, nutrient imbalance, and metal toxicity (Ma, 2004). Several other studies have also reported that Si is effective in enhancing the resistance to pests and diseases, such as sheath blight, blast, powdery mildew, leafhoppers, leaf spiders, brown plant hoppers, and white back plant-hoppers (Ma, 2004);(Saheb, et al., 2015).

However, silica together with lignin is considered as two important limiting factors in term of digestibility. This has been studied in number of papers. Hasan reported that in rice plants with resistance to lodging, where the culm was short and reinforced with larger quantities of silica, were less digestible(Hasan, et al., 1993). It was also reported in another paper that decrease in silica content to less than 10% dry weight increased the in vivo dry matter digestibility of rice straw as well as the in vitro organic matter digestibility (Balasta, et al., 1989).

1.2.2.2 Silicon form and structure in plant

Silicon (Si) is one of the basic components in most soils(Sommer, et al., 2006). Because of its steady affinity with oxygen, in nature Si always exists as silica (SiO_2) or silicates which are combined with various metals (Ma & Takahashi, 2002). In Soil, the Si content can vary dramatically from <1 to 45 % dry weight (Sommer, et al., 2006). The soil water, or the "soil solution, "contains silicon, mainly present in the form of as silicic

acid, H_4SiO_4, (or its ionized form, $Si(OH)_3O^-$, which predominates at pH > 9) allowing its uptake by plants (Currie & Perry, 2007).Plants take up silicon in the form of silicic acid and deposit it as amorphous silica. The ability to accumulate silica varies among species, with some plants containing only trace amounts, whereas in others, such as rice, silicon constitutes up to 10% the plant dry mass(Epstein, 1999). Due to the difference in the mechanisms of Si uptake and transportation, Si content differs remarkably in different plant tissue, such as roots, shoots, and leaves(Saheb, et al., 2015).

1.2.2.3 Silicon transporters

The uptake systems of Si in rice, cucumber and tomato have been investigated. The results suggest that "both transporter-mediated transport and passive diffusion of Si are involved in the radial transport of Si and that the transporter-mediated transport is an energy-dependent process, and the presence of a transporter for xylem loading is responsible for the high Si accumulation in rice" (Mitani & Ma, 2005). The first two genes encoding a Si transporter were identified by Ma et al in 2006 and 2007, named as *Low silicon rice 1 (Lsi1*) and *Low silicon rice 2 (Lsi2)*, which control silicon accumulation in rice. Rice is widely known as a typical silicon-accumulating plant which requires high Si for healthy growth and high production. This *Lsi1*belongs to the Nod26-like major intrinsic protein (NIP) subfamily of aquaporin-like proteins and is an influx transporter of silicic acid. *Lsi2* encodes a putative anion transporter and is an active efflux transporter of silicic acid. Both *Lsi1* and *Lsi2* are constitutively expressed in the roots. Suppression of *Lsi1*or *Lsi2*expression resulted in reduced silicon uptake therefore both*Lsi1* with *Lsi2* are required for efficient uptake of Si in rice(Ma, et al., 2006); (Ma, et al., 2007). The Si uptake and accumulation processes were revealed clearer by the discovery of another transporter named as Lsi6 by Yamaji et al in 2008. To start, Si transported via *Lsi1* and *Lsi2* into the stele is then translocated to the shoot by the transpirational flow through the xylem in the form of mono-silicic acid(Mitani & Ma, 2005). Finally, Si must be transported out of the xylem for the deposition as a polymer of hydrated amorphous silica.*Lsi6* was indicated to be responsible for the transport of silicic acid from the xylem into xylem parenchyma cells, thereby having influence on the subsequent Si distribution in rice shoots(Yamaji, et al., 2008).

1.2.2.4 Silicon and the cell wall

The interest in silica and its link to the cell wall have not yet been fully understood. It was first suggested by Schwarz (1973) that there are high level of intrinsic "Si" found in plant pectin (Schwarz, 1973). In 2009, Currie & Perry demonstrated the role of the Si in the rigidity the cell wall by crosslinking, showing chemical evidence for intrinsic Si within Equisetum cell walls. Nevertheless, they were not able to point out the precise chemical link of the Si to the plant cell wall polymers(Currie & Perry, 2009). On the other hand, with high silica content, rice is always considered as the model species for these studies. In 2012, a study by Yamamoto et al, revealed the changes in cell wall organic components induced by Si deficiency in rice. The cell wall has been reported to become thicker, sugar content in the cellulosic fraction and lignin content increased in the absence of Si(Yamamoto, et al., 2012). Expression of the genes involved in secondary cell wall synthesis, *OsCesA4, OsCesA7, OsPAL, OsCCR1* and *OsCAD6* was also looked up. And all of them were up-regulated under -Si condition, excepting *OsCesA1*(which is known to be mainly involved in primary cell wall synthesis). The increase in secondary cell wall biosynthesis might be explained as a key factor that compensates for the reduction in stress resistance caused by Si deficiency since it is known that Si enhances resistance to physical and biotic stress(Yamamoto, et al., 2012). This suggests that rice might expend less energy for stress resistance by using inorganic Si instead of organic material. A recent report about a hemicellulose-bound form of silicon in rice and distinctive effect of silica to cell wall composition and major wall polymer features and lignocellulosic saccharification has strengthened the idea that there may be links between Si and wall components(He, et al., 2015); (Zhang, et al., 2015). Silica levels are also reported positively correlated with three major wall polymers, indicating that silica is associated with the cell wall network(Zhang, et al., 2015).

1.3Using genomics to identify genes for biomass improvement

A possible route to improving the cost effectiveness of the production of biofuels is the generation of lignocellulosic biomass that has improved digestibility characteristics. It has been demonstrated that alteration of a number of cell wall components can affect

recalcitrance of lignocellulosic biomass and thus improve its saccharification and/or digestibility. This could be achieved by altering the plant cell wall properties through plant breeding (Foust, et al., 2008). Some work has been done indicating that improved digestibly can be achieved when cellulose crystallinity or hemicellulose composition is altered (Carroll & Somerville, 2009); (Vega-Sánchez & Ronald, 2010). However, most of the work has focused on increasing digestibility by altering the lignin structure and composition but this could lead to undesirable plant growth effects such as low biomass yield (Chen & Dixon, 2007); (Vega-Sánchez & Ronald, 2010). Although some important advances have been made to lay the foundations for plant genetic engineering for biofuel production, and the list of genes that can be manipulated for pathway engineering is growing, this science is still in its infancy and there also several challenges outside the realm of genetic engineering that need to be addressed (Sticklen, 2008).

1.3.1 Rice as a model for cereals and grasses, and as a sustainable cellulosic biomass feedstock

In recent years, interest has grown in using grass straw as a source of biomass for the production of second generation biofuels (Gomez, et al., 2008). Unfortunately, agriculturally important grasses typically have complex genomes and growth requirements that make them cumbersome for research purposes. The exception to this is rice (Oryza sativa), which is the third biggest cereal crop in the world, has a small diploid genome, and well developed molecular genetics tools (Gomez, et al., 2008). Much work has been done on understanding the synthesis and construction of cell walls in Arabidopsis (Fagard, et al., 2000) but, this research cannot directly be transferred to grasses as they differ to dicots in terms of both the composition and organisation of the various polymers found within their cell walls (Vogel, 2008).

Up to 70% of rice straw biomass is comprised of polysaccharides with equal amounts of cellulose and arabinoxylan that could be converted to sugars for fermentation. This polysaccharide network is interpenetrated and coated by lignin, a highly resistant polyphenol, which protects the sugar polymers from enzymatic attack. A notable characteristic of rice straw is that it has low lignin and high silica compared with other straws(Van Soest, 2006). After all, we can see that lignocellulosic materials not only

have potential for biofuel production but also value as animal feed, but the poor palatability and indigestibility of straw makes this inefficient.

1.3.2 Genome wide association studies

In plant and animal, studies of genetic sources of phenotypic variation have been the key to determining the cause of disease, improving agriculture and understanding adaptive processes.(Brachi, et al., 2011). Quantitative trait loci (QTL) were originally mapped in bi-parental crosses in plants(Flint-Garcia, et al., 2003). The bi-parental mapping population often have the highest genetic resolution of QTL maps ranged from 10 – 30cM, which is due to the restricted number of meiotic events captured after a cross between two parental lines (Zhu, et al., 2008). There is only a small amount of all possible alleles that could be examined for linkage analysis in the population from which the parent (Pasam, et al., 2012).

Genome wide association studies (GWAS) have been considered as a powerful approach to overcome the constraints inherent to linkage mapping, through studying in the variation in existing natural or designed population. Linkage disequilibrium (LD) mapping, or so called Association Mapping (AM) exploits historical recombination events that occurred throughout the whole genomes in the population; and all major alleles present are then taken into account to identify significant marker-phenotype associations(Pasam, et al., 2012). LD mapping was first applied in genetic mapping studies in humans (Hästbacka, et al., 1992). After over two decades, this mapping method has provided a powerful tool for fine structure localization of genes responsible for complex traits in many different species, including crops. The improvement of cost effective and high-throughput sequencing technology has made available mapped markers in many research objects. Using those sets of mapped markers and merged with the phenotypic data of studied traits, it is completely possible to identify the genomic regions with non-random associations of alleles at nearby loci (LD) which significantly associate to the trait variance in the research population. There are three key factors driving the success of mapping: the quality of phenotypic data, population size and the degree of LD present in a population(Mackay & Powell, 2007). The power of association studies generally depends on the degree of LD between genotyped marker and the functional polymorphisms(Pasam, et al., 2012). The decay of LD has been seen

varies greatly between species, among different population within one species among different loci within one species, and also among different loci within a given genome(Gupta, et al., 2005); (Caldwell, et al., 2006).

Genome wide association study (GWAS) and re-sequencing of selected candidate genes are the two strategies of LD mapping(Hirschhorn & Daly, 2005). By exploiting marker polymorphism across all chromosomes with significantly high resolution, GWAS have become increasingly popular and powerful over the last few years.(Pasam, et al., 2012). There has also been increasing number of association studies based on the analysis of candidate genes (Thornsberry, et al., 2001); (Palaisa, et al., 2003); (Stracke, et al., 2008); (Zhao, et al., 2007); (Singh, et al., 2009). Different from GWAS in humans, GWAS in crops usually use a population of diverse (and preferably homozygous) varieties that can be re-phenotyped for many traits and only needs to be genotyped once—and one can subsequently generate specific mapping populations for specific traits or QTL in crops (Huang & Han, 2014). So far, GWAS has been carried out successfully in many crops, including maize, rice, sorghum, and foxtail millet(Huang & Han, 2014). Rice is a selfing species and, like *Arabidopsis*, a good candidate for GWAS. Huang et al. identified an unbiased set of common SNPs that they used to identify strong associations between genetic loci and 14 agronomic traits, including heading date, grain size, and starch quality(Huang, et al., 2010).

The advent of high-density single-nucleotide polymorphism (SNP) genotyping allowed whole-genomes scans to identify often small haplotype blocks that are significantly correlated with quantitative trait variation (Brachi, et al., 2011). Recently, Professor Claire Halpin and Professor Robbie Waugh at the University of Dundee and Professor Simon-McQueen Mason at the Centre of Novel Agricultural Products (CNAP), University of York have been using GWAS to identify QTL for cell wall saccharification potential in barley. This involved screening multiple replicates, over two growing seasons of more than 600 barley accessions, collected from around the world, for their susceptibility to hydrolysis with commercial cellulase using high-throughput automated assay developed at York. Analysis of the data has allowed the identification of 12 significant QTL for this trait, one of which has been resolved to a cluster of lignin genes.

Based on the knowledge and achievements from the GWAS in barley, carrying out similar studies in rice will add great value to these studies. Firstly, the smaller, fully sequenced genome of rice will allow much finer resolution of QTL to individual genes in a shorter timeframe than in barley. Secondly, carrying out this work in a second cereal will allow insight into the transferability of observation between cereal species and reveal the extent of commonality between species. With the well-developed molecular genetics tools of rice, the advent of affordable large-scale DNA sequencing, and association genetic studies starting to reach their full potential, there is no doubt that the genome-wide association study (GWAS) has the potential to clarify the functional role of genes as well as to identify novel genes involved in complex traits such as cell wall synthesis and silica transportation in the rice.

The present project builds from initial studies jointly funded by the BBSRC and Vietnamese Ministry of Science and Technology (MOST) and led by Simon McQueen-Mason in the UK. In this work, I was working jointly at the Universities of Dundee and York assembling and genotyping a range of Vietnamese rice varieties to form a panel of plants for genome-wide association scans. A saccharification assay for the rice population of 116 varieties was also done in 2014, and ready as phenotypic data for association mapping. By the time I started my MSc by research in April, I already had the genotypic data of 172 varieties, which is a SNP panel of 300.000 Single Nucleotide Polymorphism. This data was collected from Genotyping by sequencing assays run on an Illumina platform in Susan Mc-Couch's Rice Laboratory, Cornel University, US. The sequencing data were run through Tassel GBS pipeline before getting the usable SNP files. In addition, I have been using straw produced from 151 of these varieties, which were grown in Vietnam.

From the availability of genotypic and straw material we decided to carry out the following experiments:

1. Determine the saccharification, lignin content and silica from 5 replicates of 151 rice genotypes included in the association panel to conduct association mapping for those traits.
2. Association analysis to identify QTL for the extended population and 2014 population.

3. Explore the biochemical basis of QTL by carrying out compositional analysis in the rice straw of the highest and lowest saccharifiers with contrasting alleles.
4. Clone the candidate genes associate with QTL from contrasting lines/varieties to identify polymorphisms.

1.4. Aim of the project

The aims of the proposed research are to carry out underpinning research to increase the attractiveness of rice straw as animal feed and for the production of biofuels by improving the quality of straw through crop breeding. To facilitate the advance in crop breeding, I will use the power of genome wide association studies (GWAS) to identify potential markers and candidate genes that: 1) increase digestibility of straw without impacting on grain yield, 2) decrease silica content without impacting on yield, and 3) increase the scientific understanding of biomass composition and biosynthesis.

Chapter 2 - Cell wall characterization

2.1 Material

The association panel comprises of 151 rice accessions from Vietnam, which originate from two Oryza sativa subspecies: *indica* and *tropical japonica*. These accessions were selected from a trial population derived from a breeding project at the Plant Biotechnology Division, Field Crops Research Institute (FCRI), 84 different varieties which are reserved in the Germplasm Bank of FCRI, 29 high-quality varieties which are popularly cultivated in different areas in Vietnam, and 38 landrace cultivars. These collected genotypes are expected to be highly inbred lines with homozygous genomic background. (*Appendix A contains the list of the population used)*

The panel was grown in the field. Each straw sample was collected in 5 replicates after plants were cut to harvest grains. The samples were taken from the main tiller (Figure 12). The straws collected were dried for two days under sunny conditions in Vietnam. Straw samples were kept in separate paper bags and sent to Centre for Novel Agricultural Product (CNAP), University of York, UK for characterisation.

At CNAP, rice stem samples (minus nodes) were cut into small pieces and then ground to a fine powder and stored in labelled tubes (Figure 13). These samples would be used for different assays including saccharification, silica content and total lignin content.

Figure 12: Rice stem sample

A B

Figure 13: (A) Cyclone mill machine; (B) Stem sample after milling

2.2. Methods

2.2.1 Saccharification assay

The method use robotic system described by Gomez et al., (Gomez, et al., 2011). This could be summarized in brief as follow: the grinded straw samples were randomised before formatting in 96 well plates. For each sample, four technical replicates of 4mg were dispensed into the 96 well plates after weighing out by the robot (Labman Automation, Stokesley, North Yorkshire, UK)(Gomez, et al., 2010). The samples were screened using the liquid handling robot (Tecan LTD, UK), which utilised a programme involving a pre-treatment of water at 70^0C for 20 minutes and an enzymatic incubation time of 8 hours at 50^0C (Figure 14). The enzyme used for digestion is 4:1 mixture of commercial enzymes Celluclast (Cellulase from Trichoderma reesei) and Novozyme 188 (cellobiose from Aspergillus niger) (Novozymes). Final step is to measure the reducing sugars release from the biomass material. This was done by doing colorimetric assay, using an adapted 3-methy-2benzothiazolinonehydrazone method (MBTH)(Gomez, et al., 2010),(Gomez, et al., 2011). Three standards of 50, 100 and 150 nmol glucose (three replicates each) and filter paper disks (four replicates) – as

control were used to account for any change in enzyme concentration or condition through time. The plates containing samples, standard, filter paper disks were read through Tecan Sunrise microplate absorbance reader at 620nm to measure the variation in absorbance.

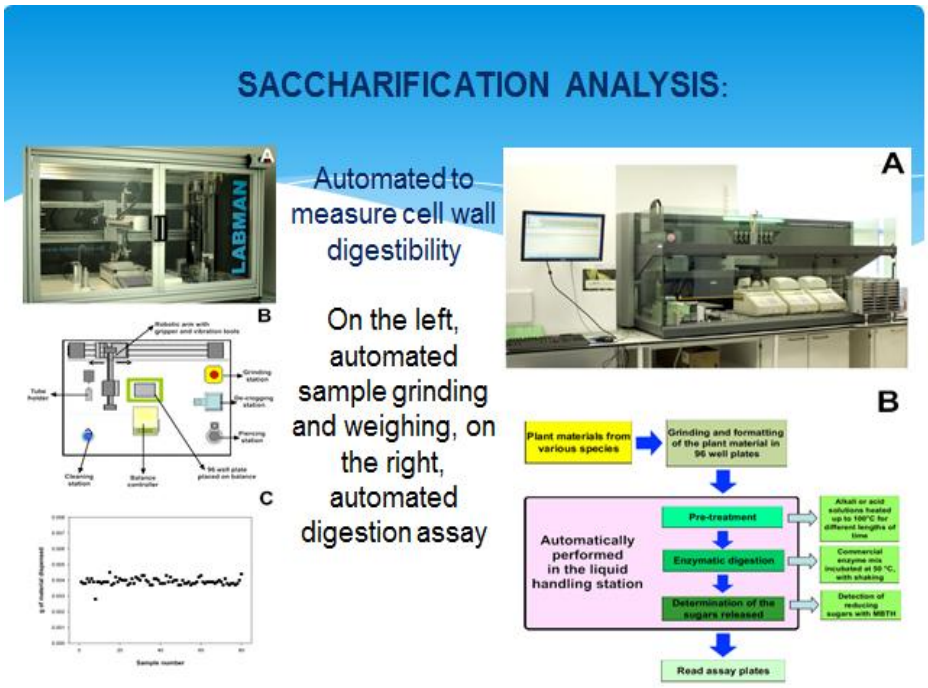


Figure 14: High-throughput saccharification assay system developed by Leonardo Gomez since 2010 in CNAP

2.2.2 Silica content measurement

For silica analyses, the samples were further milled, using a Tissue Lyser II, with the grinding jar sets and stainless steel grinding balls at a vibration rate of 3000rpm for 70 seconds.

Powder from five biological replicates of each rice accession, 755 rice stem samples in total, was pelleted using a 10 ton manual hydraulic press (Specac, UK) (Figure 15). The pellets were analysed using a portable x-ray fluorescence spectrometer (Niton XL3t900 GOLDD Analyzer; Thermo Scientific, UK) to determine the amount of silica present (Reidinger, et al., 2012) (Figure 15). A standard curve was produced using "bush" as the certified reference materials (CRM); The sample pellets were read twice (each side

once), whereas “bush” standards were scanned again after every 10 samples to check the stability and reliability of the assay.

Figure 15: The sample pellet and P-XRF machine

2.2.3 Total lignin content

Lignin content was quantified using the method reported by Fukushima and Hatfield (Fukushima & Hatfield, 2004). Briefly, 4 mg ground samples were weighed into a 2ml tubes and 250µl freshly prepared acetyl bromide solution (25% v/v acetyl bromide/75% glacial acetic acid) were added in order to break phenol bonds in the biomass. Samples were then incubated at 50°C for 2 h, followed by a further 1 h with vortexing every 15 min to solubilise the lignin. Samples were then cooled to RT before being transferred to 5ml volumetric flasks. Subsequently 1 mL of 2 M NaOH was added to hydrolyse excess acetyl bromide, followed by 175 µl freshly prepared 0.5 M hydroxylamine hydrochloride. Samples were then made up to 5 ml with glacial acetic acid, mixed, before continuing to dilute 10 times by mixing 100ul sample with 900ul acetic acid, and the absorption read using a Shimadzu UV-1800 spectrophotometer at 280 nm (Figure 16). Lignin content (µg mg-1 cell wall) was then determined using the following formula: (absorbance ÷ (coefficient x path length)) x ((total volume x 100%) ÷ biomass

weight)). Because the coefficient is different among plant when analysed, the coefficient for grass (17.75) was used for rice (Fukushima & Hatfield, 2004).

Figure 16: Spectrophotometer: The detection of total lignin was carried out with a UV-detector at 280 nm after Acetyl Bromide Assay

2.2.4 Data analysis and heritability estimation

I carried out analysis of digestibility, lignin content and silica content on straw samples (5 biological replicates) from field grown samples of the diversity panel grown in Vietnam. The raw data from these analyses were gone through post hoc analysis (Jaccard, et al., 1984) and corrected by the final LSD test linear model (Hayter, 1986) with Bonferroni's adjustment (Abdi, 2007). Based on the phenotypic data, the heritability was calculated for amount of sugar released, silica content, and for total lignin content. This is described by the document in the R program, which was used for most statistical analyses, which states that: "Given a population where each genotype is phenotyped for a number of genetically identical replicates (either individual plants or plots in a field trial), the repeatability or intra-class correlation can be estimated by V_g / (V_g + V_e), where V_g = (MS(G) - MS(E)) / r and V_e = MS(E). In these expressions, r is the number of replicates per genotype, and MS(G) and MS(E) are the mean sums of squares for genotype and residual error obtained from analysis of variance. In case MS(G) < MS(E), V_g is set to zero (Singh *et al.* 1993); (Lynch and Walsh 1998. When the genotypes have differing numbers of replicates, r is replaced by \bar r = (n-1)^{-1} (R_1 - R_2 / R_1), where R_1 = ∑ r_i and R_2 = ∑ r_i^2. Under

the assumption that all differences between genotypes are genetic, repeatability equals broad-sense heritability; otherwise it only provides an upper-bound for broad-sense heritability"

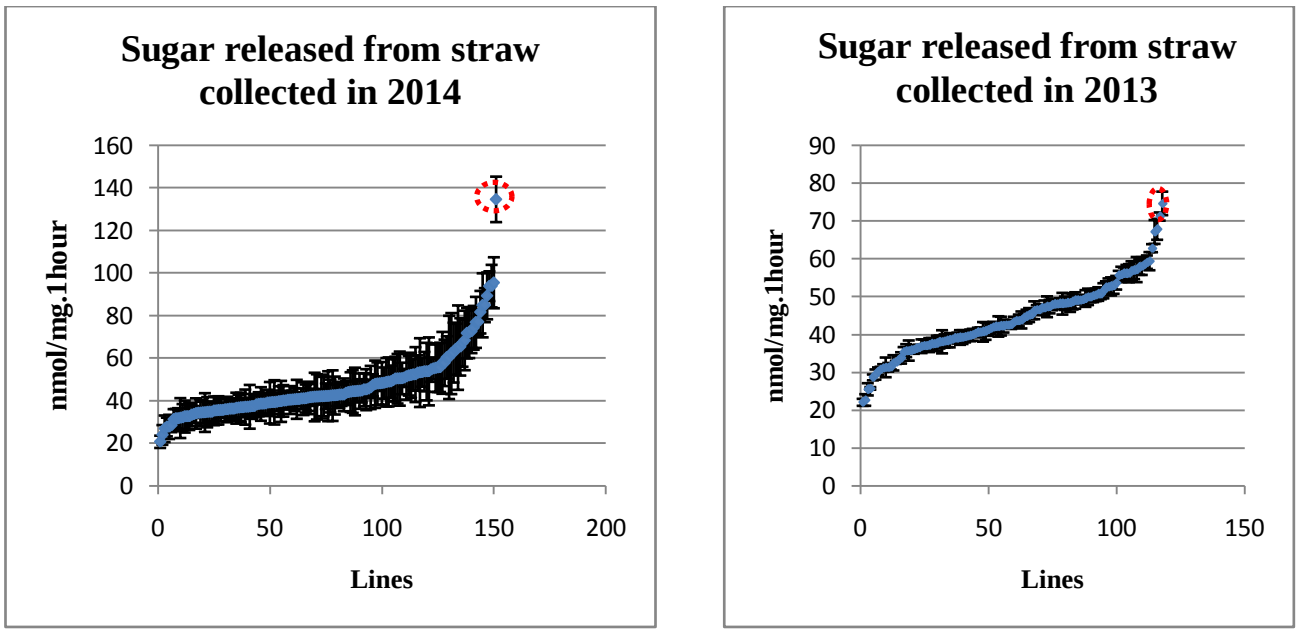
2.3 Results

3.1 Saccharification

The Saccharification results, reported in this thesis, are for two sets of data, spring season 2013 (93 lines) and summer season 2014 (151 lines). The range of digestibility in the 2014 data set is from 20 to 134 (nmol/mg.1hour), and in the 2014 data set is from 23 to 72.8 (nmol/mg.1hour) (Figure 17). There is little correlation over two year data sets when observing the same 93 lines in term of saccharification (Figure 18). The lack of such correlation between two set of data might be because of environmental effects on population. Most rice varieties are adapted for growth in a specific growing season. Some may be adapted for both season, many are just suitable to grow in one season only. Therefore if they are grown in the unsuitable season, this could cause differences in biomass quality.

Although there was no correlation between two sets of saccharification data, it is notable that the line U17 showed the highest digestibility consistently in two years (72.7 and 134.4 nmol/mg.1hour in 2013 and 2014, respectively) (Figure 17), in 2014 data set it is even easy to see its outstanding digestibility with 6 fold higher than the lowest digestible one.

These data sets show a high range of values within the population, which is essential for studying the association between the digestibility and the bi-allelic markers, which are SNPs stored in HapMap, across the genomes.



(A) (B)

Figure 17: (A) Scatter plot graph adapted standard error values of 151 rice lines, 5 biological replicates, grown in 2014, 2013 (B) Scatter plot graph adapted standard error values of 98 rice lines, 3 biological replicates, grown in 2013. The dot circled by the red oval is the line U17.

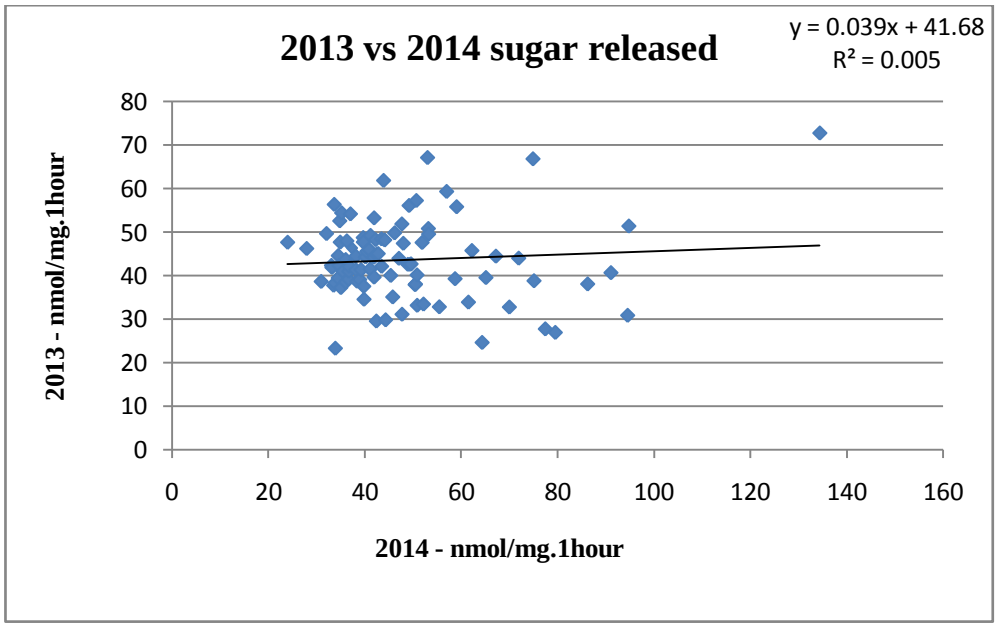


Figure 18: Correlation between 2013 vs 2014 sugar released

2.3.2 Silicon and silica content

The range of Si content in the rice population is from 0.815 to 2.57 (% in dry weight). These values of Si are multiplied by 2.14 to get the amount of silica (SiO_2). We will then have the silica content in all the samples of population, which varies from 1.57 to 5.5 (% in dry weight)(Figure 19).

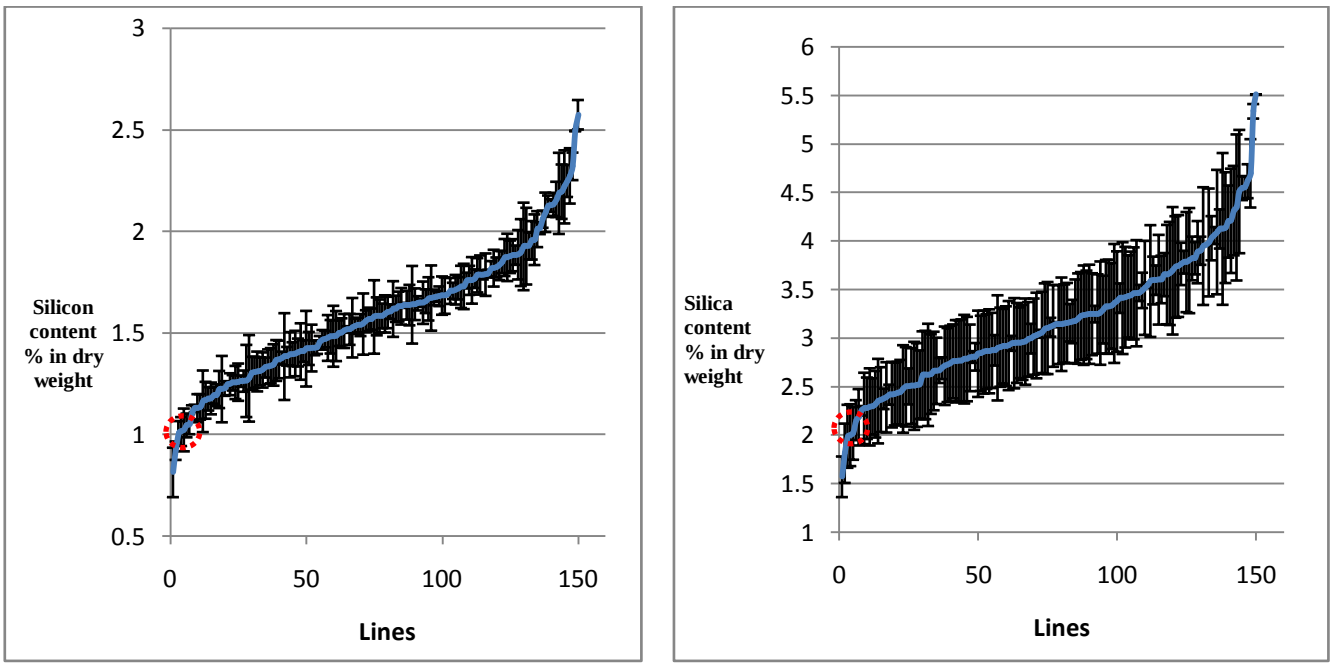


Figure 19: Scatter plot with standard error bars for silicon and silica content in 151 rice lines: 2 technical and 5 biological replicates per line. Circle in red indicates U17 Data were determined using X-ray fluorescence.

Once again, we see a good amount of variation across the rice panel and it is interesting to note that the highly digestible U17 line has one of the lowest silica concentrations seen in the diversity panel at less than 2.0 % /dry weight silica. This line has a strong stem with abundant biomass material, high digestibility and low silica. Unfortunately, we only have Si data for the straw from the 2014 field trial as there was insufficient straw from 2013 to allow this measurement to be made.

The silicon and silica content also shows sufficient variation among the varieties in the population to allow genome wide association studies to identify the QTL controlling the accumulation of silicon/silica in rice plants.

3.3 Total lignin content

Examination of lignin content using the acetyl bromide method did not reveal as high levels of variation among rice accessions as was seen in digestibility and silica content. There is 26.3 % of lignin content in the highest variety, whereas the variety that present the lowest level contains 14.3 % of lignin (Figure 20). However, it is still at a useful level of variation to be used for genome wide association studies. In addition, once again, we see that the line U17 is belongs to the group that have lowest lignin content with only 15.9 % of lignin is found in the stem biomass of this line.

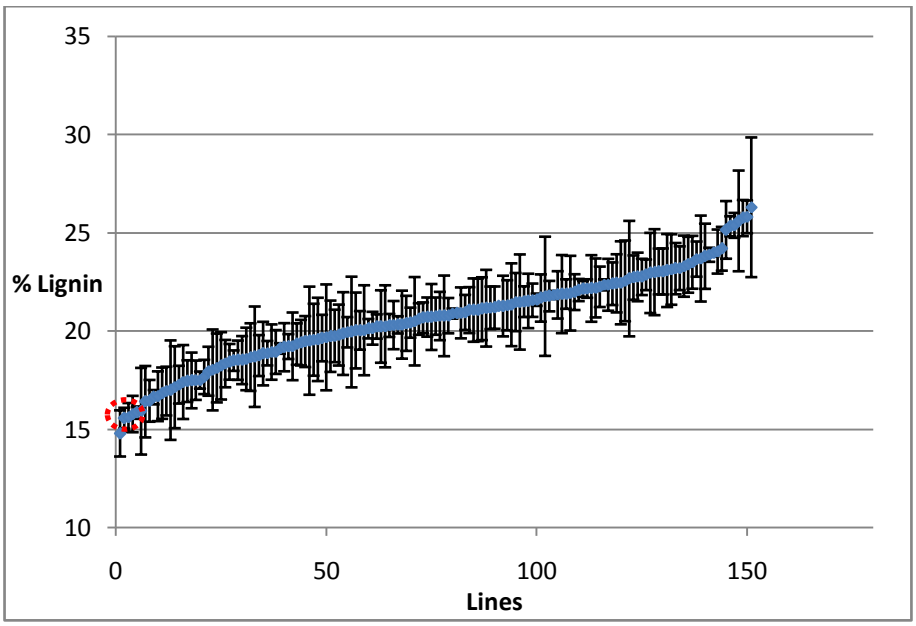


Figure 20: Total Lignin content with standard error bars. Lignin was measured using the acetyl bromide method and data are result of 151 lines, 3 biological replicates per line. Circle in red indicates U17.

3.4 Correlation analysis

Correlation analysis was performed to check if any of three studied traits above correlate with each other (Figure 21). While there is only a very weak correlation between Lignin and Silica(R^2 = 0.0036), there is a higher, but still rather weak negative correlation between Lignin content and digestibility (R^2 = 0.0224) (Figure 21). In contrast, there is a much stronger negative correlation between Silica content and digestibility (R^2 = 0.2385). This result agrees with published papers about the

relationship between silica and digestibility(Van Soest, 2006); (Binod, et al., 2010); (Zhang, et al., 2015).

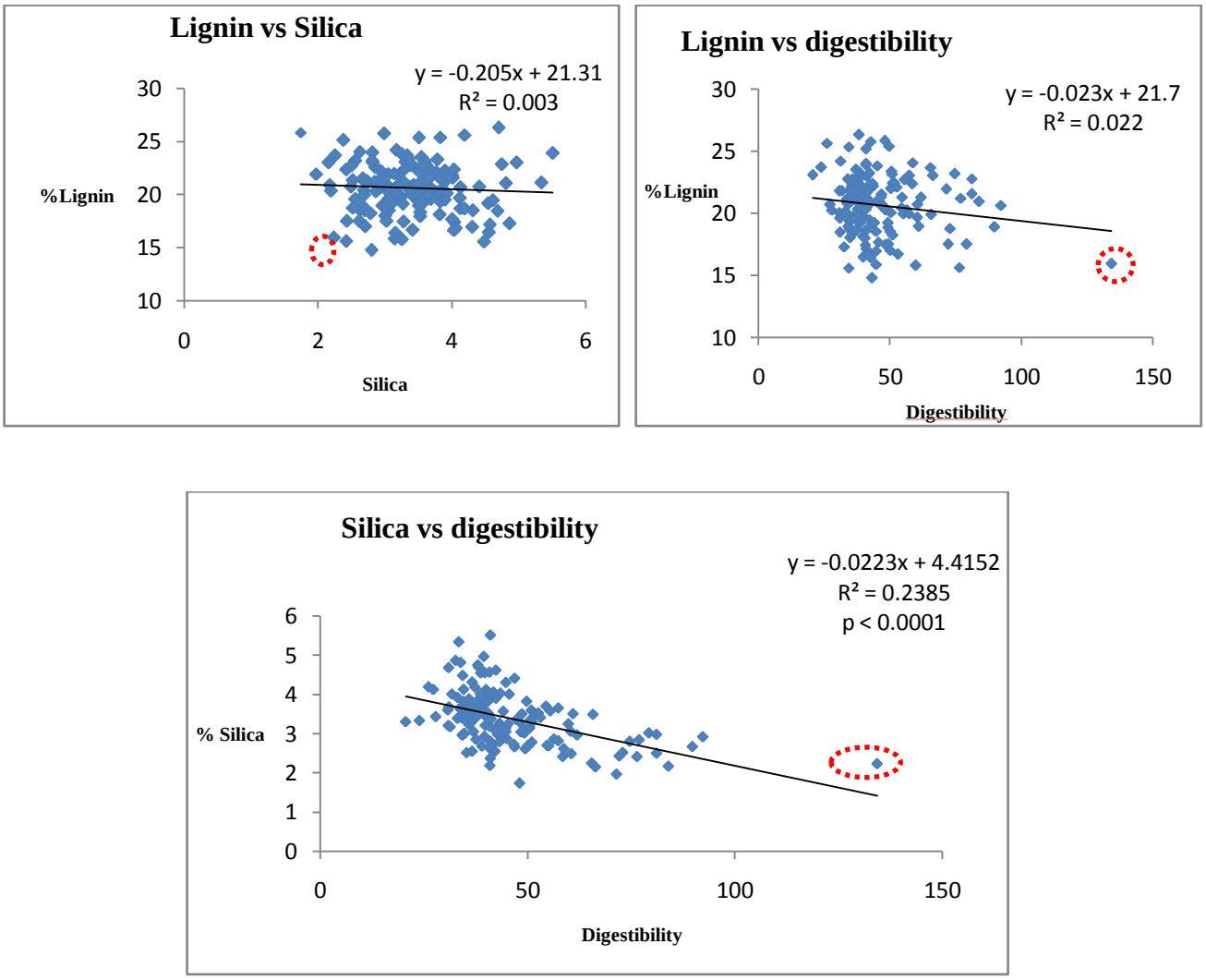


Figure 21: Correlation graph among three traits: Lignin vs Silica, Lignin vs digestibility, and Silica vs digestibility, respectively. The dot circled by the red oval is the line U17.

2.3.1 Heritability of studied traits

Based on the phenotypic data, the heritability was calculated at 0.79 (2013) and 0.65 (2014) for the amount of sugar released, 0.67 for silica content, and 0.62 for total lignin content (Table 3).Heritability is classified as low (5-10%), medium (10-30%) and high (>30%)(Dabholkar, 1992). The heritability of our studied traits is rather high and implies that on average about 62% to 79% of the individual difference observed on these traits in our population are attributable to genetic variation, and 21% to 38 % are attributable to the environmental variance. In 2010, a study looking for the QTL for silicon content in leave, using a Recombinant Inbred Lines (RIL) population has a much low heritability of silica, at 0.37(Norton, et al., 2010). This result shows similar level of heritability to other secondary compounds in evening primrose (*Oenothera biennis*) with the mean of 15 studied trait calculated at 0.68 (Johnson, et al., 2009). In wood, heritability estimates range from 0.16 to 0.97 for cell wall carbohydrate and 0.42 to 0.79 for lignin content(Poke, et al., 2006)(Ukrainetz, et al., 2008). The high heritability of these traits indicates that there is good scope to improve the digestibility through crossing and selection of the line with good quality of straw.

Table 3: Heritability of traits studied, which is based on the repeatability of phenotypic data.

Trait	Repeatability	Genetic variance	Residue variance	Repeatability of Lines	Average number of Replicate	Confident Interval
Sugar released (2013)	0.79	88.32	3.12	TRUE	7.2	0.75 -0.84
Sugar released (2014)	0.65	242.01	130.1	TRUE	20	0.60 - 0.71
Silica (2014)	0.67	0.39	0.196	TRUE	4.5	0.60 - 0.73
Lignin (2014)	0.62	3.47	2.15	TRUE	3	0.58- 0.65

2.4 Discussion

Throughout these analyses the cultivar, U17, showed remarkable straw quality through 3 different experiments, showing high digestibility and low content in both silica and lignin. Interestingly, this cultivar is a commercial variety which has been cultivated in many areas in Vietnam. This variety was originally bred from hybrid combination of IR5 ´ [(IR8 ´ 813) ´ IR 1529-640-3-2] by scientists at the Field Crops Research Institute in 1988, and has been long well-known its resistance to submergence and water logging.

The level of variation apparent in the population differs in each of the studied traits. There is much higher variation apparent in digestibility and silica content than there is in lignin content.

The lignin content in our rice accession straws are at the similar level as for grasses in general and higher than in dicot but lower than in wood species (Vogel, 2008), (Shmulsky & Jones, 2011), (Rowell, et al., 2012), (Abramson, et al., 2012).Comparing our results with the other unpublished data (using the same method)in our laboratory shows that rice is in the top high lignin content and has highest range of digestibility in the studied grasses (Table 4).

Previous reports indicate that rice typically contains 1-10% silica per dry weight (Epstein, 1999), (Ma & Takahashi, 2002); (Currie & Perry, 2007). The silica content found here is studied in the stem, whereas it is reported that leaves typically contain higher level of silica (Ma & Takahashi, 2002). Recently, silica and lignin content were also measured in rice mature straw of 42 distinct rice cell wall mutants, its silica level ranged from 1.45%to 3.89%(Zhang, et al., 2015), which is rather closed to what we see from our results.

According to the result of saccharification assay, acetyl bromide assay for total lignin, and silica measurement, we can see that silicon plays an important role in the digestibility. There is a clear correlation between silica content and digestibility, which is not seen between lignin and digestibility. This has been reported and discussed in some published papers before, indicating that, silica in rice, might have more influence to the quality of straw than lignin does, and the percentage dry weight of silica is also higher than lignin content in rice straw content, see Table 5.

Table 4: Comparison of lignin content and digestibility based on amount of sugar release among different species; published and unpublished data taken from *Simon McQueen Mason' lab*.

Species	**Lignin[(1)] (%)**	**Sugar release[(2)] (nmol.mg.1hour)**
Sorghum	28	54
Spartina Marsh grass	37	5-120
Sugarbeet	8	N/A[(3)]
Miscanthus	9	5-15
Tomato	17	2-25
Hemp	21	N/A
Brachypodium	17	20
Saccharium (sugarcane)	30	37
Rice straw	**20**	**20 - 134**

(1) Lignin contents were measured by Acetyl Bromide Soluble Lignin (ABSL) method

(2) Sugar releases were measured through saccharification assay, using water for pre-treatment.

(3) N/A: not available

The physical and chemical role of silica and its effects on biomass enzymatic saccharification remain unknown, however a recent report suggests there may be the covalent crosslinks between silicon and the hemicellulose, (He, et al., 2015), and this might explain the correlation between silica content and straw digestibility. He et al, (2015) using the ICP-MS, had found an association between silicon and hemicellulose after looking separately at different cell wall component from cell wall extraction and fractionation. Recently, Zang et al, (2015) used a silicon-supplied hydroculture analysis to demonstrate that silica distinctively affects cell wall composition and major wall polymer features, including cellulose crystallinity (CrI), arabinose substitution degree (reverse Xyl/Ara) of xylans, and sinapyl alcohol (S) proportion in three rice mutants. They found that silica levels are significantly positively correlated with three major wall polymers, indicating that silica is associated with the cell wall network (Zhang, et al., 2015). These studies together with the result from my research suggest that silica content is an important feature to consider in biotechnical application of rice straw.

Table 5: Silica and lignin content of rice plant parts, barley, oat, and wheat straws: Compiled from **(Doyle & Panday, 1990)**, **(Jones & Handreck, 1967)**, **(Soest, 1970)**, and **(Soest, 1994)**.

	Polished rice	Rice bran	**Rice straw**	Rice hulls	Rice joints	Barley straw	Oat straw	Wheat straw
SiO_2(g/kg)	0.5	50	**130**	230	350	20	20–50	10–50
Lignin (g/kg)	**–**	30	**52 ±16**	160	120	110	140	85–140

Chapter 3 - Association mapping for biomass and cell wall traits

3.1 Method

3.1.1 Genotyping by Sequencing and SNP identification

Plant growth and material

172 genotypes were grown at The James Hutton Institute (JHI), Dundee for isolating DNA from young leaves. The seeds were treated Nitric acid (0.2% HNO_3) 8-10 hours at 28^0C to break dormancy, then sterilized for 12-14 hours at 30^0C and finally germinated on petri dishes after 4 days. The seedlings were then transplanted to 96 well trays first for 5-7 days before moving to the bigger pots with rice mixed compost. Plants were grown in a temperature controlled glasshouse at 25^0C-28^0C and lighting time, 12 hours daylight/ 12 hours night (Figure 22).

Figure 22: Rice plants growing in the glass house in Dundee.

DNA sample preparation

After 3 weeks, the young leaves were collected and used for DNA extraction. The leaves after being harvested from the plants were immediately frozen in liquid nitrogen before extracting DNA, using DNeasy plant mini kit (Qiagen) according to the manufacturer's instructions (www.qiagen.com/handbooks).

DNA samples were analysed by 1% agarose gel electrophoresis (run with 0.5 x TBE) to check the quality. An example gel is shown in Figure 24.

In parallel, the DNA samples were quantified and examined for purity by NanoDrop 8000 spectrophotometer. The nucleic acid concentrations were in the range of 49.7 to 190.3 (ng/µl), and the purity of DNA and nucleic acid were accessed by A 260/280 and 260/230 Ratios (Figure 23).

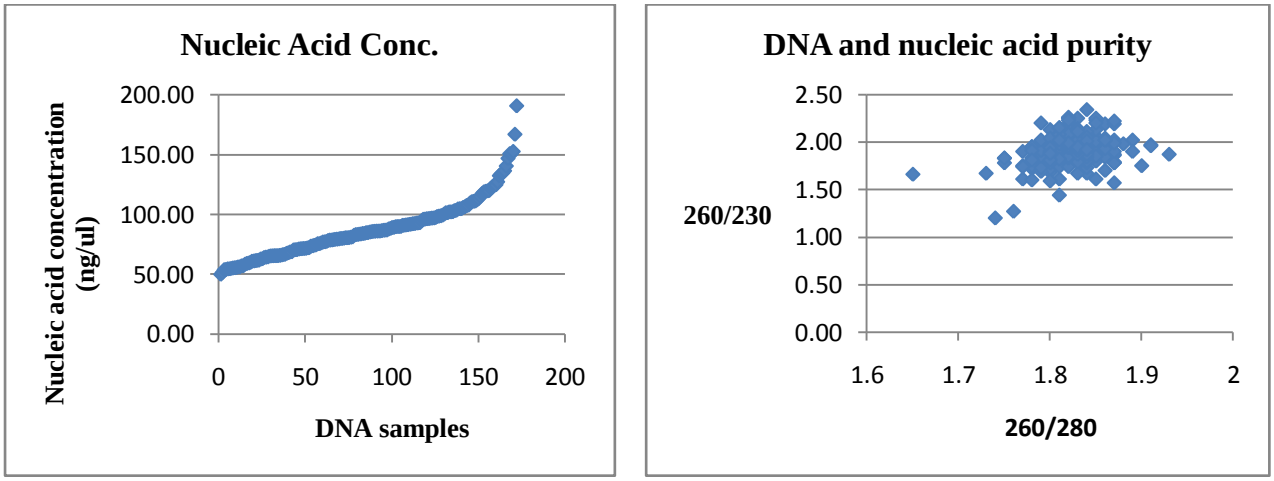


Figure 23: Nucleic acid concentration and DNA quality range at 260/280 and 260/230 Ratios

(A) (B)

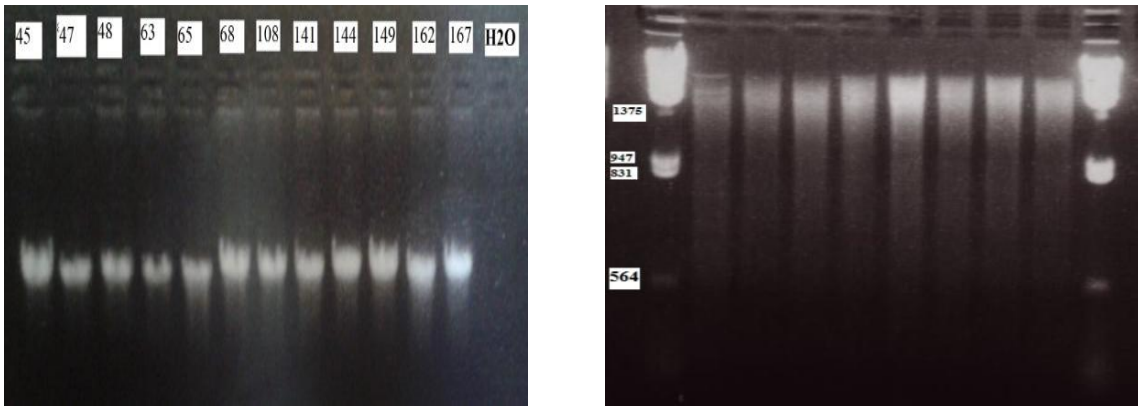


Figure 24: (A) Gel electrophoresis checking the quality of DNA extracted from young rice plants; (B) Example of digested DNA by restriction enzyme HindIII and EcoRI

Before sending the DNA samples for genotyping, I undertook trial digestion with restriction enzymes to further access the quality (Figure 24). The restriction enzyme HindIII and EcoRI were used for this experiment. The restriction digests were carried out with 10 µl of gDNA for 200 ng (20ng/µl), 0.4 µl of restriction enzyme (HindIII OR EcoRI), 3 µl of 10x NEB buffer 4, and made up to 30 µl with 16.2 µl of dH_2O. The reaction was incubated for 2 h at a temperature recommended by the manufacturer (https://www.neb.com/). Products were run on a 1% agarose gel to confirm digestion had occurred.

To process the genotyping by sequencing (GbS) assay requires at least about 100ng in 10 µl, therefore we prepared 20 ng of 30 µl of all the DNA samples in two 96 well plates and these were sent to the Sequencing Lab at Cornell University Biotechnology Resource Center (BRC), US for processing GbS.

Genotyping by sequencing assay

The Genotyping by Sequencing (GbS) assay was conducted in the McCouch Ricelab, Department of Plant Breeding & Genetics - Cornell University, who carried out library construction, sequencing, data analysis, and SNP detection from HapMap, following the methods described in (Elshire, et al., 2011).

After processing the assay, the raw data was analysed by Katie Hyma in the McCouch Lab. The GBS analysis pipeline (Tassel Version: 3.0.166 Date: April 17, 2014) was applied to analyse the data after sequencing (Glaubitz, et al., 2014). The overview of pipeline is described in Figure 26. The report of the GbS work, including all the data and results provided by Katie Hyma could be also attached as supplementary document or provided upon request. In general the analysis process of the raw data are comprised of Filtering Raw Sequence Data, DNA sequence alignments, and Mapping.

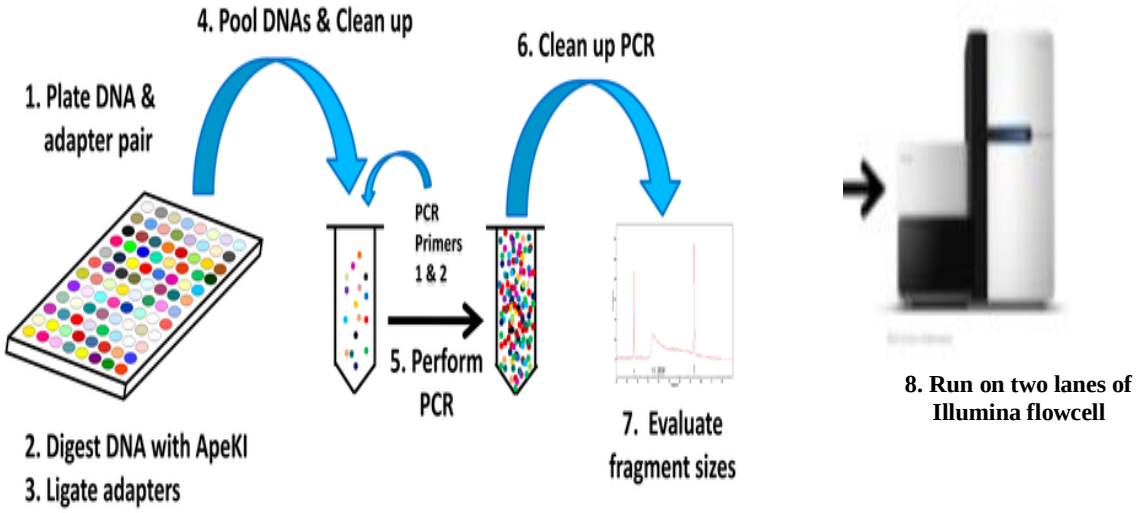


Figure 25: Steps in GBS library construction. Note: Up to 96 DNA samples can be processed simultaneously for DNA sequencing (doi:10.1371/journal.pone.0019379.g002)

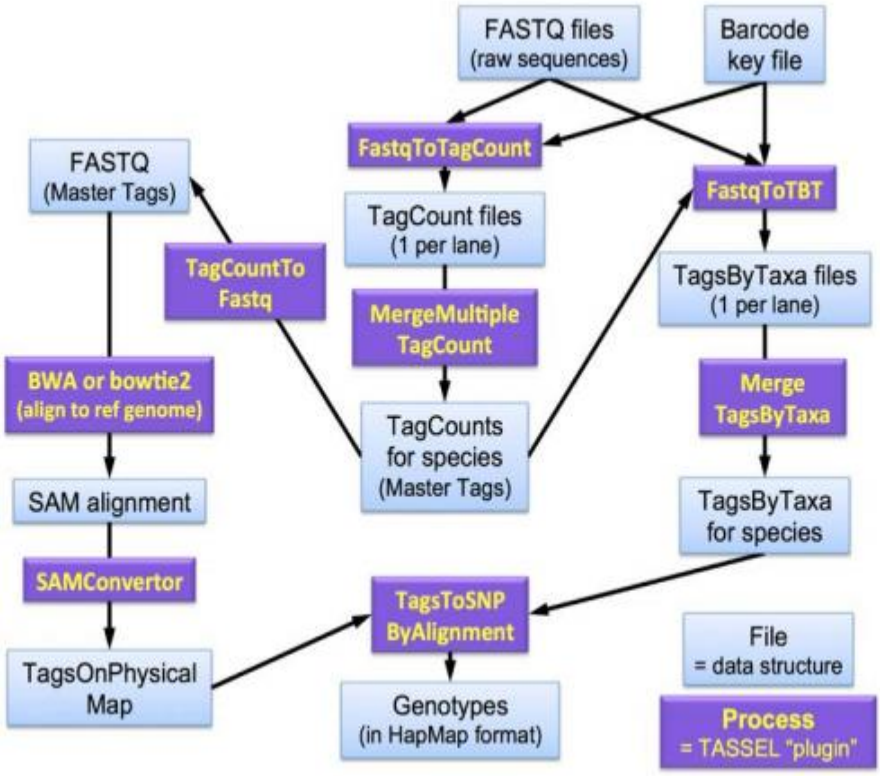


Figure 26: Flow chart showing the steps of a GBS "Discovery Pipeline" analysis link together (variations on this approach are possible). Light blue boxes represent files (or data structures) produced at each step of the analysis, and purple boxes represent the processes (Tassel3 plugins) that produced them, see the report for file description.

3.1.2 Population stratification using GAPIT

Controlling for population structure is a standard procedure in GWAS. Patterns of population structure overlap with patterns of the phenotype and with patterns of environmental variation, increasing the rate of false positives in GWAS. The genotypes used in this research were collected from many different sources and includes both *indica* and *tropical japonica* varieties. Therefore, it is necessary to check the diversity level of the population before doing GWAS, in order to know if and how much the diversity of the population would contribute to the result of GWAS.

To study stratification of the population or its structure, a phylogenetic tree was calculated and exported from GAPIT (Figure 28) (Zhang, et al., 2010); (Lipka, et al., 2012). This was done based on the kinship matrix, which accounts for the degree of genetic relatedness or coefficient of relationship between individual members of the population. Kinship among lines was calculated according to Vanraden (2008) using an R implementation (www.R-project.org) available as part of GAPIT software libraries (VanRaden, 2008); (Lipka, et al., 2012). Using resultant distances, clustering was performed in R using the internal package "hclust" with default parameters.

The codes and library used are attached below:

```
### Install all the packages and library in R program
before importing the required files for running Gapit
source("http://www.bioconductor.org/biocLite.R")
biocLite("multtest")
install.packages("gplots")
install.packages("LDheatmap")
install.packages("genetics")
install.packages("scatterplot3d")
install.packages("hclust")

library(multtest)
library(gplots)
library(LDheatmap)
library(genetics)
```

```
library(compiler) #this library is already installed in R
library("scatterplot3d")
source("http://zzlab.net/GAPIT/gapit_functions.txt")
source("http://zzlab.net/GAPIT/emma.txt")

### Import all the required files
setwd("//Biolpc2452/c$/R statistics/GWAS2015/Gapit")
myG <- read.delim("gapit_rice_snp.txt", head = FALSE)
myCV <- read.delim("Q Matrix file.txt", head = TRUE)
rice_gapit<-GAPIT(G=myG, CV=myCV, SNP.test=FALSE)
### Result in R console
[1] "---------------------- Welcome to GAPIT ---------------
-------------"
[1] "Converting genotype..."
[1] "Converting HapMap format to numerical under model of
Middle"
[1] "Perform numericalization"
[1] "Succesfuly finished converting HapMap which has bits
of 1"
[1] "Converting genotype done."
[1] "Calculating kinship..."
[1] "Number of individuals and SNPs are  146  and  328655"
[1] "Calculating kinship with VanRaden method..."
[1] "substracting P..."
[1] "Getting X'X..."
[1] "Adjusting..."
[1] "Calculating kinship with VanRaden method: done"
[1] "kinship calculated"
[1] "Creating heat map for kinship..."
[1] "Kinship heat map created"
[1] "Adding IDs to kinship..."
[1] "Writing kinship to file..."
[1] "Kinship save as file"
[1] "Kinship created!"
[1] "GAPIT.Genotype.View .Two pdf generate.successfully!"
```

3.1.3 Mixed Linear Model (MLM) using Tassel

Based on the genotypic data stored in the HapMap and the Phenotypic data collected from the analysis of saccharification (sugar released data), lignin content (% of total lignin), and silica content (% of silicon in dry weight), three GWAS were performed by merging two sets of data to examine the association between the markers and the studied trait to identify the Quantitative Trait Loci (QTL).

GWAS was performed with the compressed mixed linear model approach, which includes both fixed and random effects(Yu, et al., 2006); (Zhang, et al., 2010) carried by TASSEL (Bradbury, et al., 2007) which also implemented the Efficient Mixed-Model Association (EMMA) (Kang, et al., 2008) for performing association mapping while simultaneously correcting for relatedness and population structure.

The data were merged and manipulated in Tassel version 3.0

SNP file (Table 6), Trait file for saccharification, lignin, and silica result (Table 7), and Q Matrix file (derived from PCA analysis of population structure) (Table 8). The Q Matrix file was created by Dr Zhesi He - Bioinformatics research associate at Centre for Novel Agricultural Products (CNAP), using PSIKO (http://www.uea.ac.uk/computing/pisko) on Linux platform.

Table 6: Example SNP file - SNP are in single letter format, including ambiguity codes, and missing data have a "?" symbol

<Marker>	SNP_12345:1	SNP_54321:2	SNP_13542:3
Accession1	A	C	T
Accession2	W	A	?
Accession3	A	C	A

Table 7: Example Trait File, including sugar release (saccharification), silica, and lignin

<Trait>	Sugar release	Silica	Lignin
Accession1	59	5	21
Accession2	70	4	25
Accession3	34	2.5	18

Table 8: Example Q matrix file. The Q matrix file, derived from PCA analysis of population structure, start with <Covariate> and then on the next line <Trait> followed by names for the population clusters Q1, Q2.

<Covariate>		
<Trait>	Q1	Q2
Accession1	0.8	0.5
Accession2	0.3	0.7

3.1.4 Ferulic acid and p-coumaric acid content

Ferulic acid and p-coumaric acid content in the straw were also quantified using a protocol based on Fry's method(Fry, 1988). Briefly about this protocol, 1000ul o 1 M NAOH was added to 10mg of biomass and incubated at 250C for 16 hours with gentle mixing. The ferulate and sodium ion were then separated by adding of 100ul of 99% TFA to bring to pH 1.0. The ferulate were split up in to organic phase by addition of 0.5 ml of 1-butanol, vigorous shaking before spinning down. The ferulate in the upper organic phase was recovered in the new tubes before evaporating all the butanol in the speed vac. At last, the residues (pellets) were re-dissolved in pellet in 200 ul of 100% MeOH as original stock. Samples were run on the HPLC by Swen Langer, CNAP, the University of York. Ferulic acid and p-coumaric acid were detected and quantified with a Spectra SYSTEM® UV6000LP photo-diode array detector (Thermo Scientific), with UV–visible spectra collected at 240 – 400 nm, and analysed against a ferulic acid and p-

coumaric acid standard at 5, 25, 50, 100, 150, 200 µM mix; the sample from original stock were diluted 30 time before running because of too high concentration.

3.2 Results

3.2.1 SNP identification

Here I include some important result from the GbS pipeline analysis:

Total number of good barcoded reads in lane 1 and lane 2 are 207,579,114 and 220,050,998 respectively. Resulting tags after merging in two lanes are 25,375,094 and 18,824,732 respectively

Total number of Tags after Merging: 4,609,509

Alignment Results:

Total 4,609,509 tags

3,649,813 (79.2%) were aligned to unique positions

390,089 (8.5%) were aligned to multiple positions

569,607 (12.4%) could not be aligned

Resulting SNPs

HapMap SNPs (unfiltered): 639,891; HapMap SNPs (filtered): 328,915; Variant Called Format (VCF) SNPs: 1,631,277

VCFtools version [v0.1.11] was used to calculate Depth and Missingness from the unfiltered VCF file, taxa merged if available (Table 9).

Table 9: calculate Depth and Missingness from the unfiltered VCF file

	Mean	Median	Standard Deviation
Individual depth	7.817	7.262	2.798
Site depth	7.817	6.503	6.699
Individual missingness	0.179	0.172	0.093
Site missingness	0.179	0.034	0.263

We identified a total of 328,915 SNPs that were stored in the HapMap and used as genotypic data for the GWAS when merged with the phenotypic data (Figure 27: **Bar graph showing the distribution of identified SNP across the rice genome**), which would be the complex trait from cell wall characterisation of the same genotyped population, then the association between the markers and the trait studied would be looked for to identify the QTL.

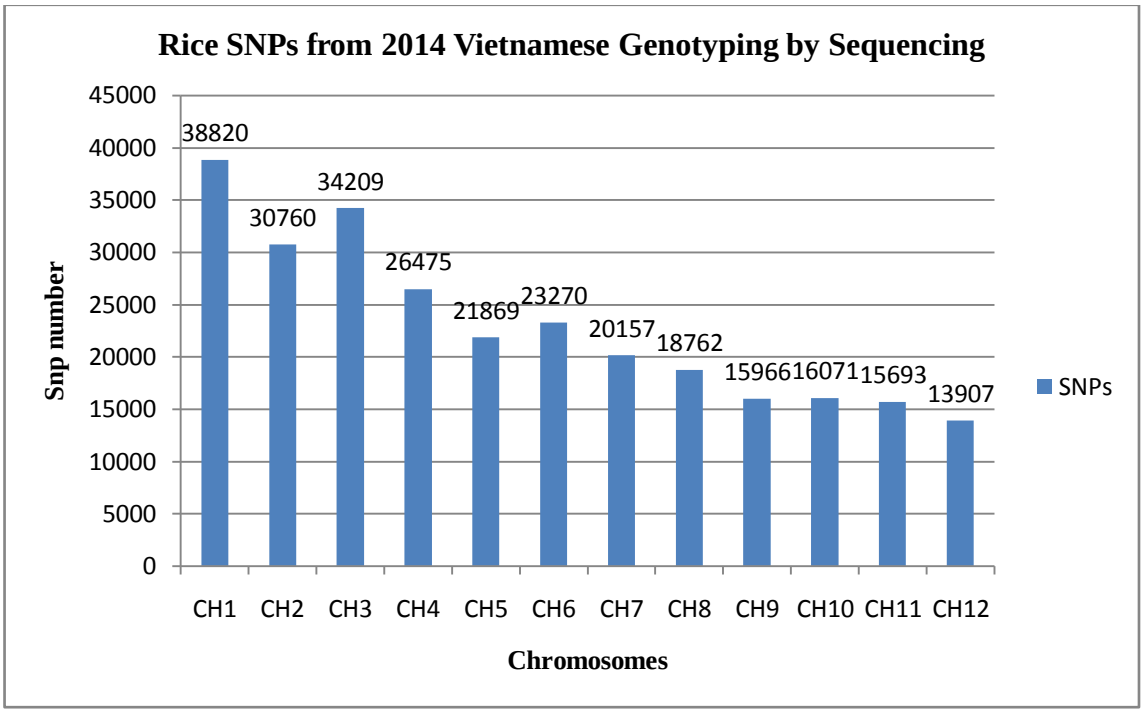


Figure 27: Bar graph showing the distribution of identified SNP across the rice genome

Based on the results of our GbS, we have the resolution at approximately 1 SNP marker at every kb (1SNP/1Kb). This represents a density level of markers across the rice genome, which will definitely help us to narrow down the QTL regions around the significant SNPs from the association map.

3.2.2 Pairwise relatedness/Population stratification

After running GAPIT, the file "GAPIT.Kin.VanRaden.pdf" was created, which is a heat map of the values in the values in the kinship matrix, showing the level of relatedness among the population**.** The results show that there are two subpopulations in our population (Figure 28). The profile indicates that the smaller subpopulation includes 22 *tropical japonica* varieties with the other comprised of 129 *indica* varieties. Separating the population in cell wall characterisation data, we saw that the range of the variation in main and sub population are rather similar in digestibility and lignin content, but the varieties from sub population (tropical japonica) seem to have lower silica content than the ones in main population (T test for P = 3.81E-11**)** (Figure 29), which is ranging about 1.4% to 4% (mean = 2.77) and 1.8% to 6 % (mean = 3.45), respectively. In term of correlation analysis, we still see a significant correlation between digestibility vs silica in either main or sub population, with R^2 = 0.166 (r = 0.407, p=1.4e-6) and R^2 = 0.3079 (r = 0.555, p < 0.00001) respectively (Figure 29). On the other hand, although we still could not see any significant correlation between lignin and digestibility in main population, we saw R^2 = 0.066 (r = 0.2569, and the p value = 0.045, just right below 0.05 which is significant point of correlation) when calculating the correlation between lignin and digestibility in the *indica* sub population.

At the end, based on the phylogenetic tree, for the next step of GWAS we decided to remove the *japonica* subpopulation in order to help to reduce the chance of confounding effects of population structure in GWAS (Brachi, et al., 2011).

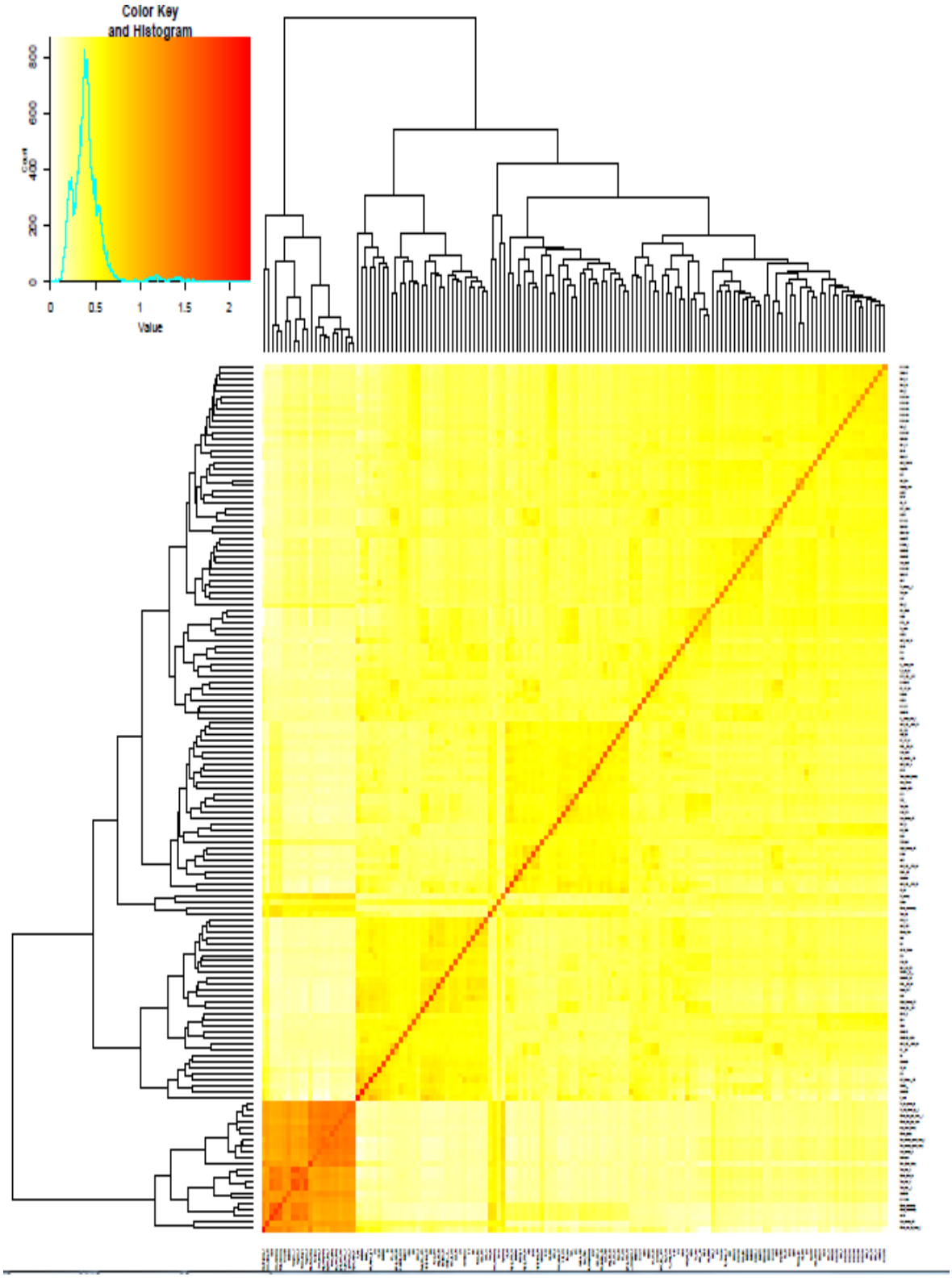
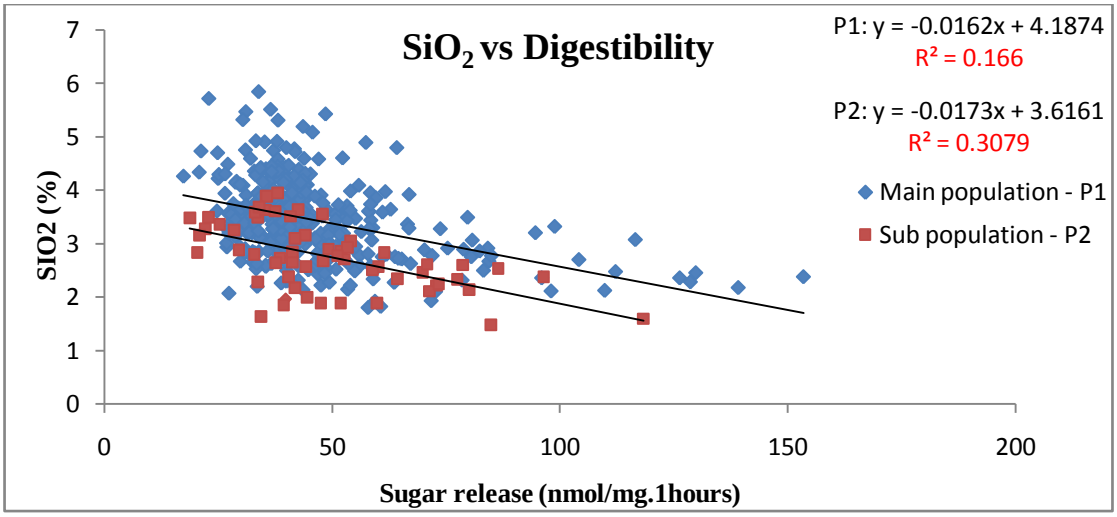
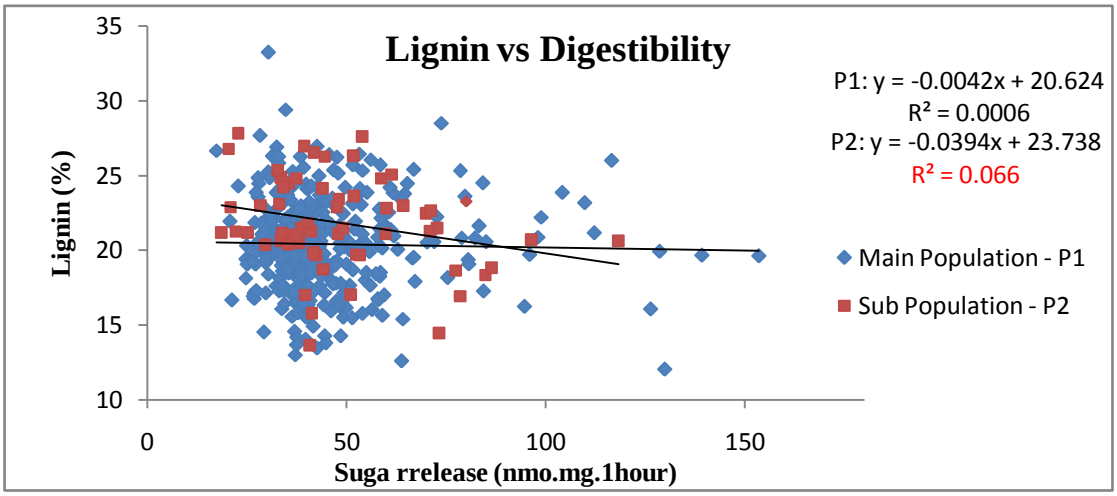


Figure 28: Phylogenetic tree in the form of kinship plot: A heat map of the values in the values in the kinship matrix. The heat map shows the level of relatedness among the population (the darker area showing highly related variety and also from different origin with the rest of the population)





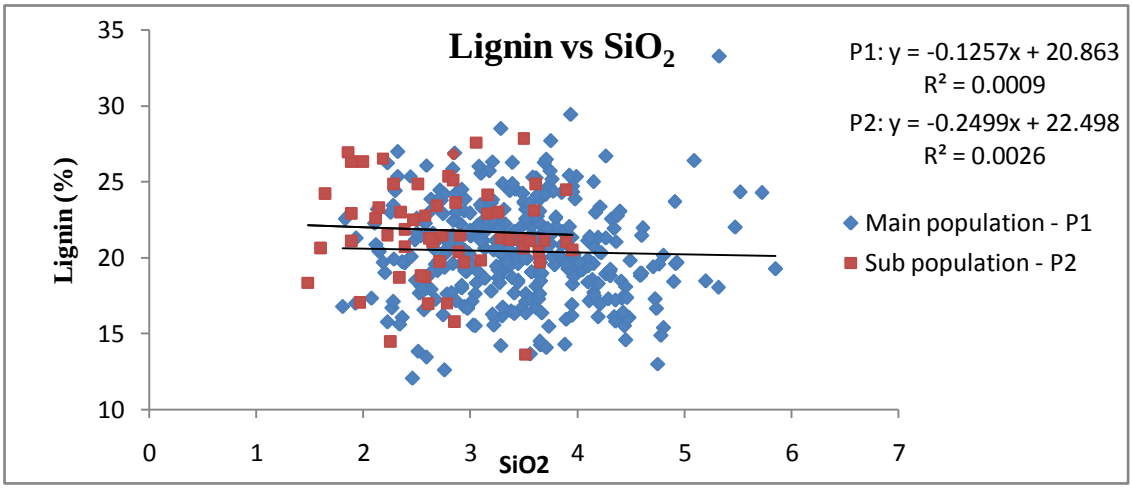


Figure 29: Correlation graph among three traits: SiO_2 vs digestibility, Lignin vs digestibility, and Lignin vs SiO_2 respectively, observing 151 line, 3 biological reps.

3.2.3 GWAS for digestibility/saccharification potential

A total of 83 significant SNPs were identified using mixed linear model (MLM) for the trait of digestibility (Table 10). The False Discovery Rate (FDR) of P-value was applied at ***FDR < 0.05*** as the cut-off for selecting significant SNPs (see Figure 30, above the red line). These significant SNPs locate in different regions in different chromosomes (CH). Genetic effects of these QTL to phenotype variance were calculated as Phenotypic variance explained (PVE) by significant SNP (see Table 10). There is a cluster of significant SNPs (a QTL) in CH1 that accounts for 37% of the phenotypic variation. There are additional SNP cluster/QTL on CH2, CH6, Ch7, CH8, and CH11 which have PVE values range from 18% (at CH2_24.6 ± 0.2 Mb) to 56% (at CH7_26.4 ± 0.4 Mb) (see Table 10).

Another GWAS, applying the same option and model (MLM) in TASSEL, was run on the digestibility data from 2013 (96 varieties overlapped with the varieties in 2014 population) merging with the same SNP data to identify the QTL and look for the common QTL regions when comparing between two year GWAS for digestibility. We identified common QTL regions, on CH2, CH6, CH7, CH8, and CH11, coming up in both year's GWAS for digestibility (see Figure 31). This helps confirm the reproducibility and reliability of the GWAS for digestibility.

To identify the candidate genes underlying the above QTL, we searched within 300 kb (±150 kb of the peak SNP) around the significantly identified loci based on the Linkage disequilibrium (LD) decay range, published in existing literature in rice(Mather, et al., 2007),(McNally, et al., 2009). MSU Rice Genome Annotation Project **(**http://rice.plantbiology.msu.edu/) database was used to search for candidates. I selected the genes as candidates in those regions based on criteria such as: cell wall related genes, expression in stems, or/and their functional studies in rice or orthologous in other species that have already been published. The list of the candidate genes is included in Table 10, in which, ***LOC_Os06g39390 (****OsAT10)* (Piston, et al., 2010) and ***Os07g49370 (****OsIRX9)***(Chiniquy, et al., 2013) are the strongest candidates as independent published worked has established a role for these in stem digestibility. The gene annotated as homologous to *OsMYB58/63*, which directly regulate *OsCESA7*(Noda, et al., 2015) is also a candidate that is worth to be more investigate.

(a) **(b)**

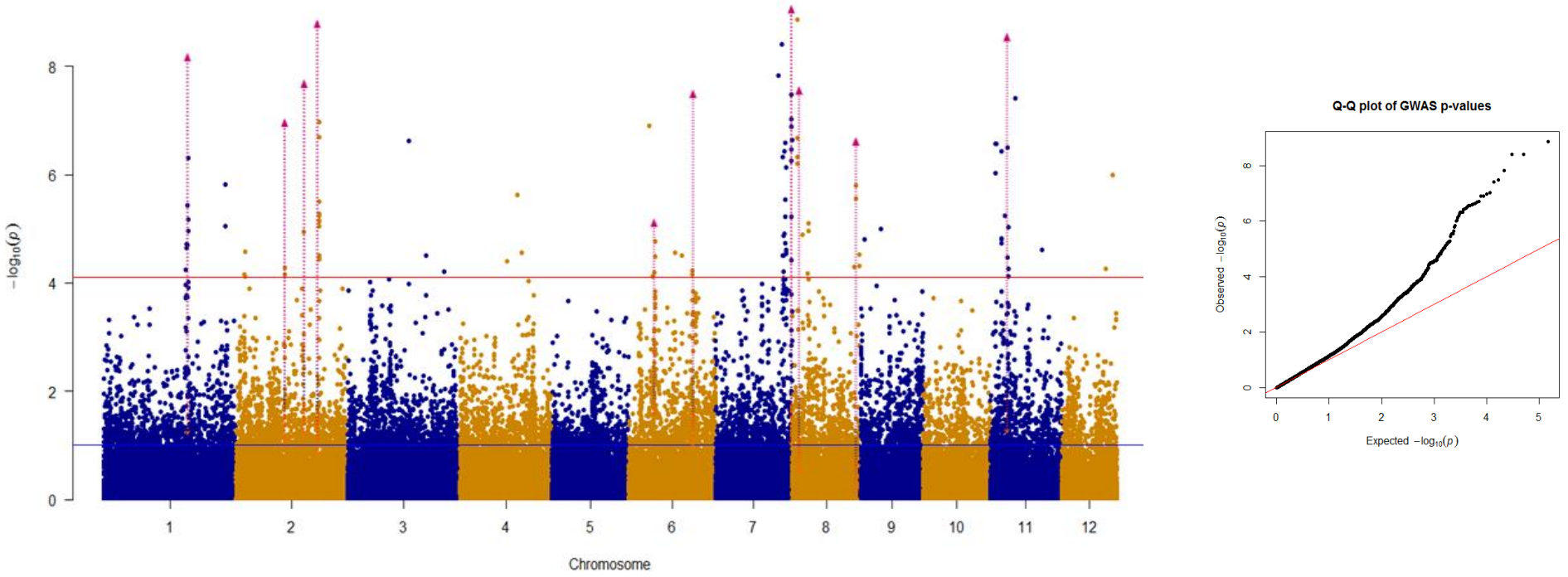


Figure 30: (a) Genome wide association study showing association between digestibility and markers across the rice genome (Significant SNPs with P<0.001; MAF>5%). **(b)** Digestibility quantile–quantile (QQ) plot determines how GWAS results compare to the expected results under the null hypothesis of no association Results above the diagonal are most likely to be significant (They have p values higher than expected by chance).

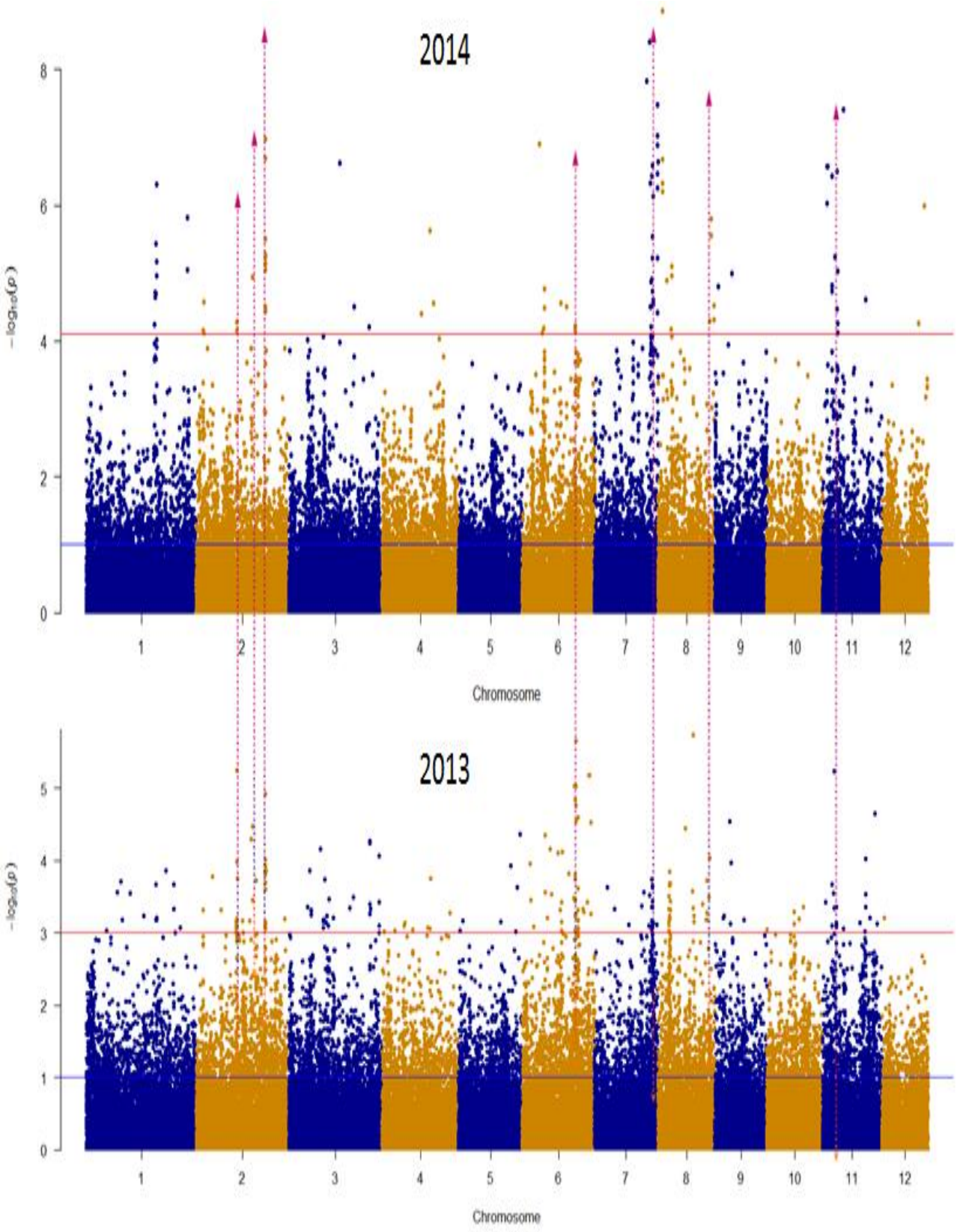


Figure 31: Genome wide association study for saccharification potential over two year studies. The red arrow indicates the common QTL.

Table 10: Digestibility QTL regions, the significant SNPs, and selected candidate genes in the QTL regions in 2014; The significant SNPs are selected by **False Discovery Rate (FDR) < 0.05**

Chromosome (CH)	**QTL regions (Mbp)**	**No of Significant SNPs in QTL regions**	**Most Significant P-Val of SNP in QTL regions**	**[(a)]MAF**	**[(b)]PVE %**	**Candidate genes**
1	CH1_29.5 ± 0.2	9	4.88488E-07	0.179	37.00	**LOC_Os01g51260** (*OsMYB26 TF*) (Guo, et al., 2014) **LOC_Os01g50720** Homologous to *BdMYB48*(Katiyar, et al., 2012)(Handakumbura, 2014)
2	CH2_19.3 ± 0.2	2	5.27001E-05	0.2	32.66	
	CH2_24.6 ± 0.4	1	1.1263E-05	0.13	18.00	**LOC_Os02g39850**(*OsHCT2*) (Kim, et al., 2012); ***http://rice.plantbiology.msu.edu/***
	CH2_28.5 ± 0.2	14	1.04809E-07	0.191	40.94	**LOC_Os02g46970** (*4CL2*) (Bruce Alberts, et al., 2002); **LOC_Os02g46780** (*OsMYB58/63 L*) (Noda, et al., 2015)
6	CH6_6.2 ± 0.2	1	1.24167E-07	0.23	46.00	
	CH6_23.3 ±	2	5.96102E-05	0.23	46.00	**LOC_Os06g39470** (BADH)

Chromosome (CH)	QTL regions (Mbp)	No of Significant SNPs in QTL regions	Most Significant P-Val of SNP in QTL regions	[(a)]MAF	[(b)]PVE %	Candidate genes
	0.2					**LOC_Os06g39390** ***(****OsAT10)* (Piston, et al., 2010)(Bartley, et al., 2013);**LOC_Os06g39970** ***(****CESA11)*(Hazen, et al., 2002)
7	CH7_26.3 ± 0.2	7	3.87549E-09	0.18	56.00	
	CH7_27.3 ± 0.2	10	2.54885E-07	0.14	30.29	
	CH7_29.4 ± 0.2	8	3.26998E-08	0.14	24.40	**Os07g49370** ***(****OsIRX9****):*** (Chiniquy, et al., 2013)
8	CH8_2.1 ± 0.2	4	1.36241E-09	0.18	42.90	
	CH8_6.7 ± 0.2	2	7.82531E-06	0.075	18.72	
	CH8_26.9 ± 0.2	2	4.95428E-05	0.088	23.74	
	CH8_27.2 ±	3	1.61358E-06	0.2	46.72	**LOC_Os08g43040** and**LOC_Os08g43020**(Orthologous

Chromosome (CH)	QTL regions (Mbp)	No of Significant SNPs in QTL regions	Most Significant P-Val of SNP in QTL regions	[(a)]MAF	[(b)]PVE %	Candidate genes
	0.2					to AT5G48930, *HCT*)
	CH8_28.0 ± 0.2	4	2.94809E-05	0.17	27.04	
11	CH11_2.3 ± 0.2	3	2.73239E-07	0.16	54.75	
	CH11_4.1 ± 0.2	4	3.65631E-07	0.2	36.12	**LOC_Os11g07960** (Orthologous to AT5G48930, *HCT*)
	CH11_5.1 ± 0.2	1	5.66175E-06	0.18	20.75	
	CH11_6.3 ± 0.2	4	3.09425E-07	0.17	27.72	

[(a)] Minimum Allele Frequency

[(b)] Phenotypic variance explained by significant SNP

3.2.4 GWAS for Silica content

For silica content significant SNPs were selected based on the cut-off at P < 0.001 and MAF > 0.05. The FDR for P value were not applied in this case because none of the SNPs could be qualified for FDR < 0.05. The clusters of significant SNPs were found in CH1, CH6, and CH11 respectively, which are also the three QTL for silica content (Figure 32). These QTL explained from 10.54% (at CH1_36.1 ± 0.3 Mb) to 15.14% (CH11_23.4± 0.5 Mb) of observed phenotypic variation (Table 12).

The QTL and cluster of significant SNPs identified in CH6 is the same one found in GWAS for digestibility. Line with contrasting alleles for the peak SNP (the marker S6_23297154) showed significant differences in term of digestibility and silica content (Figure 33), this will be examined more in a later section.

The candidate genes underlying these QTL were searched within 300 kb (±150 kb of the peak SNP) based on the Linkage disequilibrium (LD) decay range. The number of genes around the QTL regions is shown in Table 11. I looked for known genes, which already studied the function such as the silica transporters such as *Low Silicon rice 1, 2, 6* (*Lsi1*, *Lsi2*, *Lsi3*, and *Lsi6*) or their homologous. These transporters belong to the Nodulin-26-like intrinsic proteins NIP subfamily of aquaporins in rice ("facilitate transport of diverse small molecules including water and other small nutrients through biological membranes")(Park, et al., 2010), therefore any aquaporin genes might be worthy candidates for facilitating variation in silica content, but no such candidates were apparent (Table 12). Therefore, it appears that this work may lead to the identification of new genes affecting silica content in rice.

Table 11: Number of genes in QTL regions using database of MSU Rice Genome Annotation Project **http://rice.plantbiology.msu.edu/**

QTL on CH	Number of Genes in 300 Mbp around the QTL
11	143
1	84
6	130

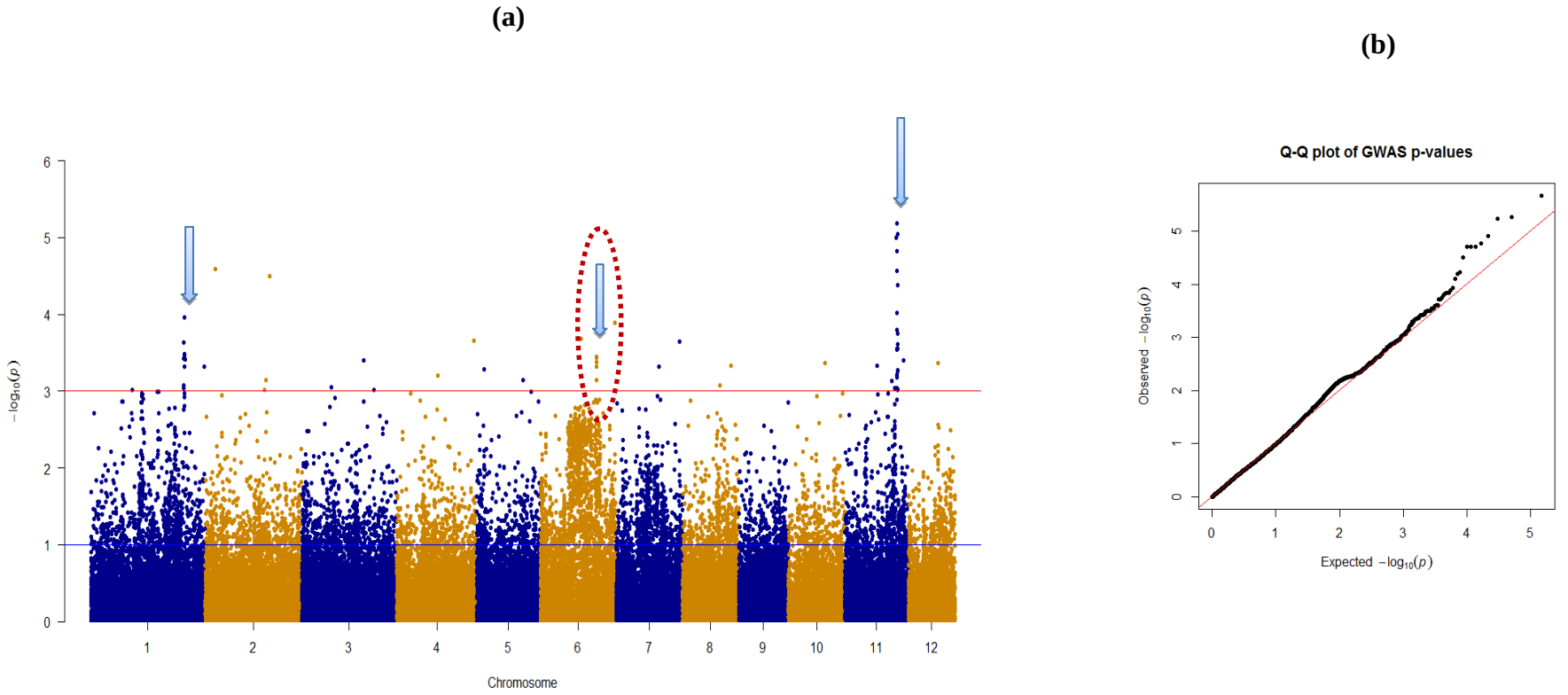


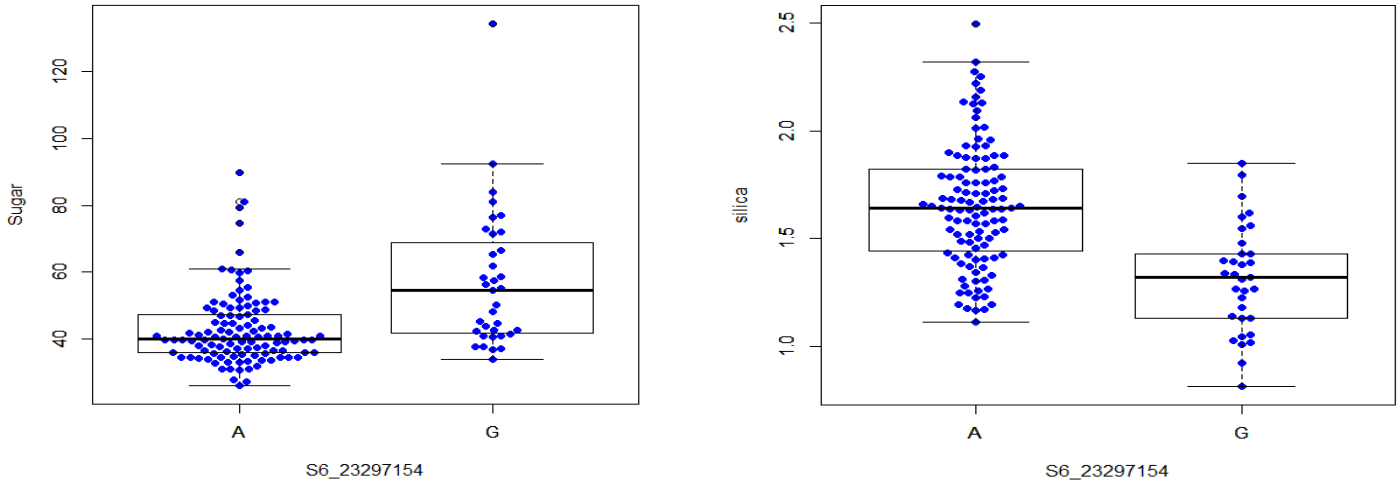
Figure 32: (a) Genome wide association study linking silica content to QTL across the rice genome (Significant SNP (P<0.001; MAF>5%). The QTL regions found in *CH6* coincides with the one in GWAS for digestibility. (b) Silica quantile–quantile (QQ) plot determines how GWAS results compare to the expected results under the null hypothesis of no association Results above the diagonal are most likely to be significant (as they have p values higher than expected by chance)

Table 12: Silica Significant SNPs (P<0.001 MAF>5%)

Chromosome (CH)	QTL regions (Mbp)	No of Significant SNPs	Most Significant P-val of SNPs	[(a)]MAF	[(b)]PVE %	Candidates genes
1	CH1_36.1 ± 0.3	3	1.00E-04	0.40	10.54	LOC_Os01g62490.1; LOC_Os01g62490.1 laccase precursor protein expressed
6	CH6_23.3± 0.4	7	2.20E-04	0.25	15.00	LOC_Os06g39390.1, LOC_Os06g39470.1: transferase family protein, putative, expressed
11	CH11_23.4± 0.5	9	1.91E-06	0.37	11.86	

[(a)] Minimum Allele Frequency

[(b)] Phenotypic variance explained by significant SNP



P - value = 3.2e-4

P -value = 8.33e-09

Figure 33: contrasting bi-allelic effect between the amount of sugar released and silica content at the SNP S6_23297154

3.2.5 GWAS for Lignin content

There were 75 significant SNPs detected based on the cut-off at P < 0.001 and MAF > 0.05. The FDR for P value were not applied in this case because none of the SNPs would be qualified for FDR < 0.05. The significant SNPs were found in CH1, CH2, CH3, CH7, CH8, CH10, and CH11 (Figure 34). These significant SNPs explain from 5.18 % (at CH10_19.2 ± 0.3 Mb) to 12.58% (at CH11_4.0 ± 0.2 Mb) of phenotypic variation (Table 13). The QTL at CH11_4.0 ± 0.2 Mb is at the same location as a QTL found on CH11 in GWAS for digestibility, although no common significant SNP were found between these two GWAS.

The candidate genes underlying lignin QTL were searched within 300 kb (±150 kb of the peak SNP) around the significantly identified loci based on the Linkage disequilibrium (LD) decay range. The list of candidate genes in the QTL regions is presented in Table 13. Several QTL regions encompass genes known to be involved in lignin biosynthesis. A Hydroxycinnamoyltransferases (HCT) gene was found in CH11 (CH11_4.0 ± 0.2 Mb), which is the common QTL region between GWAS for digestibility and Lignin. A number of peroxidase genes were found in QTL region of CH3_14.5 ± 0.4 Mb. In addition, a laccase 14 gene is located in the CH11_18.8 ± 0.3 Mb which is surrounded by many cell wall genes including a wall – associated kinase (*WAK*), kinase, and receptor-like protein kinase, and glycosyl hydrolase (Xylanase inhibitor protein).

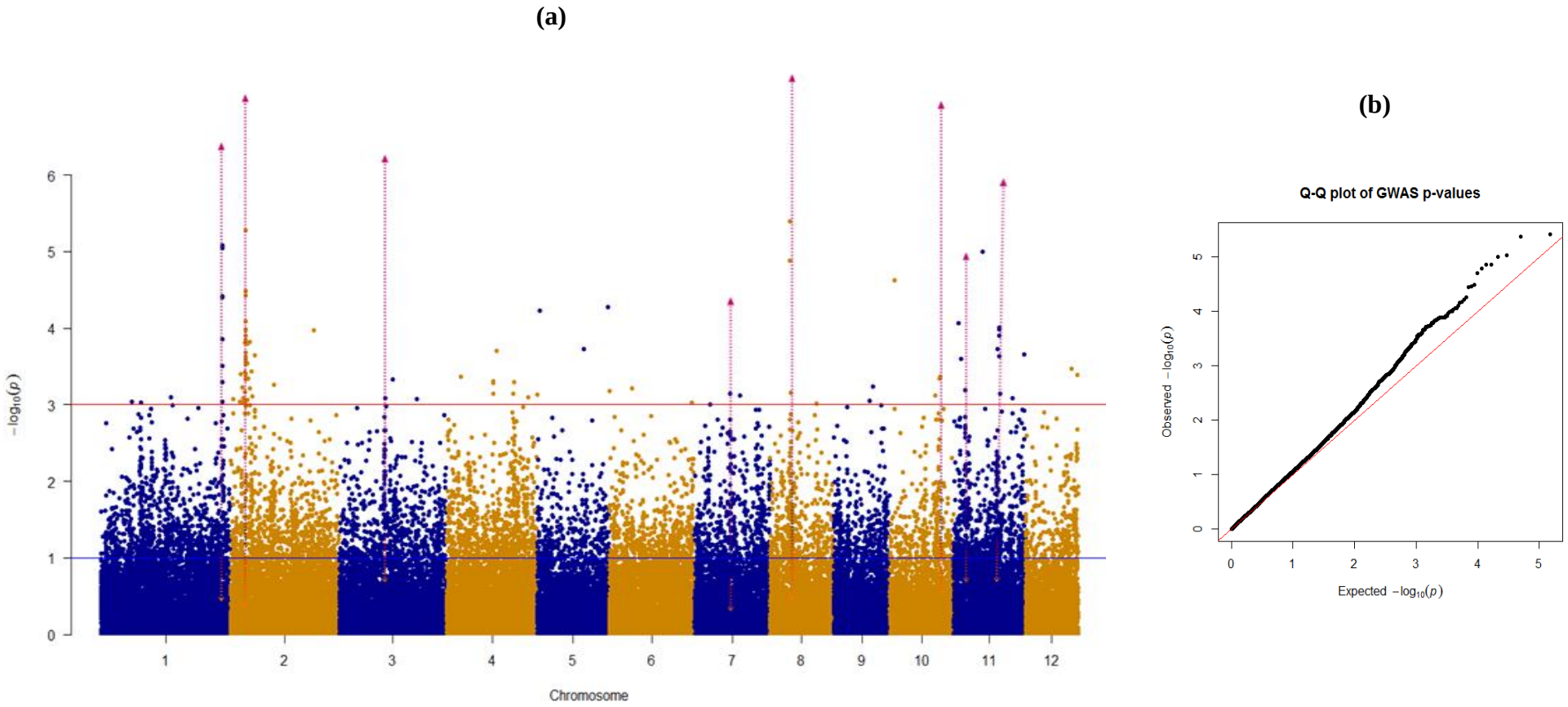


Figure 34: (a) Genome wide association study showing association between lignin content and SNP markers across the rice genome (Significant SNP (P<0.001; MAF>5%). The QTL regions found in *CH11 (HCT candidate)* coincides with the one in GWAS for digestibility. **(b)** Lignin quantile–quantile (QQ) plot determines how GWAS results compare to the expected results under the null hypothesis of no association Results above the diagonal are most likely to be significant (as they have p values higher than expected by chance).

Table 13: Lignin QTL regions, the significant SNPs, and candidates in the QTL regions in 2014; the significant SNPs are selected by P-val < 0.001 equal to Log10P-val > 3.0

Chromosome (CH)	QTL regions (Mbp)	No of Significant SNPs in QTL regions	Highest LogP-val of SNP in QTL regions	[(a)]MAF	[(b)] PVE %	Candidate genes
1	CH1_41.0 ± 0.3	8	4.86	0.24	7.09	
2	CH2_5.5 ± 0.4	24	5.37	0.47	6.94	
	CH2_6.5 ± 0.4	6	4.18	0.47	5.36	
	CH2_8.8 ± 0.2	12	3.91	0.48	5.26	
3	CH3_14.5 ± 0.4	2	3.32	0.30	6.67	7 peroxidases
7	CH7_15.8 ± 0.3	1	4.26	0.34	5.60	
8	CH8_8.8 ± 0.3	8	5.41	0.38	6.24	
10	CH10_19.2 ± 0.3	3	5.01	0.44	5.18	

Chromosome (CH)	QTL regions (Mbp)	No of Significant SNPs in QTL regions	Highest LogP-val of SNP in QTL regions	[(a)]MAF	[(b)] PVE %	Candidate genes
11	CH11_4.0 ± 0.2	2	3.13	0.17	12.58	*HCT*
	CH11_18.8 ± 0.3	9	3.88	0.38	5.84	LOC_Os11g47390.1 putative laccase 14

[(a)] Minimum Allele Frequency

[(b)] Phenotypic variance explained by significant SNP

3.3 Validation of candidate gene OsAT10 for digestibility and silica content

3.3.1 Literature review

Based on the result from searching candidates in the common QTL region of GWAS for digestibility and silica, the candidate gene LOC_Os06g39390 at CH6_23.3± 0.4 Mb had been selected as the strongest candidate. This gene and its close neighbour locus LOC_Os06g39470 both belong to family PF02458 for *GAX* feruloylation (Mitchell, et al., 2007)(Piston, et al., 2010). In 2010, Piston et al. showed that the transcript levels and ferulic acid (FA) content in the cell wall of Os06g39470 and Os06g39390 are both higher in the stems than in the leaves. In addition, down regulation of both genes by RNA interference (RNAi) construct also reduces ester-linked ferulate content in the transgenic lines (Piston, et al., 2010).

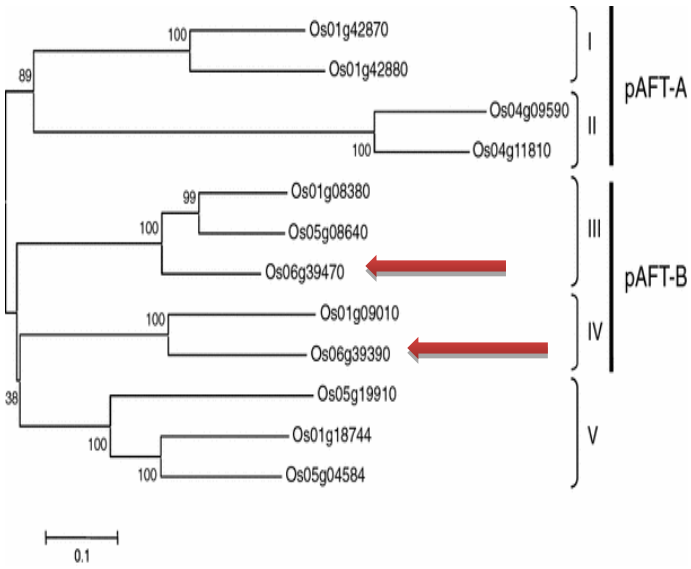


Figure 35: Neighbor-joining tree generated from the alignment of the 12 proteins from *O. sativa*PF02458 family (putative arabinoxylan feruloyl transferase) – figure taken from Piston et al., 2010.

This gene LOC_Os06g39390 has recently been named *OsAT10* which is a BAHD *acyltransferase* (Bartley, et al., 2013). Overexpression study of this gene altered cell wall hydroxycinnamic acid content, which causes 60% reduction in matrix polysaccharide-bound FA and an approximately 300% increase in p-CA in young leaf

tissue and saccharification in rice (Bartley, et al., 2013). However, no publications have described the effect of the mutant of this candidate with regards to silica content. This has raised the curiosity for us to find out if this candidate would give any effect to silica content when checking its mutant/overexpression line. We asked Bartley to send the biomass material of the *OsAT10* and its wild types to York to conduct cell wall characterisation and confirm about the effect of this candidate to cell wall and silica.

3.3.2 Cell wall characterisation of the OsAT10 overexpression lines

Studies of OsAT10 have already proposed that the gene LOC_Os06g39390 located in our QTL regions is a p-CA transferase gene, and its overexpression line helps to increase saccharification, however the silica has not been investigated in these overexpression line, therefore we asked Barley and her colleagues to send the straw over to conduct the measurement of Si using the p-XRF machine (*described in chapter 2*). In addition, I also carried out saccharification assays on these lines using the automated system at York (*method also described in chapter 2 - saccharification*). Ferulic acid and p coumaric acid content in the straw were also quantified using a protocol based on Fry's method(Fry, 1988).

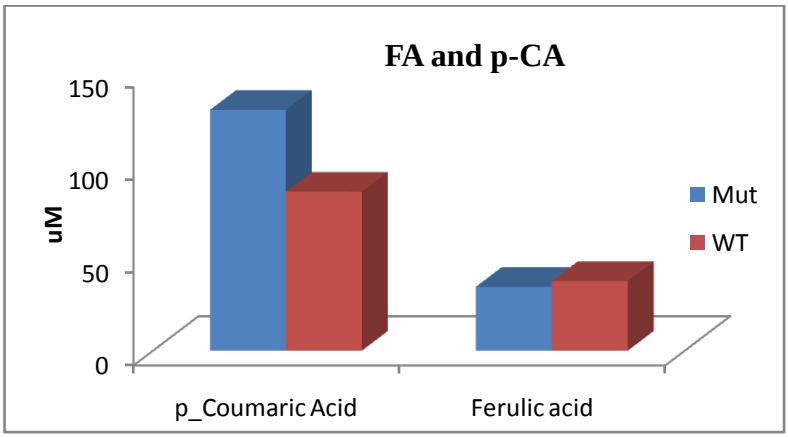
3.3.3 Result of the validation

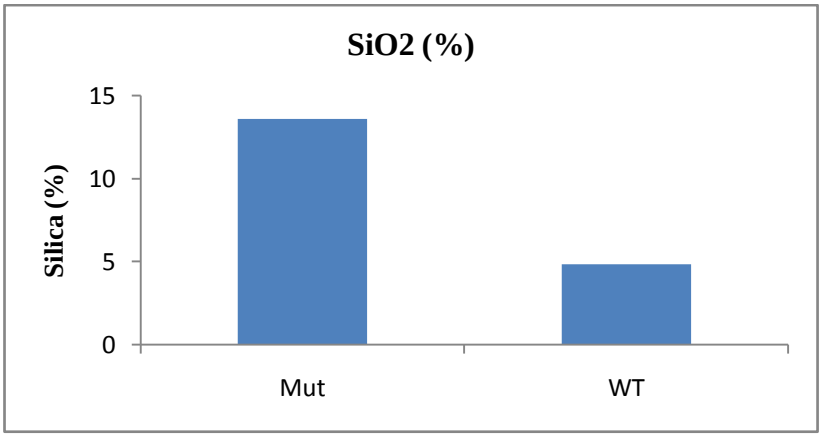
The biomass from the overexpression lines (mutants) showed 2 fold higher in digestibility than the biomass from the wild type (WT), which is about 60 nmol/mg.1hour compares with 30 nmmol/mg.1 hour (Figure 36). The silica content of the overexpressor lines was also are about 3 fold higher than the wild type, which are 14.6 % and 4.8 % respectively (Figure 36).

We also saw a higher content of p-coumaric acid in mutant (129 uM) than in the wild type (85 uM), but we could not see much difference in term of ferulic acid content (Figure 36). LOC_Os06g39470 is a gene that encoded a coumarate transferase rather than a ferulate transferase.

Overall, we could conclude that the overexpression line of LOC_Os06g39470 has effects on coumarate and silica content, as well as digestibility, supporting its potential role in determining variation in these traits around this QTL. More work needs to be

done to confirm about the link between silica and cell wall, that underlies the effects on silica content and digestibility in our population.





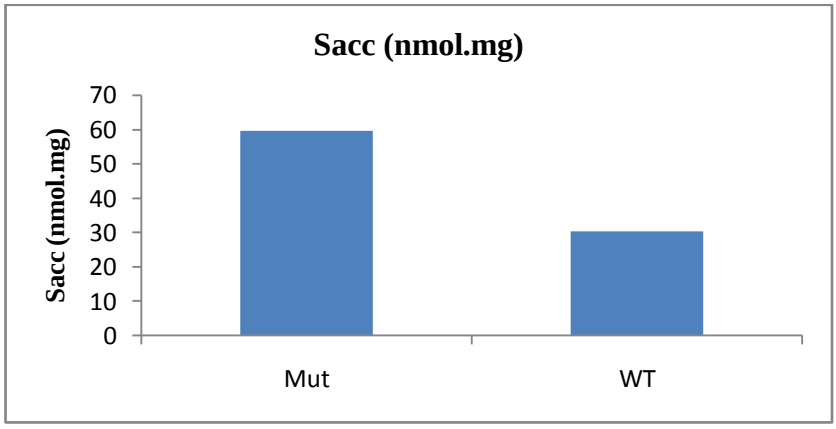


Figure 36: Measurement of FA and p-CA, silica content, and digestibility (saccharification) of the overexpression line (Mut) and wild type (WT).

3.4 Discussion

3.4.1 HapMap resolution for GWAS

Because we are interested in enabling genome wide association studies (GWAS) in rice, a species in which genome-wide linkage disequilibrium decay rates for subspecies like *indica* and *japonica* landrace are estimated at ~123 kb and ~167 kb (Huang, et al., 2010), and cultivated rice has a longer range 100 kb to over 200 kb (McNally, et al., 2009), we need to identify markers that give sufficient dense coverage such that causative polymorphisms stand a reasonable chance of being in LD with one or more markers. In our research we detected average number of 1 SNP in every kb, which is a desirable resolution of marker for association mapping, and this has helped us to identify small haplotype blocks that were significantly correlated with such complex traits as digestibility in rice straw.

3.4.2 Population stratification

In our research population, based on the pairwise studies for relatedness among the varieties, *indica* varieties were grouped into the main population and *tropical japonica* varieties were grouped into the smaller one, which is considered as the sub population. The *japonica* subpopulation was removed to reduce the chance the GWAS results were misleading by confounding factors. The false positive and false negative in GWAS can occur when patterns of population structure overlap with patterns of the phenotype and with patterns of environmental variation(Brachi, et al., 2011).

These two populations appear consistent in terms of correlation between silica content and digestibility. However, silica content was generally higher in the main population than in the sub population (Figure 29). This indicates that the *indica* varieties in our studied population have higher ability to accumulate silica than *tropical japonica* varieties. However, this study still needs to be repeated and broadened to include more varieties to confirm this. When the population was separated, we also saw a slight correlation between lignin content and digestibility in the subpopulation (Figure 29), which was not apparent in the main population and the whole population. The lignin effect on the cell wall in general and its impact on digestibility in grasses and cereal

have been recognised for a while(Zeng, et al., 2014). Other authors have suggested a role for silica in determining digestibility in some grasses, especially rice (Van Soest, 2006). It seems likely that in rice, the effects of lignin are perhaps masked in varieties that accumulate higher levels of silica, suggesting a potential redundant function.

3.4.3 QTL and candidates gene selection

One striking result is the common QTL for both digestibility and silica content on CH6, which showed a contrasting bi-allelic effect between the amount of sugar released and silica content (Figure 33). The QTL on CH11 also appears in both GWAS for digestibility and lignin content, although it is not clearly showing the bi-allelic effect between these traits. This together with the phenotypic correlation shows that, in rice, silica might be playing a more important role in straw digestibility than lignin, and this may contrast with other cereal species which have generally lower levels of silica accumulation.

Digestibility

Seven common digestibility QTL found in two separate season's data. Among these, three of the searched regions encompass candidate genes previously shown to effect stem digestibility. The first one, LOC_Os06g39390 (*OsAT10*) a p-coumaroyl coenzyme A transferase (BADH family) involved in glucuronoarabinoxylan modification altering rice cell wall hydroxycinnamic acid content and saccharification (Piston, et al., 2010); (Bartley, et al., 2013), The second, LOC_Os07g49370 (*OsIRX9)*plays the function in building the xylan backbone in the secondary and primary cell walls. The expression of *OsIRX9* in the *irx9* mutant resulted in XylT activity of stems that was over double that of wild type plants, and the stem strength of this line increased to 124% above that of wild type (Chiniquy, et al., 2013). The third is locus LOC_Os02g46780 next to the SNP - S2_28582605 (p = 1.05E-07), identified as *OsMYB58/63 L* (Hirano, et al., 2013), which is a homologous to *OsMYB58/63* which was found to directly up-regulate the expression of a rice secondary wall-specific cellulose synthase gene, cellulose synthase A7 (*OsCesA7*) (Noda, et al., 2015).

The significant SNP on CH1, S1_29469728 (p = 3.65E-06), was located inside the locus LOC_Os01g51260, which is known as the transcription factor *OsMYB26 TF*(Guo, et al., 2014). This *OsMYB26 TF* (LOC_Os01g51260) corresponds to the Arabidopsis *MYB TF* AT3G13890 (also known as *AtMYB26*), which has been identified as an activator of secondary wall thickening (Yang, et al., 2007). Another locus LOC_Os01g50720 on CH1, 50kb from S1_29188922 (p = 2.29E-05), is homologous to *BdMYB48*(Katiyar, et al., 2012), which is a grass specific activator capable of regulating secondary cell wall biosynthesis; over expression of *BdMYB48* showed upregulation level of secondary cell wall biosynthesis genes (Handakumbura, 2014).

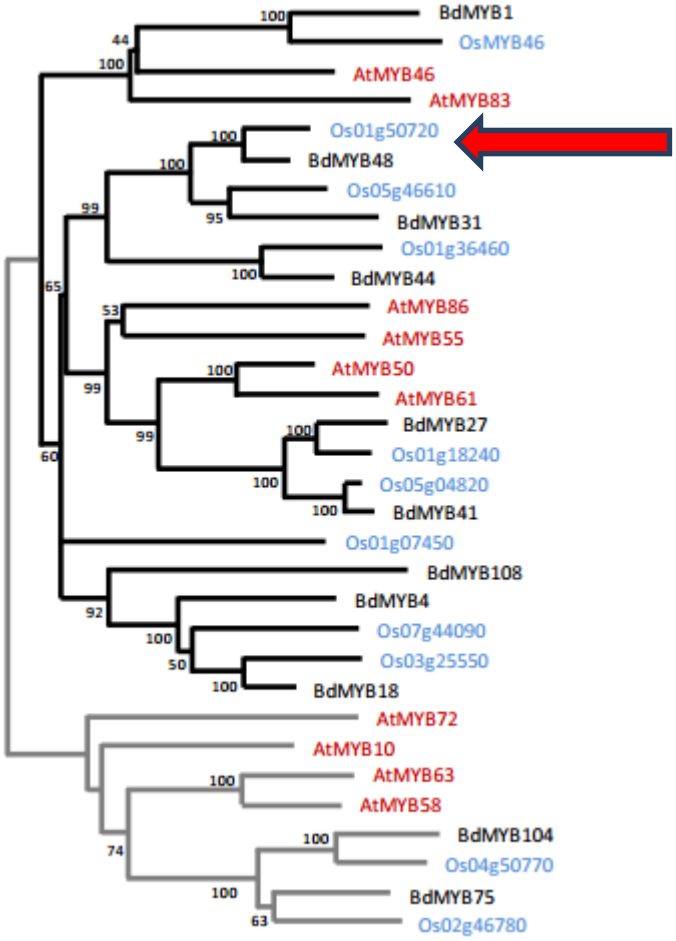


Figure 37: Phylogenetic tree of protein sequence similarity for grass specific MYB clade for A. thaliana, B. distachyon and rice. (Handakumbura, 2014). The red arrow points out the LOC_Os01g50720 and BdMYB48.

On CH2, LOC_Os02g39850, located at 400 kb from S2_24691620 (p = 1.13E-05), is the *OsHCT2*(Kim, et al., 2012). LOC_Os02g39850, which encodes a gene, is highly homologous to and clustering with Arabidopsis *HCT* (AT5G48930) as out-group in extended Mitchell Clade of BAHD Acyltransferases in grasses (Bartley, et al., 2013). By exploiting the substrate flexibility of this At *HCT*, Eudes et al., succeeded to reduce lignin content and improve biomass saccharification by engineering transgenic lines that overproduce one of the *HCT* non-canonical acceptors(Eudes, et al., 2016). The LOC_Os02g46970 found near with S2_28615156 (p = 3.08E-06) is a 4-Coumarate: coenzyme A ligase (*4CL*) catalysing the conversion of hydroxycinnamates into corresponding CoA esters. However Sun et al., pointed out that this gene was an *Os4CL2* specifically expressed in the anther and was strongly activated by UV irradiation, suggesting its potential involvement in flavonoid formation(Bruce Alberts, et al., 2002).

On Ch6, I found a potential candidate, LOC_Os06g39970, a cellulose synthase catalytic subunit**(***OsCESA11)*(Hazen, et al., 2002). However, at the time of writing, there is not any published literature on the function of this gene in rice.

On Ch8, two transferase family protein Loci, LOC_Os08g43040 and LOC_Os08g43020 were found next to the SNP - S8_27288702 (p = 2.78E-06). The results of blasting the protein sequence of these two genes in rice genome database (http://rice.plantbiology.msu.edu/analyses_search_blast.shtml) shows that they have high protein sequence similarity to the four *OsHCT 1-4*. They are also orthologous to AT5G48930, only *AtHCT* gene known in Arabidopsis. However, their functions have not been described at the time of writing.

A transferase protein family, LOC_Os11g07960, was found in the QTL region, CH11_4.0 ± 0.2, over two seasons of GWAS. This gene shows high protein sequence homolog to the four *HCT* genes found in the rice genome, including the two found in other QTL regions described above, LOC_Os08g43040 and LOC_Os08g43020. Although this candidate also has not been published on, it should be considered as another potential candidate for further functional studies.

Silica

So far, there are just a few genes that have been identified as involved in silicon uptake in rice, which are the silicon transporters: *Lsi1*, *Lsi2*, *Lsi3*, and *Lsi6*(Ma, et al., 2006),(Ma, et al., 2007), (Yamaji, et al., 2015). The other molecular mechanisms involved in silica accumulation are poorly understood. Based on the knowledge about those silicon transporters and their homologous in the Nodulin-26-like intrinsic proteins NIP subfamily of aquaporins in rice, I was trying to look for if any of those appeared in the QTL regions that we identified from Silica GWAS, but no clear candidates could be seen. I have selected the regions to look for the candidates within 300 kb (±150 kb of the peak SNP) around the significantly identified loci based on the Linkage disequilibrium (LD) decay range. There were 84, 130, and 143 genes located in the QTL regions in CH1, CH6, and CH11 respectively. This means that there might be potentially novel silica -related genes in these QTL regions. Interestingly, we found four genes that located closely or next to the peaks of CH1 and CH6. On CH1, LOC_Os01g62480.1 and LOC_Os01g62490.1, encode two laccase family proteins. Plant laccases are involved in phenolic compounds oxidation and it has been thought that peroxidases and laccase both play the role in catalysing the oxidation of monolignol (Wang, et al., 2015). On CH6, LOC_Os06g39390.1 and LOC_Os06g39470.1 encode two transferase family proteins related to arabinoxylan biosynthesis in this plant (these two transferases were also the candidates on CH6 in GWAS for digestibility). Arabinoxylans make up the major hemicellulose in rice straw. Recently, based on looking at cell wall fractions studies, He et al., found an apparent covalent association between Si and hemicellulose in rice cell wall as determined by inductively coupled plasma-mass spectrometry (ICP-MS) (He, et al., 2015). Therefore, the two transferases we found here might not only have effect on digestibility but also have effect on the silica accumulation due to the hypothetical link between silica and hemicellulose.

Lignin

I scrutinised the lignin QTL regions to look for the cell wall related genes, especially those involved in the monolignol biosynthesis or the polymerisation process of lignin formation. There is a *Hydroxycinnamoyltransferases* (*HCT*) gene on CH11 (CH11_4.0 ± 0.2 Mb), which is the common QTL region between GWAS for digestibility and

lignin content. This candidate is highly homologous with four *OsHCT* 1-4 and the other two loci, LOC_Os08g43040 and LOC_Os08g43020, found in digestibility QTL region in CH8. Reduced expression of *HCT* in Alfalfa has been shown to increase stem digestibility, but also lead to dwarfing (Chen & Dixon, 2007).

In addition, there were cluster of 7 peroxidase genes located right next to the peak in the QTL region on CH3_14.5 ± 0.4, and a laccase located in QTL region CH11_18.8 ± 0.3, LOC_Os11g47390.1. Peroxidases together with laccases have been long proposed to perform the polymerization of monolignols into lignin (Marie, et al., 1998). Down regulation or disruption of these enzymes led to the reduction of lignin content in plants (Blee, et al., 2003); (Berthet, et al., 2011); (Zhao, et al., 2013). These could be potential candidate related to the lignin biosynthesis in rice cell wall.

OsAT10 plays its role in both silica and digestibility

The availability of straw from the *OsAT10* overexpression lines and its wild types allowed me to characterize these for digestibility, ferulic and coumaric acid, and silica content. The overexpressor lines showed a 2 and 3 fold difference in both digestibility and silica content, compared to wild type. Barley et al (2013), published that the *OsAT10* had 300% increase in p-CA and 60% reduction in matrix polysaccharide-bound FA in young leaf; in my assay, we measured the FA and p-CA content in the straw and the result showed 51.7% increase in p-CA and 8% decrease in FA. . These results confirmed that *OsAT10* is a p-coumaroyl coenzyme A transferase involved in glucuronoarabinoxylan modification. Our results raise a potentially important question; if *OsAT10* is responsible for the variation in silica seen at this QTL, how does it do this if its primary function is arabinoxylan coumaroylation. The addition of feruloyl and coumaroyl esters to arabinoxylan appears to take place prior during polysaccharide biosynthesis and prior to secretion(Buanafina, 2009). One possibility is that a bond is formed between ester linked coumaric acid and silica, and this could potentially occur intracellularly or post polysaccharide secretion. Although there are no reports of such a linkage, a number of studies have indicated that covalent or other association may exist between silica and hemicellulose polymers in rice(Inanaga & Okasaka, 1995);(Inanaga, et al., 1995); (He, et al., 2015) . It seems that this observation of a potential role for *OsAT10* in determining cell wall silica content may open the way for a deeper

understanding of the role of silica in grass cell walls, and this appears a potentially fruitful area for further study.

Chapter 4 – Final Discussion

Rice straw as a research model and potential feedstock for biofuel production

The use of crop residue biomass provides a more suitable route for biofuel production than do first generation processes that compete with food for feedstock. Because rice straw is an abundantly available and globally underutilised resource, it provides an attractive feedstock for biofuel and bio-refining (Binod, et al., 2010). However, to take full advantage of this resource, we need to improve its saccharification potential and make it more easily digestible with industrial enzymes in order to allow the production of cost-competitive sustainable biofuels by fermentation. The research in this thesis focuses on rice biomass, from a diversity panel assembled from rice germplasms in Vietnam, which is one of the top four-rice exporting countries the in the world(Workman, 2016). Using rice straw as feedstock for biofuel production will also help to reduce the pollution from straw burning, as almost 90 % of this is burned in the field. In addition, rice is a model cereal with a small sized diploid genome (~430 Mb), well-developed molecular genetics tools, and has representative cell wall characteristics of grasses, making it an important science discovery tool (Yuan, et al., 2001). This is important because our understanding of the biosynthetic gene machinery and molecular structure and biosynthetic gene of plant cell walls remain incomplete and the molecular basis of biomass digestibility even more so. Rice is also notable for the high levels of silica it accumulates, and the deposition of silica and the linkage with other cell wall components is only now beginning to emerge(Zhang, et al., 2015)(He, et al., 2015).

It is generally accepted that lignin plays an important role in determining cell wall digestibility in plants, and that silica may also have a role in conferring recalcitrance in rice. The basis of the research presented in this thesis was to use a GWAS approach in rice to identify QTL for straw digestibility, lignin content, and silica content. Studies were undertaken to provide a high density SNP map for 172 rice cultivars that were then phenotyped for straw digestibility and silica content, and a much weaker association between lignin and digestibility. Using the results of this work, I was able to identify a number of QTL for all three parameters tested and proposed a number of candidate genes associated with some of these QTL.

Silica has an important role in rice biomass digestibility.

In the 2014 samples, the results show that higher variation is apparent in digestibility with lower variation in silica and lignin content. These levels of variation among the varieties in the population allow genome wide association studies to identify the QTL controlling the traits studied. We found a significant correlation between silica content and digestibility, which was not seen between lignin and digestibility. This indicates that, in rice, silica has a greater impact on digestibility than lignin does, and this agrees with the published observation that in general rice has low lignin and high silica compared with other cereal straws (Van Soest, 2006). A negative association between silica and digestibility has been reported in previous studies (Khush, et al., 1987); (Balasta, et al., 1989); (Hasan, et al., 1993); (Enishi, 2002). Hasan et al. (1993) compared straws of differing lodging resistance relative to silica contents and digestibility and found that silica content in the culm was higher in resistant types and lower in susceptible types. Correlation between silica and digestibility across all the varieties was −0.664 (Hasan, et al., 1993)

Our studies identified a common QTL shared between digestibility and silica content, and we identified OsAT10 as a potential candidate gene that might be responsible for the variation associated with this QTL. OsAT10 has previously been shown to play a role in determining rice straw digestibility through its action as a coumaroyl transferase responsible for the coumaroylation of arabinoxylans in rice (Bartley, et al., 2013). I examined the silica content of OsAT10 overexpression lines and found that they also exhibit higher levels of Si. This strongly suggests a connection between arabinoxylan coumaroylation and silica in rice cell walls. Interestingly, there is a small number of published reports that indicate that there may be chemical bonds between rice hemicellulose and silica (Inanaga & Okasaka, 1995);(Inanaga, et al., 1995); (He, et al., 2015);(Zhang, et al., 2015). However, none of these publications was able to directly demonstrate the nature of chemical link between hemicellulose and silica.

U17, a commercial cultivar, may be a good feedstock for biofuel production.

Based on the overall cell wall characterisation, the accession U17 showed outstanding straw quality with the highest digestibility in both years, and low content in both silica and lignin in the 2014 samples. This is a commercial cultivar and has been well-known for lodging and submergence resistance. This curious observation runs contrary to the association between silica, digestibility and lodging susceptibility observed by Hasan et al. (1993). However, the fact that U17 is known for lodging resistance and has low silica and high digestibility suggests that it may be possible to develop rice varieties that produce good yield, and also highly digestible straw in future breeding programs.

GWAS is a powerful tool to identify QTL for complex traits such as digestibility, silica content and lignin content

Recently, many studies have taken a reverse genetic approach, modifying the cell wall through alteration of specific cell wall synthesis genes to increase lignocellulose digestibility in *Arabidopsis*(Berthet, et al., 2011)*;* (Brown, et al., 2011)*;* (Lee, et al., 2012)*;* (Goujon, et al., 2003)*;* (Acker, et al., 2013).This thesis aimed to use a forward genetic GWAS approach to identify QTL, linked to genes that affect the digestibility of rice straw. GWAS provides a powerful alternative to traditional linkage analysis based on recombinant inbred mapping populations, which can overcome limitations of pedigree based QTL mapping because of its higher mapping resolution, reduced time, and greater number of alleles available(Oraguzie, et al., 2007); (Zhu, et al., 2008). To achieve the association studies we used a diversity panel of rice accessions from Vietnam, characterising these for digestibility, silica content and lignin content and well as producing a high density map for these lines. After this the phenotypic data were merged with a panel of SNP markers to determine any bi-allelic marker in linkage disequilibrium with the studied trait(Gupta, et al., 2005).

GWAS for digestibility: The saccharification assays were performed on the straw samples collected over two different growing seasons (spring and summer) in two years (2013 and 2014), using a high throughput platform (Gomez, et al., 2011). Little correlation was seen between data sets over two years. The heritability was calculated at

0.79 (2013) and 0.65 (2014) for the amount of sugar released. Interestingly, there were only 8 lines in the top of 25% for digestibility in the 2013 experiment, which were found in the top 25% in 2014. This might be because of environmental effects on the population and different day length requirement for different varieties (Vergara, 1985), (Krishnan, et al., 2011). The effect of the environment, including water, temperature, light, and atmosphere to the growth, developmental morphology, and biomass yield in many plant species has been long realized (Haferkamp, 1988), (Ventura, et al., 2011), (Hatfield & Prueger, 2015), (Aurangzaib, 2015). Despite this lack of correlation, it proved possible to identify seven QTL that were common for both years of analysis, when the data was analysed separately. This indicates a strong level of consistency at the population level that was not apparent in a simple correlation analysis.

There has been also number of studies using association mapping to look for QTL for digestibility in different types of plant biomass. Only a few candidate genes have been identified and validated from association mapping for saccharification so far. In alfalfa, 20 simple sequence repeat (SSR) markers were predicted to be associated with fiber-related quality traits (Heritability, H^2 = 45 to 73.6); no specific candidates gene were reported but their finding helped to facilitate marker assisted breeding programmers for introgression of alleles into locally well adapted germplasm (Wang, et al., 2016). In sorghum, by screening 703 SSR markers against the low and high saccharification (glucose release by cellulase) pools identified two markers on the sorghum chromosomes 2 (23-1062) and 4 (74-508c) associated with saccharification yield; these marker were physically close to genes encoding plant cell wall synthesis enzymes such as xyloglucan fucosyltransferase (149 kb from 74-508c) and UDP-D-glucose 4-epimerase (46 kb from 23-1062) (Wang, et al., 2011). In Maize, a GWAS for lignin abundance and sugar yield of the 282- member maize Association Panel provided candidate genes in the eleven QTL, using pyrolysis molecular-beam mass spectrometry to establish stem lignin content and an enzymatic hydrolysis assay to measure glucose and xylose yield (Penning, et al., 2014).

In my research, by looking at the regions around the significant SNPs in seven common and 3 uncommon QTL, we identified 12 candidate genes, which included transcription factors, *OsMYB26 TF, OsMYB58/63 L,* and the ortholog of *BdMYB48, OsHCT2* and three homologs of *HCT, Os4CL2, OsCESA11, OsAT8* and *OsAT10* (BAHD family), and

OsIRX9 (GT43). In this list, the *OsAT10, OsIRX9*, and *OsMYB58/63L* were considered to be the strongest candidates, which locate in the common QTL over two years. The function of OsAT10 has been already studied by characterising its overexpression lines and has been proposed to be a coumaryl transferase due to the dramatic increase in coumaric acid content in the mutant plants (Bartley, et al., 2013). The overexpression of OsIRX9 in the well-characterized Arabidopsis irregular xylem (*irx9*) mutant increased stem strength and XylT activity and proved its role in xylan biosynthesis (Chiniquy, et al., 2013). *The OsMYB58/63L* was found to be a transcriptional activators in yeast cells, and its expression level was high in culm internodes and nodes(Noda, et al., 2015). A close homolog of this candidate, *OsMYB58/63*, had already been shown directly regulating rice secondary wall-specific cellulose synthase gene, *OsCesA7*(Noda, et al., 2015). The fact that some of our candidate genes have previously been implicated in roles in biomass digestibility gives support to validate our results.

GWAS for Silica content: Unfortunately, there was insufficient straw from 2013 to allow the measurement of lignin and silica to be made. It was proposed that genotypic differences in Si concentration was associated with ecotypes (subspecies), implying that the Si uptake ability might be different between *japonica* and *indica*(Derena, et al., 1992). Our rice research accessions included both *indica* and *tropical japonica* subspecies; and *tropical japonica* varieties seem to have lower silica content than *indica* varieties (T test for *P = 3.81E-11***)**. This result is opposite to what was found before, that *japonica* rice had a higher Si concentration than *indica* rice in Si-deficient soil conditions(Winslow, et al., 1997). We decided to remove the *japonica* subpopulation in order to help reducing the chance of confounding effects of population structure in GWAS (Brachi, et al., 2011). The results of silica quantification between studies are also hard to compare when the organs used to measure are different in each study(hull, root, stem, or leaf). It was realized that silica is mostly deposited in leaves and hulls (Ma & Takahashi, 2002). In addition, the Si concentration in the soil is also one of the major factors responsible for the variation in silica content for the same genotype (Ma & Takahashi, 2002)(Ma, 2004). On the other hand, different quantification methods lead to the variation in silica content. There are several methods that have been commonly applied to determine Si: Alkaline fusion or acid digestion of the plant material (Masson, et al., 2007), spectrometric analyses of the obtained filtrate, using atomic absorption spectrometry (AAS) (Hauptkorn, et al., 2001), inductively coupled plasma spectrometry

(ICP) (Molinero, et al., 1998), or colorimetric techniques (Fox, et al., 1969); (Allen, 1989). These methods are all based on the total destruction of the plant matrix, a process that can lead to element losses due to incomplete solubilization and, particularly in the case of Si, volatilization, reducing the accuracy of the determination(Baffi C, 2002); (Reidinger, et al., 2012). In our research, we used a portable X-ray fluorescence spectrometer (P-XRF) (Reidinger, et al., 2012) to rapidly, and safely analyse the Si content, producing a very little bias. The repeatability and precision of the measurements is as good as or better than that of other methods.

The phenotypic differences in Si concentration in rice are controlled by multiple genes(Majumder, et al., 1985); (Wu, et al., 2006). In my research, the GWAS for silica discovered three QTL in chromosome 1, 6, and 11. These QTL are novel and not co-localized with other QTL that were detected in previous studies. To date, there is a number of QTL analysis on silicon content, using recombinant inbred lines (RILs) from different *indica/japonica* crosses. A total of 10 QTL for silicon content in different rice organs under paddy field conditions were detected on chromosomes 1, 5, 6, 11 and 12 (Dai, et al., 2005). Other 10 QTL for silicon uptake in rice seedlings in nutrition solution were detected on chromosomes 1, 3, 7, 8 and 9 (Wu, et al., 2006). And 2 QTL for silicon content in the leaves under paddy field condition were detected on chromosomes 5 and 10 (Norton, et al., 2010). Recently, a genetic mapping using an F6 population and a GWAS with 350 accessions from the Rice Diversity Panel 1 identified 6 SNPs (4 loci) that co-localise with previously detected QTL for rice silicon or arsenic and A single significant SNP identified in the GWAS mapping within the *tropical japonicas* accessions was detected within 200 kb of Low Silicon 2 (Lsi2) transporter gene on chromosome 3 (Talukdar, et al., 2015).

All the QTL that I discovered from my GWAS for silica content are novel and none of the known silica transporter genes appear in these QTL regions. Interestingly, the QTL on CH6 coincides with the QTL detected from GWAS for digestibility. The*OsAT10* is located close to the most significant SNPs on CH6. Notably, *OsAT10* appears to be a co-expression partner with a putative laccase gene *Lac4* (*LOC_Os01g62490*) found next to the highest peak of the QTL in CH1 (weighted PCC = 0.624856) (http://ricefrend.dna.affrc.go.jp/). This raises an interesting question as to whether there

might be a connection between the candidate genes associated with these two silica QTL.

GWAS for lignin content: Phenotyping for amount of lignin involve a range of methods, however the inconsistencies among them were found (Moreira-Vilar, et al., 2014). Recently, a report, (Moreira-Vilar, et al., 2014), showed that the Acetyl Bromide method (Hatfield, et al., 1999), applied in my research, is faster, simpler and presents better recovery of lignin in different herbaceous tissues than Klason (Bunzel, et al., 2011) and Thioglycolic Acid based methods (Suzuki, et al., 2009). Recently, the approach of association mapping based on examining individual genes and alleles at the loci responsible for lignin content has been applied to identify significant associated SNPs. An intronic SNP in the candidate gene *LpCCR1* in poplar was found significantly associated with cell wall digestibility and Klason lignin content in stem material based on association mapping (Parijs, et al., 2016). By doing association mapping across 40 candidate genes associated with lignin content measured by Pyrolysis Molecular-Beam Mass Spectrometry (PyMBMS), a total of 13 significant single marker associations were found for nine candidate genes in black cottonwood (*Populus trichocarpa*).

In this study, the GWAS identified 8 QTL regions, in which one in CH11 coincides with one found in GWAS for digestibility. This is different from the results in Maize Recombinant Inbred Population, which revealed no overlapping QTL for lignin abundance and saccharification yield (based on glucose and xylose yield) (Penning, et al., 2014). The lignin abundance together with other cell-wall polysaccharides in Penning's work was measured by PyMBMS.

A homolog of *HCT* was found in CH11 at the same regions of QTL for digestibility. Although there are no reports about functional studies of any OsHCT in rice yet, a study of a HCT by reducing its expression in *Medicago* resulted in an increase in stem digestibility, but this was accompanied by a dwarfing phenotype (Chen & Dixon, 2007). There is a cluster of 7 peroxidase genes located next to the QTL peak on CH3 and a laccase gene in the QTL region CH11_18.8 ± 0.3. There are some reports on the impact of modifying peroxidase genes in the lignin content in *Arabidopsis* and *tobacco* (Blee, et al., 2003);(Berthet, et al., 2011); (Zhao, et al., 2013). These candidates might be

playing the role in polymerization of monolignols into lignin in rice and worthy to be investigated more in functional studies.

Conclusion and future work

In conclusion, the association mapping for three traits associated with rice straw quality: digestibility, lignin and silica content, have succeeded in identifying a candidate gene that has high effect on both silica content and digestibility, and many other potential candidate genes. This forward genetic approach appears to be a powerful way to identify the known and novel genes involved in these processes. The result of this research will lead to further studies in two directions (1) to look for the linkages between silica and hemicellulose (2) to scrutinise the potential candidate genes and carry out functional studies to confirm their roles in cell wall biosynthesis in order to facilitate their application in breeding programs aiming to select plants with improved digestibility but without a yield-penalty.

Appendix A - List of studied rice accessions

No	Genotype	O. sativa species	No	Genotype	O. sativa species	No	Genotype	O. sativa species	No	Genotype	O. sativa species
1	09L140	Indica	48	BT7	Indica	95	Lua_nuong_3	Tropical Japonica	142	Te_meo	Indica
2	09L15	Indica	49	BT7_IR64	Indica	96	Lua_nuong_63	Tropical Japonica	143	Te_nuong	Indica
3	09L17	Indica	50	BTS7	Indica	97	N46	Indica	144	Te_ruong__2	Indica
4	09L28	Indica	51	BT_LT2	Indica	98	N46__B6	Indica	145	Toc_lun	Indica
5	09L33	Indica	52	BT_ST	Indica	99	N46__B6	Indica	146	Tran_chau_huong	Indica
6	1094-1	Indica	53	BT_ST	Indica	100	N91	Indica	147	U17	Indica
7	10L140	Indica	54	BT__B5	Indica	101	NH_N46	Indica	148	AC10	Indica
8	10L142	Indica	55	Belecsec	Tropical Japonica	102	Nep358	Indica	149	X_mai	Indica
9	10L144	Indica	56	Beo_Buot_Vang	Tropical Japonica	103	Nep_98	Indica	150	Xi23	Indica
10	10L150	Indica	57	CL8_AC5	Indica	104	Nghi_Huong_1	Indica	151	nep_meo_luong	Tropical Japonica
11	10L155	Indica	58	CL8_LT2	Indica	105	Nghi_Huong	Indica			
12	10L164	Indica	59	CL8_P6	Indica	106	OM2517	Indica			
13	10L182	Indica	60	Caren	Indica	107	OM3536	Indica			
14	11L108	Indica	61	D139	Indica	108	OM4325	Indica			
15	11L12	Indica	62	D381_Q5	Indica	109	OM4900	Indica			
16	11L162	Indica	63	DT7_LT2__BT7	Indica	110	OM5451	Indica			
17	11L17	Indica	64	HD1	Indica	111	OM5451	Indica			
18	142M12	Indica	65	HDT2	Indica	112	OM5494	Indica			
19	AC10_Fukus	Indica	66	HDT4	Indica	113	OM5626	Indica			
20	AC5	Indica	67	HDT5	Indica	114	OM5930	Indica			
21	AC5	Indica	68	HDT7	Indica	115	OM6613	Indica			
22	AC5_149_13	Indica	69	HDt8	Indica	116	OM68__	Indica			
23	AC5_BB1_4	Indica	70	HT1	Indica	117	OM9218	Indica			
24	AC5_CH133	Indica	71	HT6__B5	Indica	118	Om6377	Indica			
25	AC5_HT1	Indica	72	Huong_com	Indica	119	P13	Indica			
26	AC5_Q5__AC4	Indica	73	IR1561__B__	Indica	120	P6	Indica			
27	AC5_Q5__AC4	Indica	74	IR24	Indica	121	P6_ST	Indica			
28	AC5_Q5__C70	Indica	75	IRBB13	Indica	122	P6_ofodiki__P6	Indica			
29	AC5_Q5__C70	Indica	76	IR_	Indica	123	Perai_BT7	Indica			
30	AG_504	Indica	77	IR_	Indica	124	Perai_P6__HT1	Indica			
31	A__hung_cha	Indica	78	Jasmin_AC5	Indica	125	Pet_muong_canh_vang	Tropical Japonica			
32	A__hung_cha_2	Indica	79	KD18	Indica	126	Pet_muong_canh_vang_1	Tropical Japonica			
33	BB10	Indica	80	KG_4900	Indica	127	Q5	Indica			
34	BB1_10_LT2	Indica	81	Kham_Duc	Indica	128	Q5_Fukus__Q5	Indica			
35	BB1_11	Indica	82	Khau_dien_lu	Tropical Japonica	129	Que_Thom	Indica			
36	BB21	Indica	83	Khau_giang	Tropical Japonica	130	RD354	Indica			
37	BB3	Indica	84	Khau_lan_gan_mo	Tropical Japonica	131	SH4	Indica			
38	BB3_7	Indica	85	hau_lan_gan_mon_	Tropical Japonica	132	S_c_tr_ng	Indica			
39	BB4_10	Indica	86	hau_mu_lai_dong_	Tropical Japonica	133	TET4247	Indica			
40	BB4_11	Indica	87	Khau_mumeeng	Indica	134	TET_	Indica			
41	BB4_7	Indica	88	Khau_mumoong	Tropical Japonica	135	Tam_du	Indica			
42	BB5	Indica	89	Khau_munuong	Tropical Japonica	136	Tan_nhe	Tropical Japonica			
43	BB5	Indica	90	LT2	Indica	137	Tan_nhe_2	Tropical Japonica			
44	BB5_7	Indica	91	LT3	Indica	138	Tan_nhe_3	Tropical Japonica			
45	BB7	Indica	92	Lua_ngoi	Indica	139	Tan_nhe__1	Tropical Japonica			
46	BB7	Indica	93	Lua_nuong_1	Tropical Japonica	140	Te_Ka_cham_pi__1	Tropical Japonica			
47	BC15	Indica	94	Lua_nuong_2	Tropical Japonica	141	Te_ka_cham_pi	Tropical Japonica			

List of Abbreviations

4CL – Hydroxycinnamate-CoA ligase

2n – Diploid

A – Absorbance

A – Adenine

AAS _ Atomic Absorption Spectrometry

AIR – Alcohol insoluble residue

Alfalfa – *Medicago sativa*

ANOVA – Analysis of variance

Arabidopsis – *Arabidopsis thaliana*

BAHD – Superfamily named after the first four members of the family to be biochemically characterised (BEAT: benzylalcohol acetyltransferases, AHCT: anthocyanin hydroxycinnamoyl transferase, HCBT: anthranilate hydroxycinnamoyl/benzoyl transferase, DAT: deactylvindoline acetyltransferase)

bp – *Base pairs*

bm– Brown mid-rib mutant

Brachypodium – *Brachypodium distachyon*

C – Cytosine

C3H – 4-Courmarate-3-hydroxylase

C4H – Cinnamate-4-hydroxylase

CAD – Cinnamyl-alcohol dehydrogenase

CCoA-3H – 4-Hydroxycinnamoyl-CoA

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

CCR – Cinnamoyl-CoA-reductase

CH _ Chromosome

CO_2 – Carbon dioxide

Csl – Cellulose synthase-like

DNA – Deoxyribonucleic acid

E5 – Blend of fuel containing 5% ethanol

E10 – Blend of fuel containing 10% ethanol

F5H – Ferulate-5-hydroxylase

FA – Ferulic Acid

FDR – False Discovery Rate

g – Gram

G – Guanine

G – Guaiacyl lignin monomer

GAX – Glucuronoarabinoxylan

GWAS – Genome Wide Association Study

GHG – Greenhouse gas

GlcA – Glucuronic acid

GT – Glycosyltransferase

GUX – GLUCURONIC ACID SUBSTITUTION OF XYLAN

h – Hour

H – *p*-Hydroxyphenyl lignin monomer

H^2 – Broad sense heritability

H_2O – Water

H_2SO_4 – Sulphuric acid

H_4SiO_4 – Monosilic acid

HCl – Hydrochloric acid

HCT – Hydroxycinnamoyl-CoA shikimate/quinate transferase

irx – Irregular xylem mutant

ICP _ Inductively Coupled Plasma

KOR – KORRIGAN

L – Litre

LOD – Logarithm of the odds

MBTH – 3-methy-2benzothiazolinonehydrazone

M – Molar

Mbp – Mega base pairs

MeGlcA – Methyl glucuronic acid

mg – Milligram

min – Minute

ml – Millilitre

MLG – Mixed linkage glucan

Mm – Millimetre

mM – Millimolar

NaOH – Sodium hydroxide

nmol – Nanomole

OMT – *O*-methyltranserase

P – Probability

p-CA – p-Courmaric Acid

PAL – Phenylalanine-ammonia lyase

PCA – Principle component analysis

PCC _ Pearson Correlation Coefficient

PCR – Polymerase chain reaction

Poplar – *Populus*

PyMBMS _ Pyrolysis Molecular-Beam Mass Spectrometry

QTL – Quantitative trait loci

RNA – Ribonucleic acid

RIL _ Recombinant Inbred Lines

RT – Room temperature

S – Syringyl lignin monomer

SD – Standard deviation

Si – Silica

SiO_2 – Silicon dioxide

$Si(OH)_4$ – Silicic acid

SNP – Single nucleotide polymorphism

SSR _ Simple Sequence Repeat

T-DNA – Transfer DNA

v/v – Volume to volume

w/v – Weight to volume

wild-type – Wild-type

V_G – Genotype variance

V_E – Environmental variance

V_T – Total variance

XAT – XYLAN ARABINOSYLTRANSFERASE

XAX – XYLOSYL ARABINOSYL SUBSTITUTION OF XYLAN

XRF – X-ray fluorescence

°C – Degrees Celsius

References

Abdi, H., 2007. Bonferroni and Sidak corrections for multiple comparisons In: Encyclopedia of Measurement and Statistics. CA: Sage, pp. 103 - 107.

Abramson, M., Shoseyov, O., Hirsch, S. & Shani, Z., 2012. Genetic Modifications of Plant Cell Walls to Increase Biomass and Bioethanol Production. In: J. W. (. Lee, ed. *Advanced Biofuels and Bioproducts.* New York, NY, USA: Springer Science+Business Media, pp. 315 - 338.

Acker, R. V. et al., 2014. Improved saccharification and ethanol yield from field-grown transgenic poplar deficient in cinnamoyl-CoA reductase. *Proc Natl Acad Sci USA,* 111(2), pp. 845-850.

Acker, R. V. et al., 2013. Lignin biosynthesis perturbations affect secondary cell wall composition and saccharification yield in Arabidopsis thaliana. *Biotechnology for Biofuels,* Volume 6, p. 46.

AFDC, 2015. *Alternative Fuels Data Center.* [Online] Available at: http://www.afdc.energy.gov/fuels/ethanol_feedstocks.html [Accessed 20 04 2016].

Albersheim, P. et al., 2011. *Plant Cell Walls.* New York, NY, USA: Garland Science, Taylor & Francis Group.

Allen, S., 1989. *Chemical analysis of ecological materials.* 2nd ed. Oxford, UK: Blackwell Scientific Publications..

Appenzeller, L. et al., 2004. Cellulose synthesis in maize: isolation and expression analysis of the cellulose synthase (CesA) gene family. *Cellulose,* 11(3), pp. 287-299.

Aurangzaib, M., 2015. Developmental morphology, biomass yield and compositional differences among upland and lowland switchgrass *(Panicum virgatum L.) ecotypes grown as a bioenergy feedstock crop,* Ames, Iowa, United States: Iowa State University.

Baffi C, B. M. B. G. S. S., 2002. Comparison of different analytical procedures in the determination of trace elements in lichens. *Chemosphere,* 48(3), pp. 299 - 306.

Balasta, M. L. F. C. et al., 1989. Effects of silica level on some properties of Oryza sativa straw and hull. *Canadian Journal of Botany*, 67(8), pp. 2356-2363.

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

Bartley, L. E. et al., 2013. Overexpression of a BAHD Acyltransferase, OsAt10, Alters Rice Cell Wall Hydroxycinnamic Acid Content and Saccharification. *Plant Physiology,* 161(4), p. 1615–1633.

BEC, 2012. *Biomass Energy Center (BEC).* [Online] Available at: http://www.biomassenergycentre.org.uk/portal/page?_pageid=76,15049&_dad=portal&_schema=PORTAL
[Accessed 17 March 2016].

Bell-Lelong, D. A., Cusumano, J. C., Meyer, K. & Chapple, C., 1997. Cinnamate-4-Hydroxylase Expression in Arabidopsis (Regulation in Response to Development and the Environment). *Plant Physiology,* 113(3), pp. 729-738.

Bentley, D. R. et al., 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature,* Volume 456, pp. 53-59.

Berthet, S. et al., 2011. Disruption of LACCASE4 and 17 Results in Tissue-Specific Alterations to Lignification of Arabidopsis thaliana Stems. *The Plant Cell*, 23(3), pp. 1124 - 1137.

Binod, P. et al., 2010. Bioethanol production from rice straw: An overview. *Bioresource Technology,* Volume 101, pp. 4767 - 4774.

Binod, P. et al., 2010. Bioethanol production from rice straw: An overview. *Bioresource Technology,* Volume 101, p. 4767–4774.

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

Boerjan, W., Ralph, J. & Baucher, M., 2003. Lignin Biosynthesis. *Annu. Rev. Plant Biol,* Volume 54, p. 519–546.

Brachi, B., Morri, G. P. & Borevitz, J. O., 2011. Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biology,* 12(10), p. 232.

Bradbury, P. J. et al., 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics,* 23(19), pp. 2633-2635.

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

Bruce Alberts, A. J. et al., 2002. The plant cell wall. In: *Molecular Biology of the Cell (4th ed.)*. Newyork: Garland Science.

Buanafina, M. M. d. O., 2009. Feruloylation in Grasses: Current and Future Perspectives. *Molecular Plant,* 2(5), pp. 861 - 872.

Buanafina, M. M. d. O. et al., 2008. Expression of a fungal ferulic acid esterase increases cell wall digestibility of tall fescue (Festuca arundinacea). *Plant Biotechnology Journal,* 6(3), p. 264–80.

Bunzel, M., Schüßler, A. & Saha, G. T., 2011. Chemical characterization of Klason lignin preparations from plant-based foods. *J Agric Food Chem.,* 59(23), pp. 12506 - 13.

Burton, R. A. et al., 2004. The CesA Gene Family of Barley. Quantitative Analysis of Transcripts Reveals Two Groups of Co-Expressed Genes. *American Society of Plant Biologists,* 134(1), pp. 224-236.

Caldwell, K. S., Russell, J., Langridge, P. & Powell, W., 2006. Extreme Population-Dependent Linkage Disequilibrium Detected in an Inbreeding Plant Species, Hordeum vulgare. *Genetics,* 172(1), pp. 557-567.

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(3), p. 330–337.

Carroll, A. & Somerville, C., 2009. Cellulosic Biofuels. *Annual Review of Plan Biology,* Volume 60, pp. 165-182.

Cătălin Voiniciuc, *. M. G., Schmidt, M. H.-W. & Usadel, B., 2015. Highly Branched Xylan Made by IRREGULAR XYLEM14 and MUCILAGE-RELATED21 Links Mucilage to Arabidopsis Seeds. *Plant Physiology,* 169(4), pp. 2481-2495.

Chapple, C., 1998. Molecular- Genetic analysis of plant cytochrome P450-Dependent Monooxygenases. *Annual Review of Plant Physiology and Plant Molecular Biology,* Volume 49, pp. 311-343.

Chen, F. & Dixon, R. A., 2007. Lignin modification improves fermentable sugar yields for biofuel production. *Nature Biotechnology,* Volume 25, pp. 759 - 761.

Chiniquy, D., Varanasi, P., [...] & Ronald, P. C., 2013. Three Novel Rice Genes Closely Related to the Arabidopsis IRX9, IRX9L, and IRX14 Genes and Their Roles in Xylan Biosynthesis. *Front Plant Sci,* 10(4), p. 83.

Chiniquy, D. et al., 2013. Three Novel Rice Genes Closely Related to the Arabidopsis IRX9, IRX9L, and IRX14 Genes and Their Roles in Xylan Biosynthesis. *Front Plant Sci,* Volume 4, p. 83.

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(4), pp. 1355-1366.

Cocuron, J. C. et al., 2007. A gene from the cellulose synthase-like C family encodes a β-1,4 glucan synthase. 104(20), p. 8550–8555.

Cooke, J. & Leishman, M. R., 2011. Is plant ecology more siliceous than we realise. *Trend in Plant Science,* 16(2), p. 61–68.

Cosgrove, D. J., 2014. Re-constructing our models of cellulose and primary cell wall assembly. *Cell biology,* Volume 22, p. 122–131.

Currie, H. A. & Perry, C. C., 2007. Silica in Plants: Biological, Biochemical and Chemical Studies. *Annals of Botany,* 100(7), pp. 1383-1389.

Currie, H. A. & Perry, C. C., 2009. Chemical evidence for intrinsic 'Si' within Equisetum cell walls. *Phytochemistry,* 70(17-18), pp. 2089-2095.

Dabholkar, A. R., 1992. In: *Elements of Biometrical Genetics.* New Delhi, India: Concept Publishing Company, pp. 138-140.

Dai, W.-M.et al., 2005. Genetic Dissection of Silicon Content in Different Organs of Rice. *Crop Science,* Volume 45, pp. 1345 - 1352.

Davis, . S., Pierce, . J. & Hay, . W., 2014. *Biomass in the energy industry: An introduction.* London: BP p.l.c.

Del Bem, L. E. . V. & Vincentz, M. G., 2010. Evolution of xyloglucan-related genes in green plants. *BMC Evolutionary Biology,* 10(341).

Delmer, D. . P., 1999. CELLULOSE BIOSYNTHESIS: Exciting Times for A Difficult Field of Study. *Annual Review of Plant Physiology and Plant Molecular Biology,* Volume 50, pp. 245-276.

Demirbas, A., 2009. Biofuels securing the planet's future energy needs. *Energy Conversion and Management,* 50(9), p. 2239–2249.

Derena, C. W., Datnoffa, L. E. & Snydera, G. H., 1992. Variable silicon content of rice cultivars grown on everglades histosols. *Journal of Plant Nutrition,* 15(11), pp. 2363 - 2368.

Dhugga, K. S., 2012. Biosynthesis of non-cellulosic polysaccharides of plant cell walls. *Phytochemistry,* Volume 70, p. 8–19.

Dhugga, K. S. et al., 2004. Guar Seed ß-Mannan Synthase Is a Member of the Cellulose Synthase Super Gene Family. *Science,* 303(5656), pp. 363-366.

Djerbi, S. et al., 2005. The genome sequence of black cottonwood (Populus trichocarpa) reveals 18 conserved cellulose synthase (CesA) genes. *Planta,* 221(5), pp. 739-746.

Doblin MS, K. I. J.-W. D. D., 2012. Cellulose biosynthesis in plants: from genes to rosettes. *Plant Cell Physiol,* Volume 43, pp. 1407-1420.

Do, C.-T.et al., 2007. Both caffeoyl Coenzyme A 3-O-methyltransferase 1 and caffeic acid O-methyltransferase 1 are involved in redundant functions for lignin, flavonoids and sinapoyl malate biosynthesis in Arabidopsis. *Planta,* 226(5), pp. 1117-1129.

Doyle, P. & Panday, S., 1990. The feeding value of cereal straws for sheep. III. Supplementation with minerals or minerals and urea. *Animal Feed Science and Technology,* 29(1-2), pp. 29-43.

Duan, L. et al., 2014. Multiple phytohormones and phytoalexins are involved in disease resistance to Magnaporthe oryzae invaded from roots in rice. *486–500,* 152(3), p. 486–500.

Ebringerová, A., Hromádková, Z. & Heinze, T., 2005. Hemicelluloses. In: *Polysaccharides I.* s.l.:Springer, pp. 1-67.

Edye, L. A. & Doherty, W. O. S., 2015. *Fractionation of a lignocellulosic material,* Queesland: Queensland University Of Technology.

Ehlting, J., Shin, J. J. K. & Douglas, C. J., 2001. Identification of 4-coumarate:coenzyme A ligase (4CL) substrate recognition domains. *the plant journal,* 27(5), p. 455–465.

Elshire, R. J. et al., 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE,* 6(5), p. e19379.

Encyclopedia, S., 2009. *Science Encyclopedia.* [Online] Available at: http://science.jrank.org/pages/1333/Cellulose-Structure-cellulose.html [Accessed 29 12 2015].

Enishi, O., 2002. Reevaluation on feed characteristics of rice straw.. *J. Jap. Grassl. Sci.,* Volume 48, pp. 371 - 378.

Epstein, E., 1999. Silicon. *Annual Review of Plant Physiology and Plant Molecular Biology,* Volume 50, pp. 641-664.

Erdtman , H., 1972. Occurrence, formation, structure and reactions. *Journal of Polymer Science Part B: Polymer Letters,* 10(3), pp. 228-230.

Eudes, A. et al., 2016. Exploiting the Substrate Promiscuity of Hydroxycinnamoyl-CoA:Shikimate Hydroxycinnamoyl Transferase to Reduce Lignin. *Plant Cell Physiol,* 57(3), pp. 568 - 579.

Evans, G., 2008. *"International Biofuels Strategy Project. Liquid Transport Biofuels - Technology Status Report, NNFCC 08-017,* York: National Non-Food Crops Centre.

Fagard, M., Höfte, H. & Vernhettes, S., 2000. Cell wall mutants. *Plant Physiology and Biochemistry,* 38(1), pp. 15-25.

Fernandes, A. N. et al., 2011. Nanostructure of cellulose microfibrils in spruce wood. *PNAS,* 108(47), p. E1195–E1203.

Flint-Garcia, S. A., Thornsberry, J. M. & IV, E. S. B., 2003. Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology,* Volume 54, pp. 357-374.

Foust, T. D. et al., 2008. The Biorefinery. In: M. Himmel, ed. *Biomass recalcitrance: Deconstructing the plant cell wall for Bioenergy* . Oxford: Wiley-Blackwell, pp. 7-37.

Fox, R. L., Silva, J. A., Plucknett, D. L. & Teranishi, D. Y., 1969. Soluble and total silicon in sugar cane. *Plant and Soil,* 30(1), pp. 81 - 92.

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

Fry, S. C., 1988. *The Growing Plant Cell Wall: Chemical and Metabolic Analysis.* 1 ed. s.l.:Longman Scientific & Technical.

Fry, S. C., 1989. The Structure and Functions of Xyloglucan. *Journal of Experimental Botany,* 40(210), pp. 1-11.

Fukushima & Hatfield, R., 2004. Comparison of the acetyl bromide spectrophotometric method with other analytical lignin methods for determining lignin concentration in forage samples. *J. Agric. Food Chem,* 52(12), pp. 3713-3720.

Gengshou, X., 2013. Repression of Lignin Synthesis in Rice by C4H and 4CL using RNAi. *International Journal of Bioscience, Biochemistry and Bioinformatics.*, 3(3), pp. 226-228.

Gerya, C. et al., 2011. Natural variation in the freezing tolerance of Arabidopsis thaliana: Effects of RNAi-induced CBF depletion and QTL localisation vary among accessions. *Plant Science*, 180(1), p. 12–23.

Gibson, L. J., 2012. The hierarchical structure andmechanics of plant materials. *J. R. Soc. Interface,* 9(76), p. 2749–2766.

Gillmor, S. C. et al., 2005. Glycosylphosphatidylinositol-Anchored Proteins Are Required for Cell Wall Synthesis and Morphogenesis in Arabidopsis. *American Society of Plant Biologists,* 17(4), pp. 1128-1140.

Glaubitz, J. C. et al., 2014. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE,* 9(2), p. e90346.

Goettemoeller, J. & Goettemoeller, A., 2007. *Sustainable Ethanol: Biofuels, Biorefineries, Cellulosic Biomass, Flex-fuel Vehicles, and Sustainable Farming for Energy Independence.* Missouri: Prairie Oak Publishing.

Gomez, L. D., Bristow, J. K., Statham, E. R. & McQueen-Mason, S. J., 2008. Analysis of saccharification in Brachypodium distachyon stems under mild conditions of hydrolysis. *Biotechnology for Biofuels,* pp. 1-15.

Gomez, L. D., Stelle-King, C. G. & Mason, S. M., 2008. Sustainable liquid biofuels from biomass: the writing's on the walls. *New Phytologist,* 178(3), p. 473–485.

Gomez, L. D. et al., 2010. Automated saccharification assay for determination of digestibility in plant materials. *Biotechnology for Biofuels,* 3(23).

Gomez, L. D., Whitehead, C., Roberts, P. & McQueen-Mason, S. J., 2011. High-throughput Saccharification Assay for Lignocellulosic Materials. *Journal of Visualized Experiment,* Volume 53, p. 3240.

Goubet, F. et al., 2009. Cell wall glucomannan in Arabidopsis is synthesised by CSLA glycosyltransferases, and influences the progression of embryogenesis. *Plant Journal*, 60(3), pp. 527-538.

Goujon, T. et al., 2003. Down-regulation of the AtCCR1 gene in Arabidopsis thaliana: effects on phenotype, lignins and cell wall degradability. *Planta*, 217(2), pp. 218-228.

Govinda R. , T. & Shrestha, A., 2011. How much hope should we have for biofuels?. *Energy*, 36(4), pp. 2055-2069.

Grabbe, J. H., 2004. How Do Lignin Composition, Structure, and Cross-Linking Affect Degradability? A Review of Cell Wall Model Studies. *Crop Science*, 45(3), pp. 820 - 831.

Grabber, J. H., 2005. How do lignin composition, structure, and cross-linking affect degradability? A review of cell wall model studies. *Crop science*, 45(3), pp. 820-831.

Groombridge, B. & Jenkins, M., 2002. *World Atlas of Biodiversity: Earth's Living Resources in the 21st Century*. Cambridge: World Conservation Monitoring Centre;World Conservation Press.

Gu, J., Shen, J. & Li, L., 2011. Functional Characterization of Evolutionarily Divergent 4-Coumarate:Coenzyme A Ligases in Rice. *Plant physiology*, 157(2), pp. 574-586.

Guo, K. et al., 2014. An integrated genomic and metabolomic framework for cell wall biology in rice. *BMC genomic*, 15(1), p. 596.

Gupta, P. K., Rustgi, S. & Kulwal, P. L., 2005. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. *Plant Molecular Biology*, 57(4), pp. 461-485.

Haferkamp, M., 1988. *Environmental factors affecting plant productivity*. Bozeman, Montana, Fort Keogh Research symposium, pp. 27 - 36.

Halpin, C. et al., 1998. Brown-midrib maize (bm1) – a mutation affecting the cinnamyl alcohol dehydrogenase gene. *the plant journal*, 14(5), p. 545–553.

Handakumbura, P., 2014. *Understanding the transcriptional regulation of secondary cell wall biosynthesis in the model grass Brachypodium distachyon,* Massachusetts, US: University of Massachusetts - Amherst.

Hasan, S., Shimojo, M. & Goto, I., 1993. Chemical components influencing lodging resistance of rice plant and its straw digestibility in vitro. *Asian-Australasian Journal of Animal Sciences,* 6(1).

Hästbacka, o. et al., 1992. Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nature Genetics,* Volume 2, pp. 204 - 211.

Hatfield, J. L. & Prueger, J. H., 2015. Temperature extremes: Effect on plant growth and development. *Weather and Climate Extremes,* 10(A), pp. 4 - 10.

Hatfield, R. D., Grabber, J., Ralph, J. & Brei, K., 1999. Using the Acetyl Bromide Assay To Determine Lignin Concentrations in Herbaceous Plants: Some Cautionary Notes. *J. Agric. Food Chem.*, 47(2), pp. 628 - 632.

Hauptkorn, S., Pavel, J. & Seltner, H., 2001. Determination of silicon in biological samples by ICP-OES after non-oxidative decomposition under alkaline conditions. *Fresenius' Journal of Analytical Chemistry,* 370(2), pp. 246 - 250.

Hayter, A. J., 1986. The Maximum Familywise Error Rate of Fisher's Least Significant Difference Test. *Journal of the American Statistical Association,* 81(1986), pp. 1000 - 1004.

Hazen, S. P., Scott-Craig, J. S. & Walton, J. D., 2002. Cellulose Synthase-Like Genes of Rice. *Plant Physiology,* 128(2), pp. 336-340.

He, C., Ma, J. & Wang, L., 2015. A hemicellulose-bound form of silicon with potential to improve the mechanical properties and regeneration of the cell wall of rice. *New Phytologist,* Volume 206, p. 1051–1062.

Highina, B. K. & Bugaje, I. M., 2014. A review on second generation biofuel: A comparison of its carbon footprints. *European Journal of Engineering and Technology,* 2(2), pp. 117 - 125.

Himmel, M. E. et al., 2007. *Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production,* US: Science .

Hirano, K. et al., 2012. OsCAD2 is the major CAD gene responsible for monolignol biosynthesis in rice culm. *Plant Cell Rep,* 31(1), pp. 91-101.

Hirano, K. et al., 2013. Identification of Transcription Factors Involved in Rice Secondary Cell Wall Formation. *Plant Cell Physiol,* 54(11), pp. 1791 - 1802.

Hirschhorn, J. N. & Daly, M. J., 2005. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics,* Volume 6, pp. 95-108.

Hoch, G., 2007. Cell wall hemicelluloses as mobile carbon stores in non-reproductive plant tissues. *Functional Ecology,* 21(823–834).

Hoeven, M. V., 2014. *CO2 Emissions from fuel combustion Highlight ,* Lima: International Energy Agency.

Hoffmann, L. et al., 2004. Silencing of Hydroxycinnamoyl-Coenzyme A Shikimate/Quinate Hydroxycinnamoyltransferase Affects Phenylpropanoid Biosynthesis. *Plant Cell,* 16(6), p. 1446–1465.

Holley, M. S., Higson, A. & Evans, G., 2007. Bring on the biorefinery. *The Chemical Engineer,* Issue 795, p. 46.

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

Huang, X. & Han, B., 2014. Natural Variations and Genome-Wide Association Studies in Crop Plants. *Annual Review of Plant Biology,* Volume 65, pp. 531-551.

Huang, X. et al., 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature genetic,* 967(11), p. 961.

I. Ceres, 2015. *Carbon dioxide/Net energ,* California, United States: Cares, Inc.,.

Inanaga, S. & Okasaka, A., 1995. Calcium and silicon binding compounds in cell walls of rice shoots. *Soil Science and Plant Nutrition,* 41(1), pp. 103-110.

Inanaga, S., Okasaka, A. & Tanaka, S., 1995. Does silicon exist in association with organic compounds in rice plant?. *Soil Science and Plant Nutrition,* 41(1), pp. 111 - 117.

Jaccard, J., Becker, M. A. & Wood, G., 1984. Pairwise multiple comparison procedures: A review. *Psychological Bulletin,* 96(3), pp. 589 - 596.

Johnson, M. T. J., Agrawal, A. A., Maron, J. L. & Salminen, d.-P., 2009. Heritability, covariation and natural selection on 24 traits of common evening primrose (Oenothera biennis) from a field experiment. *Journal of Evolutionary Biology,* 22(6), pp. 1295 - 1307.

Jones, L. & Handreck, K., 1967. Silica in soils, plants and animals. A review. *ADVANCES IN AGRONOMY,* Volume 19, pp. 107-149.

Kang, H. M. et al., 2008. Efficient Control of Population Structure in Model Organism Association Mapping. *GENETICS,* 178(3), pp. 1709-1723.

Katiyar, A. et al., 2012. Genome-wide classification and expression analysis of MYB transcription factor families in rice and Arabidopsis. *BMC Genomics,* Volume 13, p. 544.

Kawahara, Y. et al., 2013. Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. *Rice,* 6(1), p. 4.

Keegstra, K. & Walton, J., 2006. β-Glucans--Brewer's Bane, Dietician's Delight. *Science,* 311(5769), pp. 1872-1873.

Khush, G. S., Juliano, B. O. & Roxas, D. B., 1987. *Genetic selection for improved nutritional quality of rice straw-a plant breeder's viewpoint.* Addis Ababa, Ethiopia, s.n.

Kiemle, S. N. et al., 2014. Role of (1,3)(1,4)-β-Glucan in Cell Walls: Interaction with Cellulose. *Biomacromolecules,* 15(5), p. 1727–1736.

Kim, A., Kim, B.-G., Kim, M. & Ahn, J.-H., 2012. Characterization of hydroxycinnamoyltransferase from rice and its application for biological synthesis of hydroxycinnamoyl glycerols. *Phytochemistry,* Volume 76, pp. 25-31.

Kim, I. A., Kim, B.-G., Kim, M. & Ahn, J.-H., 2012. Characterization of hydroxycinnamoyltransferase from rice and its application for biological synthesis of hydroxycinnamoyl glycerols. *Phytochemistry,* Volume 76, pp. 25-31.

Kim, J.-H.et al., 2006. Cloning, Characterization, and Expression of Two cDNA Clones for a Rice Ferulate-5-Hydroxylase Gene, a Cytochrome P450-Dependent Monooxygenase. *Journal of Plant Biology*, 49(3), pp. 200-204 .

Kim, S. J., Zemelis, S., Keegstra, K. & Brandizzi, 2015. The cytoplasmic localization of the catalytic site of CSLF6 supports a channeling model for the biosynthesis of mixed-linkage glucan.. *Plant Journal,* 81(4), pp. 537-547.

Kim, S.-J.et al., 2004. Functional reclassification of the putative cinnamyl alcohol dehydrogenase multigene family in Arabidopsis. *Proc Natl Acad Sci USA,* 101(6), pp. 1455-1460.

Kneusel, R. E., Matern, U. & Nicolay, K., 1989. Formation of trans-caffeoyl-CoA from trans-4-coumaroyl-CoA by Zn2+-dependent enzymes in cultured plant cells and its activation by an elicitor-induced pH shift. *Archives of Biochemistry and Biophysic*, 269(2), p. 455–462.

Krishnan, P., Ramakrishnan, B., Reddy, K. R. & Reddy, V. R., 2011. High-Temperature Effects on Rice Growth, Yield, and Grain Quality. *Advances in Agronomy,* Volume 111, pp. 87 - 206.

Lane, D. R. et al., 2001. Temperature-Sensitive Alleles of RSW2 Link the KORRIGAN Endo-1,4-β-Glucanase to Cellulose Synthesis and Cytokinesis in Arabidopsis. *American Society of Plant Physiologists,* 126(1), pp. 278-288.

Langan, P. et al., 2014. Common processes drive the thermochemical pretreatment of lignocellulosic biomass. *Green Chemistry,* Volume 16, p. 63–68.

Lauvergeat, V. et al., 2001. Two cinnamoyl-CoA reductase (CCR) genes from Arabidopsis thaliana are differentially expressed during development and in response to infection with pathogenic bacteria. *Phytochemistry,* 57(7), p. 1187–1195.

Lee, C., Teng, Q., Zhong, R. & Ye, Z.-H., 2012. Arabidopsis GUX proteins are glucuronyltransferases responsible for the addition of glucuronic acid side chains onto xylan. *Plant Cell Physiol,* 53(7), pp. 1204 - 1216.

Leisola, M., Pastinen, O. & Axe, D. D., 2012. Lignin--Designed Randomness. *BIO-Complexity,* Volume 3, pp. 1-11.

Leplé, J.-C.et al., 2007. Downregulation of Cinnamoyl-Coenzyme A Reductase in Poplar: Multiple-Level Phenotyping Reveals Effects on Cell Wall Polymer Metabolism and Structure. *The Plant Cell,* 19(11), pp. 3669-3691.

Liepman, A. H. et al., 2007. Functional Genomic Analysis Supports Conservation of Function Among Cellulose Synthase-Like A Gene Family Members and Suggests Diverse Roles of Mannans in Plants. *Plant Physiology,* 143(4), pp. 1881-1893.

Liepman, A. H., Wilkerson, C. . G. & Keegstra , K., 2005. Expression of cellulose synthase-like (Csl) genes in insect cells reveals that CslA family members encode mannan synthases. *Proc Natl Acad Sci U S A.,* 102(6), p. 2221–2226.

Lipka, A. E. et al., 2012. GAPIT: genome association and prediction integrated tool. *Bioinformatics,* 28(18), pp. 2397-2399.

Li, S., Bashline, L., Lei, . L. & Gu, Y., 2014. Cellulose Synthesis and Its Regulation. *The Arabidopsis Book / American Society of Plant Biologists.*, Volume e0169.

Liu, C. J. et al., 2014. Tailoring lignin biosynthesis for efficient and sustainable biofuel production. *Plant Biotechnology Journal,* 12(9), p. 1154–1162.

Li, X. et al., 2013. Downregulation of caffeoyl-CoA O-methyltransferase (CCoAOMT) by RNA interference leads to reduced lignin production in maize straw. *Genetic and Molecular Biology,* 36(4), p. 540–546.

Li, X. et al., 2009. FLEXIBLE CULM 1 encoding a cinnamyl-alcohol dehydrogenase controls culm mechanical strength in rice. *Plant Molecular Biology,* 69(6), p. 685–697.

Luttenegger, D. G. & Nevins, D. J., 1985. Transient Nature of a (1 m 3), (1 -m 4)-f-D-Glucan in Zea mays Coleoptile Cell Walls. *Plant Physiology,* 77(1), pp. 175-178.

Mackay, I. & Powell, W., 2007. Methods for linkage disequilibrium mapping in crops. *Trends in Plant Scicence,* 12(2), pp. 57 - 63.

Maeda, Y., Awano, T., Takabe, K. & Fujita, M., 2000. Immunolocalization of glucomannans in the cell wall of differentiating tracheids inChamaecyparis obtusa. *Protoplasma,* 213(3), pp. 148-156.

Ma, J. F., 2003. Functions of Silicon in Higher Plants. In: W. E. G. (. Müller, ed. *Silicon Biomineralization.* Berlin: Springer, pp. 127 - 147.

Ma, J. F., 2004. Role of silicon in enhancing the resistance of plants to biotic and abiotic stresses. *Soil Science and Plant Nutrition,* 50(1), pp. 11-18.

Ma, J. F. & Takahashi, E., 2002. *Soil, Fertilizer, and Plant Silicon Research in Japan.* Amsterdam: Elsevier Science & Technology.

Ma, J. F. et al., 2006. A silicon transporter in rice. *Nature,* Volume 440, pp. 688-691.

Ma, J. F. et al., 2007. An efflux transporter of silicon in rice. *Nature,* Volume 448, pp. 209-212.

Majumder, N. D., Rakshit, S. C. & Borthakur, D. N., 1985. Genetics of silica uptake in selected genotypes of rice (O. sativa L.). *Plant and Soil,* 88(3), pp. 449 - 453.

Mann, P., Gahagan, L. & Gordon, M. B., 2003. Tectonic setting of the world's giant oil and gas fields. In: O. Tulsa, ed. *Giant Oil and Gas Fields of the Decade, 1990–1999.* USA: American Association of Petroleum Geologists (AAPG), pp. 55-105.

Marie, B., Monties, B., Montagu, M. V. & Boerjan, W., 1998. Biosynthesis and Genetic Engineering of Lignin. *Critical Reviews in Plant Sciences,* 17(2), pp. 125-197.

Marriott, P. E., Leonardo , . G. D. & McQueen-Mason, S. J., 2015. Unlocking the potential of lignocellulosic biomass through plant science. *New Phytologist.*

Martone, P. T. et al., 2008. Discovery of Lignin in Seaweed Reveals Convergent Evolution of Cell-Wall Architecture. *Current Biology,* 19(2), p. 169–175.

Masson, P., Dauthieu, M., Trolard, F. & Denaix, L., 2007. Application of direct solid analysis of plant samples by electrothermal vaporization-inductively coupled plasma atomic emission spectrometry: Determination of Cd and Si for environmental purposes. *Spectrochimica Acta Part B: Atomic Spectroscopy,* 62(3), pp. 224 - 230.

Mather, K. A. et al., 2007. The Extent of Linkage Disequilibrium in Rice (Oryza sativa L.). *Genetics,* 177(4), pp. 2223 - 2232.

McNallya, K. L. et al., 2009. Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc Natl Acad Sci U S A,* 106(30), pp. 12273 - 12278.

McNally, K. L. et al., 2009. Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *PNAS,* 106(30), p. 12273–12278.

Mitani, N. & Ma, J. F., 2005. Uptake system of silicon in different plant species. *Journal of Experimental Botany,* 56(414), pp. 1255-1261.

Mitchell, R. A., Dupree, P. & Shewry, P. R., 2007. A Novel Bioinformatics Approach Identifies Candidate Genes for the Synthesis and Feruloylation of Arabinoxylan. *American Society of Plant Biologists,* 144(1), pp. 43 - 53.

Molinero, A. L., Martinez, L., Villareal, A. & Castillo, J., 1998. Silicon determination by inductively coupled plasma atomic emission spectrometry after generation of volatile silicon tetrafluoride. *Talanta,* 45(6), pp. 1211 - 1217.

Moore, P. . J. & Staehelin, A. L., 1988. Immunogold localization of the cell-wall-matrix polysaccharides rhamnogalacturonan I and xyloglucan during cell expansion and cytokinesis inTrifolium pratense L.; implication for secretory pathways. *Planta,* 174(4), pp. 433-445.

Moreira-Vilar, F. C. et al., 2014. The Acetyl Bromide Method Is Faster, Simpler and Presents Best Recovery of Lignin in Different Herbaceous Tissues than Klason and Thioglycolic Acid Methods. *PLoS ONE,* 9(10), p. e110000.

MSU, R. G. A. P. T., 2009. *Rice Genome Annotation Project.* [Online] Available at: http://rice.plantbiology.msu.edu/ [Accessed 2015].

Naik, S. N., Goud, V. V., Rout, P. . K. & Dalai, A. K., 2010. Production of first and second generation biofuels: A comprehensive review. *Renewable and Sustainable Energy Reviews,* 14(2), p. 578–597.

Neil, E., 2006. *Chemistry and industry* , Issue 17, pp. 24-25.

Newman, R. H., Hill, S. J. & Harris, P. J., 2013. Wide-Angle X-Ray Scattering and Solid-State Nuclear Magnetic Resonance Data Combined to Test Models for Cellulose Microfibrils in Mung Bean Cell Walls. *Plant Physiol,* 163(4), p. 1558–1567.

Noda, S. et al., 2015. The expression of a rice secondary wall-specific cellulose synthase gene, OsCesA7, is directly regulated by a rice transcription factor, OsMYB58/63. *planta,* 242(3), pp. 589-600.

Norton, G. J. et al., 2010. Genetic mapping of the rice ionome in leaves and grain: identification of QTL for 17 elements including arsenic, cadmium, iron and selenium. *Plant and Soil,* 329(1), pp. 139 - 153.

NREL, 2005. *National Renewable Energy Laboratory.* [Online] Available at: http://www.nrel.gov/biomass/biorefinery.html [Accessed 28 12 2015].

Oikawa, A., Lund, C. H., Sakuragi, Y. & Scheller, H. . V., 2013. Golgi-localized enzyme complexes for plant cell wall biosynthesis. *Trends Plant Sci,* 18(1), pp. 49-58.

Ookawa, T. et al., 2014. Increased lodging resistance in long-culm, low-lignin gh2 rice for improved feed and bioenergy production. *Scientific Reports,* Volume 4, p. 6567.

Oraguzie, N. C., Rikkerink, E. H. A., Gardiner, D. S. E. & Silva, H. N. D., 2007. *Association Mapping in Plants.* 1 ed. New York: Springer-Verlag New York.

Pagant, S. et al., 2002. KOBITO1 Encodes a Novel Plasma Membrane Protein Necessary for Normal Synthesis of Cellulose during Cell Expansion. *Plant Cell,* 14(9), pp. 2001-2013.

Pakusch, A.-E. & Matern, U., 1991. Kinetic characterization of caffeoyl-coenzyme a-specific 3-o-methyltransferase from elicited parsley cell suspensions. *plant physiology*, 96(1), pp. 327-330.

Palaisa, K. A., Morgante, M., Williamsb, M. & Rafalski, A., 2003. Contrasting Effects of Selection on Sequence Diversity and Linkage Disequilibrium at Two Phytoene Synthase Loci. *The Plant Cell*, 15(8), pp. 1795-1806.

Parijs, F. R. D. v. et al., 2016. Association Mapping of LpCCR1 with Lignin Content and Cell Wall Digestibility of Perennial Ryegrass. In: I. Roldán-Ruiz, J. Baert & D. Reheul, eds. *Breeding in a World of Scarcity*. s.l.:Springer International Publishing, pp. 219 - 224.

Park, H. L. et al., 2013. Transcriptomic analysis of UV-treated rice leaves reveals UV-induced phytoalexin biosynthetic pathways and their regulatory networks in rice. *Phytochemistry*, Volume 96, p. 57–71.

Park, W., Scheffler, B. E., Bauer, P. J. & Campbell, B. T., 2010. Identification of the family of aquaporin genes and their expression in upland cotton (Gossypium hirsutum L.). *BMC Plant Biol*, 10(142).

Pasam, R. K. et al., 2012. Genome-wide association studies for agronomical traits in a world wide spring barley collection. *BMC Plant Biology*, Volume 12, p. 16.

Penning, B. W. et al., 2014. Genetic Determinants for Enzymatic Digestion of Lignocellulosic Biomass Are Independent of Those for Lignin Abundance in a Maize Recombinant Inbred Population. *Plant Physiology*, 165(4), pp. 1475 - 1487.

Penning, B. W. et al., 2014. Genetic Determinants for Enzymatic Digestion of Lignocellulosic Biomass Are Independent of Those for Lignin Abundance in a Maize Recombinant Inbred Population. *Plant Physiology*, 165(4), pp. 1475 - 1487.

Piston, F., Uauy, C. & Dubcovsky, J., 2010. Down-regulation of four putative arabinoxylan feruloyl transferase genes from family PF02458 reduces ester-linked ferulate content in rice cell walls. *planta*, 231(3), pp. 677-691.

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

Poke, F. S., Potts, B. M., Vaillancourt, R. E. & Raymond, C. A., 2006. Genetic parameters for lignin, extractives and decay in Eucalyptus globulus. *Annals of Forest Science,* 63(8), pp. 813 - 821.

Pradea, R. A., 1996. Xylanases: from Biology to BioTechnology. *Biotechnology and Genetic Engineering Reviews,* 13(1), pp. 101-132.

Puls, J., 1997. Chemistry and biochemistry of hemicelluloses: Relationship between hemicellulose structure and enzymes required for hydrolysis. *Macromolecular Symposia,* 120(1), p. 183–196.

Raes, J. et al., 2003. Genome-Wide Characterization of the Lignification Toolbox in Arabidopsis. *Plant Physiology,* 133(3), pp. 1051-1071.

Reidinger, S., Ramsey, M. H. & Hartley, S. E., 2012. Rapid and accurate analyses of silicon and phosphorus in plants using a portable X-ray fluorescence spectrometer. *New Phytologist,* 195(3), p. 699–706.

RFA, 2013. *Renewable Fuel Association.* [Online] Available at: http://ethanolrfa.org/pages/World-Fuel-Ethanol-Production [Accessed 14 September 2015].

RFA, R. F. A., 2014. *Going Global - 2015 Ethanol Industry Outlook.* Washington : s.n.

Rohde, A. et al., 2004. Molecular Phenotyping of the pal1 and pal2 Mutants of Arabidopsis thaliana Reveals Far-Reaching Consequences on Phenylpropanoid, Amino Acid, and Carbohydrate Metabolism. *The Plant Cell,* 16(10), pp. 2749-2771.

Rosss.W, 2013. *wikipedia.* [Online] Available at: https://en.wikipedia.org/wiki/List_of_countries_by_oil_production [Accessed 14 September 2015].

Roudier, F. et al., 2005. COBRA, an Arabidopsis Extracellular Glycosyl-Phosphatidyl Inositol-Anchored Protein, Specifically Controls Highly Anisotropic Expansion through Its Involvement in Cellulose Microfibril Orientation. *Plant Cell,* 17(6), pp. 1749-1763.

Rowell, R. M., Pettersen, R. & Tshabalala, M. A., 2012. Cell Wall Chemistry. In: R. M. Rowell, ed. *Handbook of Wood Chemistry and Wood Composites.* Boca Raton, FL, USA: CRC Press, Taylor and Francis Group, pp. 33 - 72.

Rulli, M. C. et al., 2016. The water-land-food nexus of firstgeneration. *Scientific Reports,* Volume 6.

Saathoff, A. J. et al., 2011. Downregulation of cinnamyl-alcohol dehydrogenase in switchgrass by RNA silencing results in enhanced glucose release after cellulase treatment. *PLoS One,* 6(1), p. e16416..

Saballos, A. et al., 2012. Brown midrib2 (Bmr2) encodes the major 4-coumarate:coenzyme A ligase involved in lignin biosynthesis in sorghum (Sorghum bicolor (L.) Moench). *the plant journal,* 70(5), p. 818–830.

Saheb, M. et al., 2015. Importance of Silicon and Mechanisms of Biosilica Formation in Plants. *BioMed Research International,* Volume 396010, p. 16 pages.

Sattler, S. E., Funnell-Harrisa, D. L. & Pederse, J. F., 2010. Brown midrib mutations and their importance to the utilization of maize, sorghum, and pearl millet lignocellulosic tissues. *Plant Science,* 178(3), p. 229–238.

Scheller, H. V. & Ulvskov, P., 2010. Hemicelluloses. *PLANT BIOLOGY,* Volume 61, pp. 263-289.

Schilmiller, A. L. et al., 2009. Mutations in the cinnamate 4-hydroxylase gene impact metabolism, growth and development in Arabidopsis. *the plant journal,* 60(5), p. 771–782.

Schindelman, G. et al., 2001. COBRA encodes a putative GPI-anchored protein, which is polarly localized and necessary for oriented cell expansion in Arabidopsis. *Genes & Development,* 15(9), pp. 1115-1127.

Schultink, A. et al., 2013. The Identification of Two Arabinosyltransferases from Tomato Reveals Functional Equivalency of Xyloglucan Side Chain Substituents. *Plant Physiology,* 163(1), pp. 86-94.

Schwarz, K., 1973. A bound form of silicon in glycosaminoglycans and polyuronides.. *Proc Natl Acad Sci U S A.,* 70(5), pp. 1608-1612.

Shmulsky, R. & Jones, P. D., 2011. *Forest Products and Wood Science An Introduction.* 6th ed. Chichester, UK: Wiley-Blackwell.

Sibouta, R. et al., 2005. CINNAMYL ALCOHOL DEHYDROGENASE-C and -D are the primary genes involved in lignin biosynthesis in the floral stem of Arabidopsis. *The Plant Cell,* 17(7), pp. 2059-2076.

Singh, A. et al., 2009. Allelic variation at Psy1-A1 and association with yellow pigment in durum wheat grain. *Theoretical and Applied Genetics,* 118(8), pp. 1539-1548.

Sixta, H., 2006. *Handbook of Pulp.* 1st ed. Germany: Wiley-VCH Verlag GmbH & Co. KGaA.

Smith-Moritz, A. M. et al., 2015. Structural characterization of a mixed-linkage glucan deficient mutant reveals alteration in cellulose microfibril orientation in rice coleoptile mesophyll cell walls. *Plant Science,* Volume 6, p. 628.

Soest, P. V., 1970. *The role of silicon in the nutrition of plants and animals.* NY, Proceedings of the Cornell Nutrition Conference, pp. 103-109.

Soest, P. V., 1994. *Nutritional Ecology of the Ruminant.* 2nd ed. NY: Cornell University Press.

Somerville, C., 2006. Cellulose Synthesis in higher plants. *Annu. Rev. Cell Dev. Biol,* Volume 22, pp. 53-78.

Sommer, M., Kaczorek, D., Kuzyakov, Y. & Breuer, J., 2006. Silicon pools and fluxes in soils and landscapes—a review. *J. Plant Nutr. Soil Sc*

i., Volume 169, p. 310–329.

Souza, A. P. D. et al., 2013. Composition and Structure of Sugarcane Cell Wall Polysaccharides: Implications for Second-Generation Bioethanol Production. *BioEnergy Research,* 6(2), pp. 564-579.

Sticklen, M. B., 2008. Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol. *Nature Reviews Genetics,* Volume 9, pp. 433-443.

Stracke, S. et al., 2008. Association mapping reveals gene action and interactions in the determination of flowering time in barley. *Theoretical and Applied Genetics,* 118(2), pp. 259-273.

STS, 2012. *Sustainable Transport Solutions.* [Online] Available at: http://www.sts-technology.com/docs/Bioethanol-Fact-Sheet-Final.pdf [Accessed 14 September 2015].

Sun, H. et al., 2013. Analysis of five rice 4-coumarate:coenzyme A ligase enzyme activity and stress response for potential roles in lignin and flavonoid biosynthesis in rice. *Biochemical and Biophysical Research Communications,* 430(3), p. 1151–1156.

Suzuki, S. et al., 2009. High-throughput determination of thioglycolic acid lignin from rice. *Plant Biotechnology,* 26(3), pp. 337 - 340.

Szyjanowicz, P. M. J. et al., 2004. The irregular xylem 2 mutant is an allele of korrigan that affects the secondary cell wall of Arabidopsis thaliana. *The plant journar: for cell and molecular biology,* 37(5), pp. 730-740.

Talukdar, P., Douglas, A., Price, A. H. & Norton, G. J., 2015. Biallelic and Genome Wide Association Mapping of Germanium Tolerant Loci in Rice (Oryza sativa L.). *PLoS One,* 10(9), p. e0137577.

Tanaka, K. et al., 2003. Three Distinct Rice Cellulose Synthase Catalytic Subunit Genes Required for Cellulose Synthesis in the Secondary Wall. *Plant Physiology,* 133(1), pp. 73 - 83.

Taylor, N. G., Laurie, S. & Turner, S. R., 2000. Multiple Cellulose Synthase Catalytic Subunits Are Required for Cellulose Synthesis in Arabidopsis. *American Society of Plant Physiologists,* 12(12), pp. 2529-2539.

Thomas, L. H. et al., 2013. Structure of Cellulose Microfibrils in Primary Cell Walls from Collenchyma. *Plant Physiol,* 161(1), p. 465–476.

Thornsberry, J. M. et al., 2001. Dwarf8 polymorphisms associate with variation in flowering time. *Nature Genetics,* Volume 28, pp. 286-289.

Ukrainetz, i. K. et al., 2008. Heritability and phenotypic and genetic correlations of coastal Douglas-fir (Pseudotsuga menziesii) wood quality traits. *Canadian Journal of Forest Research,* 38(6), pp. 1536-1546.

US.DOE, 2015. *Lignocellulosic Biomass for Advanced Biofuels and Bioproducts: Workshop Report,* US: Office of Biological and Environmental Research.

Van Soest, P., 2006. Rice straw, the role of silica and treatments to improve quality. *Animal Feed Science and Technology,* 130(3-4), pp. 137-171.

Vanholme, R. et al., 2010. Lignin Biosynthesis and Structure. *Plant Physiology,* 153(3), pp. 895-905.

VanRaden, P., 2008. Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science,* 91(11), pp. 4414-4423.

Vega-Sánchez, M. E. & Ronald, P. C., 2010. Genetic and biotechnological approaches for biofuel crop improvement. *Current Opinion in Biotechnology,* 21(2), p. 218–224.

Ventura, Y. et al., 2011. Effects of day length on flowering and yield production of Salicornia and. *Scientia Horticulturae,* Volume 130, pp. 510 - 516.

Vergara, B. S. &. T.-T. C., 1985. *The flowering response of the rice plant to photoperiod: a review of the literature.* 4th ed. Los Banos, Laguna, Philippines: Int. Rice Res. Inst.

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

Wagner, A. et al., 2009. Suppression of 4-coumarate-CoA ligase in the coniferous gymnosperm Pinus radiata. *Plant Physiology*, 149(1), pp. 370-383.

Wang, J. et al., 2015. Lignin engineering through laccase modification: a promising field for energy plant improvement. *Biotechnology for Biofuels*, 8(145).

Wang, Y., Mortimer, J. C., Davis, J. & Dupree, P., 2013. Identification of an additional protein involved in mannan biosynthesis. *Plant Journal*, 73(1), p. 105–117.

Wang, Y.-H., Poudel, D. & Hasenstein, K. H., 2011. Identification of SSR markers associated with saccharification yield using pool-based genome-wide association mapping in sorghum. *Genome*, 54(11), pp. 883 - 889.

Wang, Z. et al., 2016. Association Mapping for Fiber-Related Traits and Digestibility in Alfalfa (Medicago sativa). *Plant Sciences*, Volume 7, p. 331.

Wende, G. & Fry, . S. . C., 1997. 2-O-β-D-xylopyranosyl-(5-O-feruloyl)-L-arabinose,, a widespread component of grass cell walls. *Phytochemistry*, 44(6), p. 1019–1030.

Weng, J.-K., Mo, H. & Chapple, C., 2010. Over-expression of F5H in COMT-deficient Arabidopsis leads to enrichment of an unusual lignin and disruption of pollen wall formation.. *The plant journal*, 64(6), pp. 898-911.

Winslow, M. D., Okada, K. & Correa-Victoria, F., 1997. Silicon deficiency and the adaptation of tropical rice ecotypes. *Plant and Soil*, 188(2), pp. 239 - 248.

Workman, D., 2016. *World's Top exports.* [Online] Available at: http://www.worldstopexports.com/rice-exports-country/ [Accessed 30 May 2016].

Wu, Q.-S.et al., 2006. Genetic dissection of silicon uptake ability in rice (Oryza sativa L.). *Plant Science*, 171(4), pp. 441 - 448.

Wu, X. et al., 2013. Phylogenetic, Molecular, and Biochemical Characterization of Caffeic Acid o-Methyltransferase Gene Family in Brachypodium distachyon. *International Journal of Plant Genomics*, Volume 2013, p. 12 pages.

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

Xu, Z. et al., 2009. Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. *BMC Bioinformatics,* 10(11), p. S3.

Yamaji, N., Mitatni, N. & Ma, J. F., 2008. A Transporter Regulating Silicon Distribution in Rice Shoots. *The Plant cell,* 20(5), pp. 1381-1389.

Yamaji, N., Sakurai, G., Mitani-Ueno, N. & Ma, J. F., 2015. Orchestration of three transporters and distinct vascular structures in node for intervascular transfer of silicon in rice. *Proc Natl Acad Sci,* 112(36), pp. 11401 - 11406.

Yamamoto, T. et al., 2012. Effect of silicon deficiency on secondary cell wall synthesis in rice leaf. *Journal of Plant Research,* 125(6), pp. 771-779.

Yang, C. et al., 2007. Arabidopsis MYB26/MALE STERILE35 Regulates Secondary Thickening in the Endothecium and Is Essential for Anther Dehiscence. *The Plant Cell,* 19(2), pp. 534 - 548.

Yoon, J., Choi, H. & An, G., 2015. Roles of lignin biosynthesis and regulatory genes in plant development. *Journal of Integrative Plant Biology,* 57(11), pp. 902-912.

Yuan, Q. et al., 2001. Rice Bioinformatics. Analysis of Rice Sequence Data and Leveraging the Data to Other Plant Species. *Plant Physiology,* 125(3), pp. 1166 - 1174.

Yu, J. et al., 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics,* Volume 203-208, p. 38.

Zabotina, O. A., 2012. Xyloglucan and Its Biosynthesis. *Front Plant Sci,* Volume 3, p. 134.

Zeng, J., Singh, D., Gao, D. & Chen, S., 2014. Effects of lignin modification on wheat straw cell wall deconstruction by Phanerochaete chrysosporium. *Biotechnology for Biofuels,* 7(161).

Zhang, J. et al., 2015. Silica distinctively affects cell wall features and lignocellulosic saccharification with large enhancement on biomass production in rice. *Plant Science*, Volume 239, pp. 84-91.

Zhang, K. et al., 2006. GOLD HULL AND INTERNODE2 Encodes a Primarily Multifunctional Cinnamyl-Alcohol Dehydrogenase in Rice. *Plant Physiology,* 140(3), pp. 972-983.

Zhang, Z. et al., 2010. Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics,* Volume 42, p. 355–360.

Zhao, H. et al., 2004. Characterization of three riceCCoAOMT genes. *Chinese Science Bulletin,* 49(15), pp. 1602-1606.

Zhao, K. et al., 2007. An Arabidopsis Example of Association Mapping in Structured Samples. *PLoS Genetics,* 3(1), p. e4.

Zhao, Q. et al., 2013. Laccase is necessary and nonredundant with peroxidase for lignin polymerization during vascular development in Arabidopsis. *The Plant Cell,* 25(10), pp. 3976 - 3987.

Zhu, C., Gore, M., Buckler, E. S. & Yu, J., 2008. Status and Prospects of Association Mapping in Plants. *The Plant Genome,* 1(1), pp. 5-20.