ABSTRACT

Understanding what contributes to the genetic control of stem mechanical strength in important crop species, has potential for the improvement of lodging resistance. This may be particularly important given the recent interest in the exploitation of crop residues as feedstock for lignocellulosic ethanol production, where the manipulation of cell wall chemical composition to improve feedstock processibility may lead to reduced stem strength. Lodging is a widespread problem in both *Brassica napus* (*B. napus*) and wheat, and is a key contributor to yield loss. The development of genetic markers that could be used in the selection of elite accessions with high stem mechanical strength is an approach that could compliment current strategies in place for lodging control.

The level of genetic variation available for stem mechanical strength and related traits was assessed across wheat and *B. napus* diversity panels. Following the detection of high levels of variation, these traits were analysed using Associative Transcriptomics. Important variation at both the sequence and the gene expression level was identified.

This analysis revealed a potential importance of xylan acetylation in wheat. For *B. napus* a subset of candidate genes were validated in an *Arabidopsis thaliana* mutant screen, which revealed *GAUT5, SAUR72* and a pectin methylesterase (AT3G12880), as key contributors to stem mechanical strength. Xylan acetylation and pectin methylesterification are both known to impair saccharification efficiency of lignocellulosic biomass, suggesting that there may be conflict between feedstock processibility and stem mechanical strength.

A subset of markers detected through Associative Transcriptomics were further tested an independent panel of *B. napus* and wheat breeding material. Through genotyping and subsequent mechanical testing, markers with high durability for the selection of stem mechanical strength were identified in both crop species. These experiments demonstrate the power of Associative Transcriptomics for the identification of durable markers for complex traits of agronomic importance.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTARY FIGURES (SEE ATTACHED CD)

Supplementary Figure 3. 1. Associative Transcriptomics SNP (A) and GEM (B) results for Fmax for JIC-grown *B. napus.* Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C.

Supplementary Figure 3. 2. Associative Transcriptomics SNP (A) and GEM (B) results for F/V for JIC-grown *B. napus*. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C.

Supplementary Figure 3. 3. Associative Transcriptomics SNP (A) and GEM (B) results for second moment of area for JIC-grown *B. napus.* Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C

Supplementary Figure 3. 4. Associative Transcriptomics SNP (A) and GEM (B) results for stem diameter for JIC-grown *B. napus.* Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C

Supplementary Figure 3. 5. Associative Transcriptomics SNP (A) and GEM (B) results for MOE for JIC-grown *B. napus.* Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C

Supplementary Figure 3. 6. Associative Transcriptomics SNP (A) and GEM (B) results for MOR for JIC-grown *B. napus*. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C.

Supplementary Figure 3. 7. Associative Transcriptomics SNP (A) and GEM (B) results for stem parenchyma area for JIC-grown *B. napus*. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C

Supplementary Figure 3. 8. Associative Transcriptomics SNP (A) and GEM (B) results for stem weight for JIC-grown *B. napus.* Significance of marker associations

is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C

Supplementary Figure 3. 9. Associative Transcriptomics SNP (A) and GEM (B) results for stem outer cortex thickness for JIC-grown *B. napus*. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C

Supplementary Figure 3. 10. Associative Transcriptomics SNP (A) and GEM (B) results for plant height for JIC-grown *B. napus*. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C

Supplementary Figure 3. 11. Associative Transcriptomics SNP results for Fmax (A) and F/V (B) for KWS-2010-grown *B. napus.* Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C

Supplementary Figure 3. 12. Associative Transcriptomics SNP results for second moment of area (A) and Log10 second moment of area (B) for KWS-2010-grown *B. napus*. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C

Supplementary Figure 3. 13. Associative Transcriptomics SNP results for stem outer cortex thickness (A) and stem diameter (B) for KWS-2010-grown *B. napus*. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C

Supplementary Figure 3. 14. Associative Transcriptomics SNP results for MOE (A) and Log10MOE (B) for KWS-2010-grown *B. napus*. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C

Supplementary Figure 3. 15. Associative Transcriptomics SNP results for MOR (A) and Log10MOR (B) for KWS-2010-grown *B. napus*. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C

Supplementary Figure 3. 16. Associative Transcriptomics SNP results for parenchyma area (A) and Log10 parenchyma area (B) for KWS-2010-grown *B. napus*. Significance of marker associations is shown as –Log10P and markers are

plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C

Supplementary Figure 3. 17. Associative Transcriptomics SNP results for stem weight for KWS-2010-grown *B. napus.* Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C.

Supplementary Figure 3. 18. Associative Transcriptomics SNP results for Fmax (A) and F/V (B) for KWS-2011-grown *B. napus*. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C

Supplementary Figure 3. 19. Associative Transcriptomics SNP results for second moment of area (A) and Log10 second moment of area (B) for KWS-2011-grown *B. napus*. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C

Supplementary Figure 3. 20. Associative Transcriptomics SNP results for outer cortex thickness (A) and Log10 outer cortex thickness (B) for KWS-2011-grown *B. napus*. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C

Supplementary Figure 3. 21. Associative Transcriptomics SNP results for MOE (A) and Log10 MOE (B) for KWS-2011-grown *B. napus*. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C

Supplementary Figure 3. 22 Associative Transcriptomics SNP results for MOR (A) and Log10 MOR (B) for KWS-2011-grown *B. napus*. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C

Supplementary Figure 3. 23. Associative Transcriptomics SNP results for parenchyma area (A) and Log10 parenchyma area (B) for KWS-2011-grown *B. napus*. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C

Supplementary Figure 3. 24. Associative Transcriptomics SNP results for stem weight (A) and stem diameter (B) for KWS-2011-grown *B. napus.* Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The two *B. napus* genomes can be seen marked as A and C

Supplementary Figure 4. 1. Associative Transcriptomics SNP (A) and GEM (B) results for Fmax in wheat . Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The three wheat genomes can be seen marked as A, B and D.

Supplementary Figure 4. 2. Associative Transcriptomics SNP (A) and GEM (B) results for F/V in wheat. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The three wheat genomes can be seen marked as A, B and D

Supplementary Figure 4. 3. Associative Transcriptomics SNP (A) and GEM (B) results for MOE in wheat. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The three wheat genomes can be seen marked as A, B and D

Supplementary Figure 4. 4. Associative Transcriptomics SNP (A) and GEM (B) results for MOR in wheat. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The three wheat genomes can be seen marked as A, B and D.

Supplementary Figure 4. 5. Associative Transcriptomics SNP (A) and GEM (B) results for LogMOR in wheat. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The three wheat genomes can be seen marked as A, B and D

Supplementary Figure 4. 6. Associative Transcriptomics SNP (A) and GEM (B) results for outer cortex thickness in wheat. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The three wheat genomes can be seen marked as A, B and D

Supplementary Figure 4. 7. Associative Transcriptomics SNP (A) and GEM (B) results for Log10 outer cortex thickness in wheat. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The three wheat genomes can be seen marked as A, B and D

Supplementary Figure 4. 8. Associative Transcriptomics SNP (A) and GEM (B) results for parenchyma area in wheat. Significance of marker associations is shown

as –Log10P and markers are plotted in pseudomolecule order. The three wheat genomes can be seen marked as A, B and D

Supplementary Figure 4. 9. Associative Transcriptomics SNP (A) and GEM (B) results for stem hollow area in wheat. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The three wheat genomes can be seen marked as A, B and D

Supplementary Figure 4. 10. Associative Transcriptomics SNP (A) and GEM (B) results for plant height in wheat. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The three wheat genomes can be seen marked as A, B and D

Supplementary Figure 4. 11. Associative Transcriptomics SNP (A) and GEM (B) results for main stem threshed weight in wheat. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The three wheat genomes can be seen marked as A, B and D

Supplementary Figure 4. 12. Associative Transcriptomics SNP (A) and GEM (B) results for main stem grain weight in wheat. Significance of marker associations is shown as –Log10P and markers are plotted in pseudomolecule order. The three wheat genomes can be seen marked as A, B and D

LIST OF SUPPLEMENTARY DATA FILES (SEE ATTACHED CD)

Supplementary Data file 2.1-script used in performing stem mechanical strength calculations.

Supplementary Data file 3.1:

- **a)** Summary statistics and ANOVA outputs for mechanical strength and related Traits for JIC-grown *B. napus*
- **b)** Trait summaries, ANOVA and correlations analysis for stem mechanical and related traits for KWS-grown *B. napus* accessions
- **c)** Summary statistics and ANOVA output for *B. napus* lodging simulation experiment

Supplementary Data file 3.2:

- **a)** Determining K value estimate for *B. napus* Associative Transcriptomics panel using the Evanno et al (2005) method
- **b)** SNP data file used in *B. napus* Associative Transcriptomics analyses
- **c)** Unigene transcript abundance data file used in *B. napus* Linear Regress (GEM analyses)
- **d)** Scripts used for Linear Regression analyses and for plotting SNP and GEM results in *B. napus* Associative Transcriptomics analyses
- **e)** *B. napus* pseudomolecule spreadsheet
- **f)** Candidate gene information for stem mechanical strength and related traits in *B. napus.*

Supplementary Data file 3.3:

- **a)** Marker genotype information across 96 *B. napus* ASSYST accessions
- **b)** Trait data summary and ANOVA output across all ASSYST accessions mechanically tested
- **c)** Trait summaries and ANOVA outputs for each marker locus

Supplementary Data file 3.4:

- **a)** Summary data and T-test results for Arabidopsis T-DNA insertion lines
- **b)** FT-IR data output for Arabidopsis T-DNA insertion lines

Supplementary Data file 4.1:

- **a)** Testing for year by year interactions and fitting REML models for mechanical strength and related traits in wheat
- **b)** REML-predicted trait means across 100 wheat accessions

c) Trait summary and AVOVA output for wheat stem lodging simulation experiment

d) Mechanical strength and chemical composition correlation analysis in wheat

Supplementary Data file 4.2:

a) Determining K value estimate for wheat Associative Transcriptomics panel using the Evanno et al (2005) method

b) SNP data file used in wheat Associative Transcriptomics analyses

c) Unigene transcript abundance data file used in wheat Linear Regress (GEM analyses)

d) Scripts used for Linear Regression analyses and for plotting SNP and GEM results in wheat Associative Transcriptomics analyses

e) Wheat pseudomolecule spreadsheet

Supplementary Data file 4.3:

a) Genotype information across 96 WAGTAIL wheat accessions

b) Trait summary data for MOR across 30 WAGTAIL accessions and ANOVA outputs for each marker locus

ACKNOWLEDGEMENTS

I would like to thank everybody who has given their kind support throughout the duration of my PhD project and the preparation of this thesis. My supervisors, Professor Ian Bancroft, Dr Judith Irwin, Dr Andrea Harper and Professor Keith Waldron, have provided valuable guidance throughout my PhD. They have each taught me a great deal, and have played a huge role in my continued development as a scientist. I would particularly like to thank Andrea Harper, who has been victim to a barrage of questions over the past four years and who has always been there to help me through the most stressful times. Dr Martin Trick has also been a great help with his bioinformatic wizardry, and again, a great deal of patience with my many questions. I am also very grateful to the many research and industry members of the IBTI club, who have always shown great interest in my research and have been the source of much interesting discussion. I would also like to thank Colin Morgan for his help with statistical analyses and general great company. Dr Rachel Wells has been a huge help in many ways, and has helped me to become more confident as a scientist. Eleni Soumpourou has also been a great support (and a great supplier of chocolate) and has never failed to put a smile on my face. This PhD project would not have been possible without the help of many plant breeders, particularly those at KWS, Thriplow, as well as the valuable support provided by the JIC horticulture team. Professor James Brown has also been a great help and has always been there to answer the more complex statistical analyses questions. Also, thank you to John Humble for his help in designing and creating the "Humble Pi" contraption. I would like to say a special thank you to Dr Judith Irwin, who very happily took me under her wing following the relocation of my supervisors to York University. Finally, I would like to thank my beautiful friends and family for being proud and always being there to make me smile. I would particularly like to thank Ian Wood, Pauline Haleux, Jack Dumenil and Scott Berry, who have made the past few years some of the most memorable yet. The greatest thanks however, must go to my wonderful Robin. He has always believed in me and has and has never failed to make me feel like the luckiest girl in the world.

OVERVIEW OF THESIS CONTENT

This thesis is concerned with understanding the genetic control of stem mechanical properties and related traits in *Brassica napus* (*B. napus*) and wheat. The first Chapter begins by introducing the two crop species studied. It will then discuss the importance of stem mechanical properties for improved lodging resistance and discuss the potential importance of these traits when looking to breed feedstock crops for lignocellulosic ethanol production. It will then focus on examples of previous work carried out to explore stem mechanical properties in plants. The Chapter concludes with the discussion of the different ways in which genetic markers can be utilised to select for traits of agronomic importance. The second Chapter outlines the general methods used in the present study for both crop species. Methods more specific to a single crop are described in Chapters 3 and 4 for *B. napus* and wheat respectively. Chapter three describes the work carried out to explore stem mechanical strength in *B. napus.* It first explores the level of existing variation for stem strength and related traits within a selected panel of 79 *B. napus* accessions and then focuses on the genetic control of these traits using Associative Transcriptomics. Based on conserved synteny between *B. napus* and Arabidopsis, potential candidate genes in close linkage to the associated markers are discussed. The following sections of this Chapter describe experiments aimed at marker and candidate gene validation. Chapter 3 closes with an overall discussion of the results obtained in *B. napus*. Chapter 4 focuses on the results obtained for wheat. As for *B. napus,* it begins by exploring variation in stem mechanical strength and related traits across a panel of 100 wheat accessions and looks at the relationship between these traits. It then explores genetic markers detected for a subset of the most important traits identified through Associative Transcriptomics. The final results section describes a marker validation study to evaluate the potential value of these markers in crop improvement. Chapter 4 ends with a general discussion of the results obtained for wheat. The final Chapter is a general discussion. A number of interesting comparisons are made between the genetic control of stem mechanical strength in *B. napus* and wheat. It discusses the merits and limitations of the methods employed and explores potential areas of improvement for future research.

Chapter 1. Introduction

1.1. *B. NAPUS* AND WHEAT-CROPS OF GLOBAL IMPORTANCE

B. napus is a recently formed allopolyploid species (5,000-10,000 mya), resulting from interspecific hybridisation events between two diploid species, *B. rapa* (contributing the A genome) and *B. oleracea* (contributing the C genome). Comparative linkage mapping (Lagercrantz and Lydiate, 1996) and, more recently, physical mapping studies (O'Neill and Bancroft, 2000) have shown that the parental genomes of *B. napus* are extensively triplicated (indicative of a hexaploid ancestor). This adds further to the complex genome structure of *B. napus*, where for every gene present in the genome of *Arabidopsis thaliana* (hereafter referred to as Arabidopsis), we may expect to find up to six copies (Trick et al., 2009; Wang et al., 2011). *B. napus*, like the extensively studied model plant species, Arabidopsis, belongs to the family Brassicaceae. *B. napus* is a crop species of global economic importance. It comprises several crop types grown for seed, leaf and roots. Oilseed rape (winter, spring and Chinese semi winter types), is widely grown for the high oil content of the seed and is second only to soybean as an oil crop (Bancroft et al., 2011). Kale, including Siberian kale, Japanese kale and animal fodder types, are grown for their edible leaves, and rutabagas are grown for the turnip-like root mass which is eaten as a vegetable (Harper et al., 2012).

Triticum aestivum, or bread wheat, is a hexaploid crop thought to have arisen from two independent hybridisation events. The initial hybridisation event, believed to have occurred approximately 10,000 years ago, is thought to have taken place between two diploid progenitor species, *T. urartu* (contributing the A genome) and *T. speltoides* (contributing the B genome) (Brenchley et al., 2012). This hybridisation event gave rise to a tetraploid species. The hexaploid wheat that we know today is thought to have arisen from a further hybridation event, taking place between the newly formed tetraploid species (*T. urata* x *T. speltoides*) and a third diploid species, *T. tauschii* (contributing the D genome) (Eckardt, 2010; Peng et al., 2011). Bread wheat (hereafter referred to as wheat) is expected to have one of the largest genomes of higher plant species. This large size is, at least in part, due to the presence of large amounts of repetitive sequence (Choulet et al., 2010). Wheat, together with rice and sorghum, is considered to be one of the most important staple

cereal crops grown world-wide. Bread wheat alone is estimated to account for as much as 20% of human dietary intake worldwide (Agriculture, 2012).

Given the global importance of these crop species, and the expected continual rise in population growth, it is important that breeders are able to continue to improve the yield and general agronomic performance of these crops. One key aspect of agronomic performance which can have a devastating effect on crop yield is lodging.

1.2 LODGING

1.2.1 Lodging resistance - an important agronomic trait

Lodging can be defined as the permanent displacement of a crop plant from its usually upright position (Berry and Spink, 2012). This phenomenon is divided into two categories: lodging caused by anchorage failure, also known as root lodging; and lodging caused by stem breakage (typically at the base of the stem), also referred to as stem lodging, or "brackling" (van Delden et al., 2010). Lodging is a complex trait which is the result of a combination of environmental, agronomic and genetic factors. Lodging is a phenomenon of great concern in both *B. napus* and wheat, and is a key contributor to yield losses in both species (Islam and Evans, 1994; Berry and Spink, 2012). For example, it is estimated that lodging of UK-grown oilseed rape, can cause yield losses of up to 50% (Baylis and Wright, 1990; Armstrong and Nicol, 1991). In wheat, it is estimated that, in years when lodging is a problem, it effects on average 20% of UK-grown winter wheat. Estimates suggest that on average an annual loss of ~£50M can be expected resulting from the lodging of UK winter wheat alone (Berry and Spink, 2012).

The yield loss caused by lodging is primarily due to the reduced photosynthetic capability of the crop, where penetration of light to the lower green tissue is hindered (Berry and Spink, 2012). Lodging has also been linked with reduced product quality in both *B. napus* and wheat crops. In wheat for example, lodging reduces the overall flour yield of the crop. Lodging has also been found to increase the ash content of the flour, which may result in a reduced milling score (Weibel and Pendleton, 1964). In oilseed rape, lodging can result in reduced the oil content of the seed and has been linked with increased seed glucosinolate levels (Baylis and Wright, 1990). Finally, the irregular canopy caused by lodging can make mechanical harvesting very

difficult and not only results in increased harvest time, but also increased cost (Wang et al., 2012).

1.2.2 Current strategies implemented to reduce lodging susceptibility

Currently, the two most common methods used in controlling lodging are dwarfing genes and plant growth regulators (PGRs). However, although both of these methods have helped towards reducing lodging severity, they have not been successful in completely eradicating the problem, and it has been questioned as to whether they will continue to be the most effective solution in years to come (Flintham et al., 1997).There is for example, evidence to suggest that there may be a yield penalty related to dwarfing in wheat when grown in suboptimal conditions (Uddin and Marshall, 1989). Furthermore, studies have shown that the minimum height to which wheat can grow whilst maintaining high yield, is between 0.7 and 1.0 m. Given that many current wheat accessions already fall within this height range, the scope for future improvements through increased dwarfing is limited (Berry et al., 2007). PGRs on the other hand are very expensive and as it is difficult to predict when they will be needed, they are often used as a precaution, which incurs unnecessary costs (Garthwaite and Hart, 1994).

With the threat of climate change and an ever-growing global population size, breeders remain under continual and increasing pressure to increase crop yields. However, the desired increase in seed/grain weight in these higher yielding crops, will add further strain to the supporting body of the plant which may result in an increased risk of lodging susceptibility. Given this, and the described limitations of the methods currently employed, it is important that an alternative (or indeed complimentary) approach to improving lodging resistance is employed.

1.3 STEM MECHANICAL PROPERTIES

1.3.1 Improved stem mechanical properties as an alternative approach to lodging resistance

A number of studies have suggested that lodging resistance could be improved through maximising the mechanical strength of the crop stem. For example, studies have found that there is a positive correlation between lodging resistance and stem mechanical strength in wheat (Atkins, 1938; Crook and Ennos, 1994). Given the proposed increases in wheat yield, a lodging resistant ideotype is considered to be

essential, and that one key focus of the development of this ideotype should be the improvement of the mechanical strength of the plant stem (Berry et al., 2007). It has been suggested that in order to achieve this, the improvement of stem structural, morphological and chemical attributes will be necessary (Berry et al., 2007). Although considered to be a trait of agronomic importance, stem mechanical strength has been largely neglected in crop improvement. This is, at least in part, because the importance of stem strength in improving lodging resistance has only been realised in recent years. Also, the selection of crops with improved mechanical properties is complex (HGCA, 2003). A number of studies have looked into what contributes to the mechanical strength of plant stems. This includes research focused towards the detection of potential morphological and anatomical traits which may contribute to overall stem strength, and also studies which have worked to underpin the genetic components of stem mechanical strength. In addition, more recently, a number of studies have explored the role of cell wall composition in contributing to stem mechanical strength. An overview of some of the most relevant research is given below.

1.3.2 The role of anatomical and morphological stem traits in contributing to stem mechanical strength and lodging resistance

A number of studies have focused on detecting anatomical and morphological traits which may be used as phenotypic markers for the selection of improved stem mechanical strength. Studies looking into lodging resistance in rice suggest that stem diameter may positively contribute to overall stem strength and could therefore serve as a proxy for lodging resistance (Kashiwagi et al., 2008). However this is not the case in work carried out in barley, where stem diameter was found to negatively correlate with stem strength (Dunn and Briggs, 1989). Previous studies indicate that stem outer cortex, or wall thickness may also be important (Wang et al., 2006). In addition, internode length has been found to be important in crops such as wheat, rye and barley (Kaack and Schwarz, 2001), with an increase in internode length (particularly the lower internodes) correlating with increased lodging risk. Several studies have looked at the role of the parenchyma, or pith tissue, in determining stem mechanical properties. While some studies have found that the amount of parenchyma tissue is negatively correlated to stem strength (Wang et al., 2006), other studies suggest that this tissue may have a role in maintaining the strong

cylindrical shape of the stem, therefore suggesting an overall positive contribution to stem strength (Spatz et al., 1998).

1.3.3 The importance of cell wall composition in determining stem mechanical strength

More recently, several studies have looked at the potential role of cell wall composition and structure in maintaining the structural integrity of the plant stem. The plant cell wall is made up of four main components. The first of these, considered to be the most characteristic component of the plant cell wall, is cellulose (Keegstra, 2010). Cellulose is made up of chains of beta-1-4 linked glucan subunits. Cellulose is found in both crystalline and non-crystalline form. Crystaline cellulose is formed following the joining of glucan chains with one another via hydrogen bonds (Keegstra, 2010; Park et al., 2010). The plant cell wall also contains hemicelluloses which form linkages with cellulose and other cell wall components (Eckardt, 2008). There are a number of different hemicellulosic components including xyloglucans, glucomannans and xylans (Keegstra, 2010). Xylans are the most common hemicelluloses found in the plant cell walls of most species. They are made up of a linear backbone of xylose which may be decorated with different side-groups, such as acetyl groups (Han et al., 2012). Pectin is also a common feature of plant cell walls. It is typically considered that homogalacturonan is the most abundant of the cell wall pectins in plants (Atmodjo et al., 2011). Pectin forms tight linkages with hemicellulose, helped by the presence of the hemicellulosic side chains. Pectic polysaccharides are made up of galacturonic acid units, which may be decorated with various side groups, including methyl-ester groups (Mohnen, 2008; Harholt et al., 2010). Figure 1.1 gives an overview of the diverse types of decorations seen in pectins. Together, pectin and hemicellulose form a matrix within which cellulose is embedded (Figure 1.2).

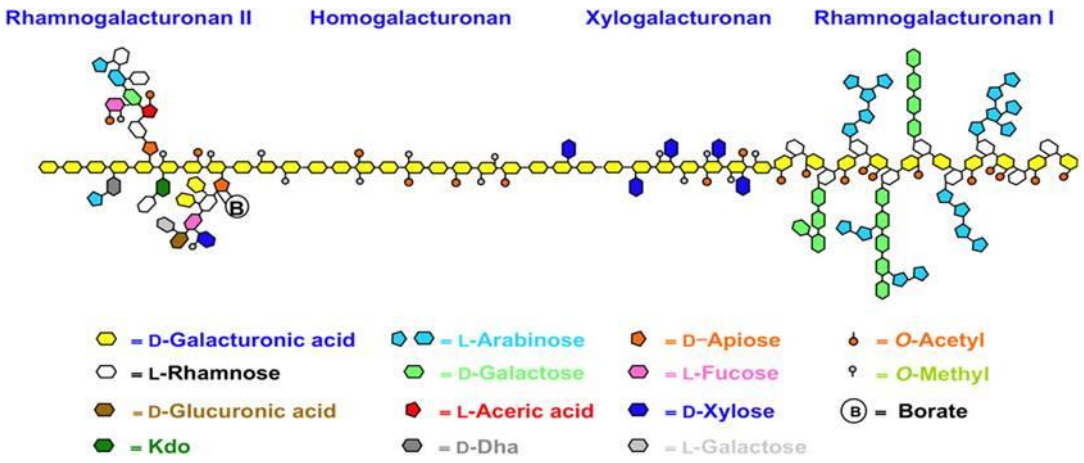

Figure 1. 1. Variation in pectin structure seen in plant cell walls (Harholt et al., 2010).

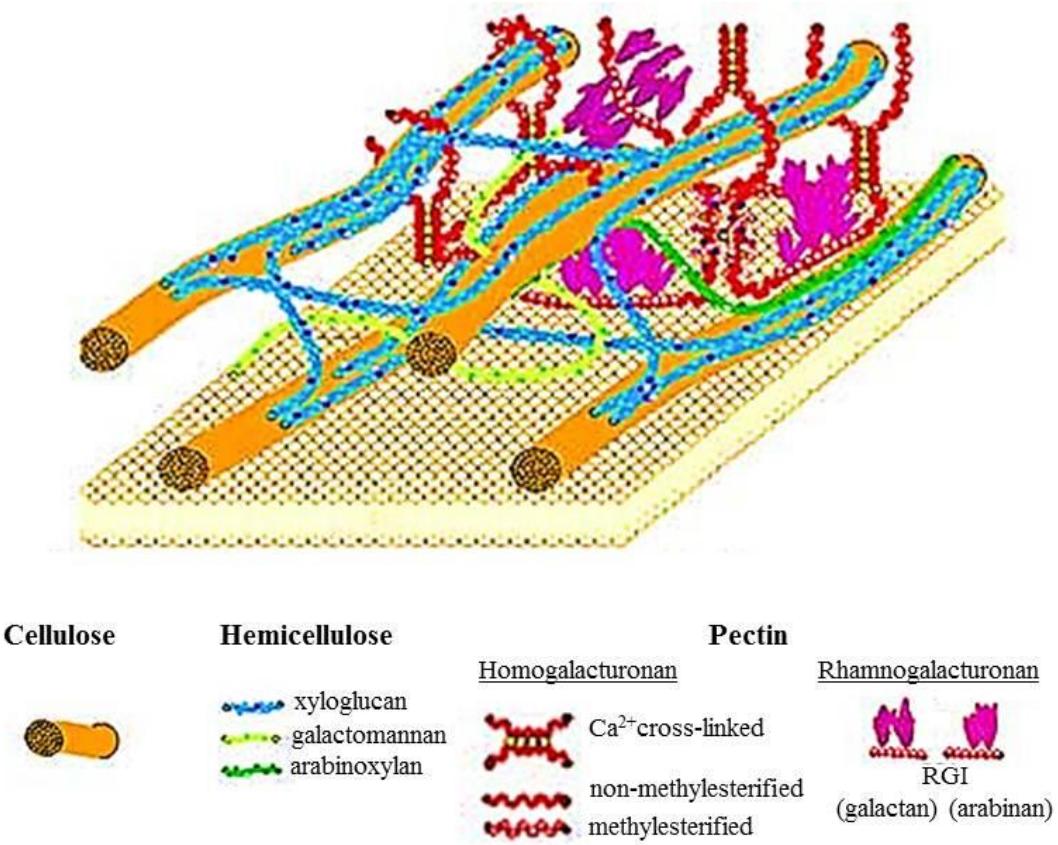

Figure 1. 2. The complex interaction between cellulosic, hemicellulosic and pectic polysaccharides in the plant cell wall (image adapted from: http://pmb.berkeley.edu/profile/mpauly).

This matrix is further reinforced by the impregnation of lignin. Lignin is a complex phenylpropanoid polymer found predominantly in the xylem, fiber cells and interfascicular fibers of the plant stem (Zhong et al., 2000) (Figure 1.3). Together these four major components make up a very complex and structurally very strong, yet dynamic cell wall (Keegstra, 2010).

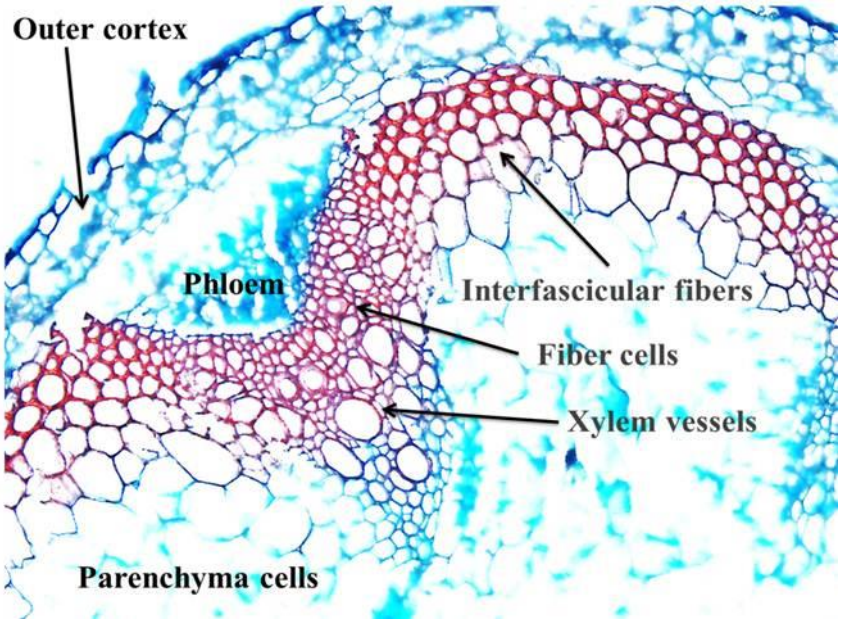

Figure 1. 3. A cross section of a Col-0 Arabidopsis stem showing highly lignified cells in the interfascicular fiber cells, fiber cells and xylem vessels. The lignin-rich components can be seen stained pink with phloroglucinol.

Previous studies suggest that lignin may have an important role in providing structural stability and rigidity to the plant stem (Tien and Tu, 1987; Downes and Turvey, 1990). Lignin is known to be important in providing mechanical support to xylem vessels and is thought to be vital in allowing these vessels to withstand the high internal pressures created during water and mineral transportation (Kitin et al., 2010). Several lignin mutants have been studied in Arabidopsis. One of these, *irregular xylem 4*, was found to have a 50% reduction in lignin content. This mutant displays a collapsed xylem phenotype and exhibits a pendant stem phenotype, suggesting that the lignification of these vessels is an important contributor to overall stem strength. (Jones et al., 2001). Figure 1.4 shows the reduced lignification seen in the xylem vessels of this mutant.

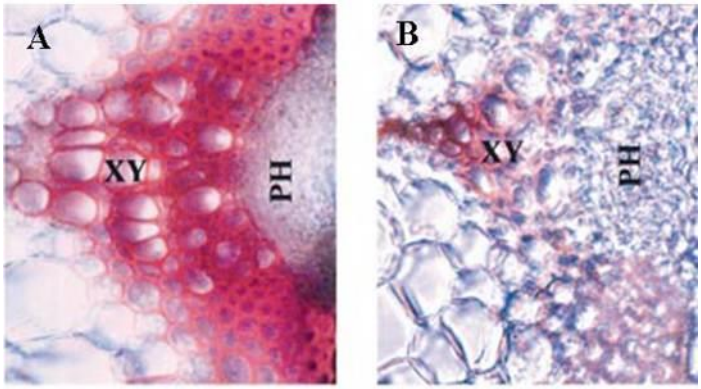

Figure 1. 4. Xylem phenotype of the *irregular xylem 4* Arabidopsis mutant (B) in comparison to WT (A). Highly lignified tissues can be seen in pink. PH indicates the phloem cells and XY indicates the xylem cells (Jones et al., 2001).

A second Arabidopsis mutant exhibiting a pendant stem phenotype is *interfascicular fibreless 1* (*ifl1*). *IFL1* is predicted to encode a Homeodomain-leucine zipper protein and plants defective for this gene exhibit abnormal differentiation of interfascicular fiber cells (Figure 1.5A). The lack of these lignin-rich structures results in a pendant stem phenotype (Figure1.5B). These phenotypes are successfully rescued following complementation of the mutant with a functional copy of *IFL1* (Figure 1.5C and D) (Zhong and Ye, 1999). More recently, it has been suggested that the phenotypes observed in this mutant, may be related to defective auxin polar transport, which is known to be important in xylem and interfascicular fiber cell development and lignin deposition (Zhong and Ye, 2001).

Figure 1. 5. Defective interfascicular fiber cells (marked with a white arrow) of the *interfascicular fiberless-1* Arabidopsis mutant (A). These mutants exhibit a pendant stem phenotype (B). These phenotypes are rescued following complementation with a function copy of *IFL1* (C and D).

In contrast, work conducted in *Zea mays L* showed that hybrids with greater lignin levels were more susceptible to stem fracture in comparison to low lignin types (Li, 1997).

In addition to research focused on the importance of lignin in determining stem mechanical properties, there is also evidence to suggest that cellulose may too have an important role to play (Wang et al., 2006). Unlike the *irx4* mutant in Arabidopsis, *irx1-3*, exhibit no significant changes in lignin content and yet there is a clear

reduction in stem mechanical strength. These mutants have significantly reduced cellulose content in comparison to WT plants (Wang et al., 2006). Other studies suggest that it is not the overall cellulose content that provides the mechanical support, but more specifically the degree of cellulose crystallinity which is important (Burton et al., 2010).

A number of studies have also focused on the role of the finer cell wall structures in contributing to stem mechanical strength. There is evidence to suggest that the methylesterification state of the cell wall pectin, homogalacturonan (see Figure 1.1 for an illustration of methyl groups in homogalacturonan), may have a vital role to play in the maintenance of stem mechanical integrity (Hongo et al., 2012). The exact mechanisms through which demethylesterification (the removal of methyl ester side-groups) alters cell wall integrity is not clear and there are theories which suggest both an increasing and decreasing effect on strength. The first proposed role is that following demethylesterification, the newly exposed galacturonic acid residues are able to form intermolecular Ca linkages (Figure 1.2), which forms a rigid gel-like structure. This is thought to induce strengthening of the cell wall. In contrast, it is also thought that demethylesterification can render the homogalacturonan more susceptible to hydrolysis through the action of cell wall pectin-degrading enzymes, which would likely have the opposite effect on the cell wall mechanical strength (Hongo et al., 2012; Müller et al., 2013). Pectin demethylesterification is carried out by a family of enzymes called pectin methylesterases (PMEs). The loss-of-function Arabidopsis mutant, *pme-35,* exhibits reduced levels of homogalacturonan demethylesterification. This mutant is also unable to maintain an upright growth habit, suggesting that the methylesterification state of the cell wall in Arabidopsis is important in determining stem mechanical strength (Hongo et al., 2012).

It is clear that stem mechanical strength is an incredibly complex trait which may be the result of morphological, anatomical and chemical components. It is also clear, that despite much effort, our understanding of what contributes to mechanical strength is poor, with studies showing conflicting results and with results varying greatly between species. If we are to successfully breed crops with improved stem mechanical properties, a deeper understanding of which components are of greatest importance in the crop species of interest will be needed.

1.4 STEM LODGING SUSCEPTIBILITY-AN IMPORTANT CONSIDERATION WHEN BREEDING FEEDSTOCK CROPS FOR LIGNOCELLULOSIC ETHANOL PRODUCTION

In recent years, there has been increasing recognition of the opportunity to exploit crop residues, such as the straw of *B. napus* and wheat crops, as a biomass resource for the production of lignocellulosic ethanol. Not only would the use of such biomass eliminate the "food vs fuel" dilemma seen with first generation bioethanol production, but it would also add value to the selected feedstock crop. One of the key limiting factors in lignocellulosic ethanol production is the recalcitrance of the lignocellulosic matrix to saccharification (Kim and Dale, 2004). To produce lignocellulosic ethanol, the sugar-containing cell wall components, cellulose and the hemicelluloses, must be broken down by hydrolysing enzymes. One cell wall component which reduces the efficiency of this hydrolysis is lignin, which forms a rigid barrier around the hemicellulosic and cellulosic components. This barrier impedes hydrolysis by blocking the hydrolysing enzymes from their target substrate. For this reason, there is great interest in the production of feedstock crops with reduced lignin content. Recent work in transgenic poplar found that down regulation of cinnamoyl-CoA reductase, a gene involved in lignin biosynthesis, resulted in an increase in saccharification efficiency (Van Acker et al., 2014). Research also suggests that saccharification efficiency may be improved through the reduction of cellulose crystallinity (Harris et al., 2012) and possibly by a reduction in xylan branching (Mortimer et al., 2010). However, as previously discussed, cell wall composition may have an integral role to play in determining stem mechanical properties. It is therefore possible that an increase in processability, through altered cell wall composition, could come at the cost of reduced stem strength. If lignocellulosic ethanol from food crop residues is to be considered as a feasible fuel source in the future, it is vital that high ethanol yields can be achieved with no resulting cost in agronomic performance. An improved understanding of what contributes to stem mechanical strength is therefore of great relevance when breeding feedstock crops for lignocellulosic ethanol production.

1.5 GENETIC MARKERS FOR MARKER ASSISTED SELECTION

1.5.1 Genetic markers for improved stem mechanical properties

Although the use of phenotypic markers may seem desirable, they are limited to only easily scored phenotypes, such as stem diameter, internode length and height for example. There are many other components which may be influencing stem mechanical strength which cannot be easily screened, such as lignin content or cellulose crystallinity. Furthermore, the selection of traits based on phenotypic markers may be erroneous due to the effect that environmental conditions may have on the trait. An alternative method of trait selection is through the use of variation at the genetic level. The past 40 years has seen tremendous efforts and success in the development of genome sequencing technologies which have been paramount in the advances made in the utilisation of variation at the genomic sequence level. The most widely used method of sequencing continues to be the Sanger sequencing method which, since its development in 1977, has been continually optimized (Ellegren, 2008). This 'chain-termination' approach, has allowed much progress to be made in genome sequencing, and has been named the 'gold standard' technology (Gharizadeh et al., 2006). However, the throughput of this method is modest and cost per sequenced base is relatively high (Kircher and Kelso, 2010). Because of these limitations, this method of sequencing is largely limited to small-scale sequencing projects and may not be an option for many smaller research teams, where money may be an issue. In an attempt to overcome such limitations, much work has been focused towards the development of so-called 'next generation' sequencing technologies, which promise higher throughput at lower cost (Kircher and Kelso, 2010). An example of this can be seen with the Illumina sequencing method. This technology utilises a sequencing approach similar to that of Sanger sequencing - the chain termination method. Illumina sequencing however, firstly involves the binding of DNA fragments to a primer-filled slide. Through bridge-amplification, these DNA strands are multiplied, forming dense clusters of a single DNA sequence (Mardis, 2008). Fluorescently labelled Deoxyribonucleotide triphosphates (dNTPs) are then attached and the resulting fluorescence of the clusters imaged to determine which nucleotide base has been added. This high-throughput technology allows the sequencing of 50Gb per day (Caporaso et al., 2012). Although this method may suffer somewhat from short read lengths of just 100 nucleotides (much shorter than

that achieved through Sanger sequencing), Illumina sequencing is possible at the cost of just $0.5/Mb, making it a very attractive choice, especially for large-scale projects (Kircher and Kelso, 2010). The low cost and high throughput nature of such next generation sequencing methods, has allowed for great advances in Single Nucleotide Polymorphism (SNP) discovery, where individuals within a population/species can be sequenced and polymorphisms between them detected.

For such polymorphisms to be detected, a reference genome sequence is required, to which the newly sequenced genomes can be aligned. Unfortunately, there is currently no such reference sequence for the majority of the important crop species. Recent work carried out by Trick et al (2009) has resulted in the development of a novel method of SNP discovery, focusing on transcribed sequences, in species for which a genome sequence in not currently available. This work was firstly conducted to improve SNP detection in *B. napus*. A set of 94,558 unigenes, put together using ESTs from several Brassica species, were used to act as a proxy reference genome. Through use of a linkage map and sequence information from genome scaffolds of *B. rapa* (Wang et al., 2011) and *B. oleracea* (Liu et al., 2014) as well as conserved synteny with Arabidopsis, these unigene sequences were arranged into a hypothetical gene order, making up what has been termed a "pseudomolecule". Through Illumina sequencing, transcriptome sequences for two *B. napus* double haploid (DH) accessions, Tapidor and Ningyou 7, were obtained. The alignment of the sequence reads obtained for Tapidor and Ningyou 7 against the pseudomolecule, identified 41,593 SNPs. SNP detection in a polyploidy species (although simplified in this study through the use of DH accessions), can be complicated by the presence of what have been termed "interhomoeologue polymorphisms" (IHPs). IHPs represent variation between homoeologues and do not represent the useful, between accession variations, which are most valuable in mapping studies. Due to the high level of sequence similarity between the A and C genome homoeologues in *B. napus* (known to vary on average by only 3.5 %), in most cases, the SNPs identified were in the form of what has been termed "hemi-SNPs". Hemi-SNPs represent useful variation between accessions, coming from a single genome. Due to sequence similarity between genomes, it was not possible to assign this variation to a specific genome (Trick et al., 2009). Despite this, the detection of such a high level of useful variation in this complex species holds great promise for future studies in *B. napus* genetics.

Figure 1.6 provides an illustration of the different types of allelic variation identified in this study.

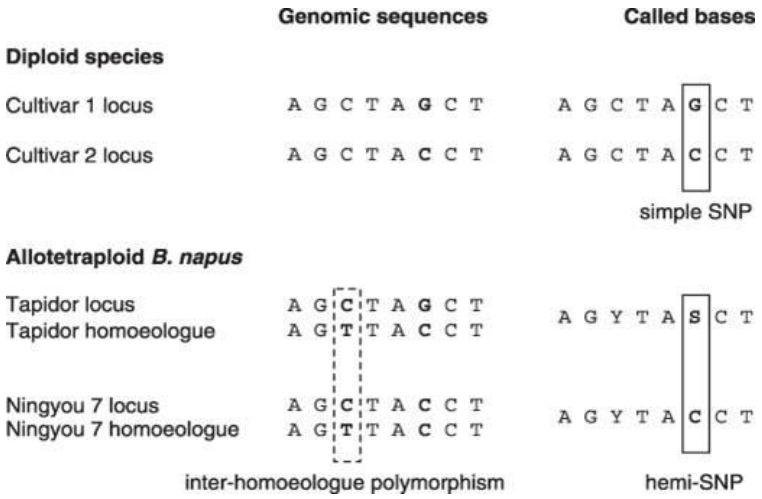

Figure 1. 6. Different types of marker variation identified in *B. napus* (Trick et al., 2009).

A very similar approach to that described for *B. napus*, has more recently been taken in hexaploid wheat (Harper et al., Under review).This involved the construction of a reference pseudomolecule from the sequencing of the three progenitor ancestral genomes (comprising the A genome, inherited from *Triticum urartu*; the B genome inherited from a close relative of *Aegilops speltoides* and the D genome from *Aegilops tauschii* (Chantret et al., 2005)) of hexaploid wheat through Illumina transcriptome sequencing. The utilisation of a high density linkage map (Paragon x Chinese Spring) together with information regarding *Brachypodium distachyon* (a closely related model grass species) gene order, allowed the assembly of sequence reads into a hypothetical gene order. This transcriptome reference sequence consisted of 105,069, 132,363 and 85,296 unigenes representing the A, B and D genomes of hexaploid wheat respectively. Using sequence similarities with rice, *Brachypodium*, Sorghum and Arabidopsis, where possible, unigenes models were annotated with probable gene function (Harper et al., Under review). This reference transcriptome provides a powerful tool which may be utilised in future genetic studies in this important crop species.

1.5.2 Bridging the gap between phenotype and genotype

Following the identification of genetic polymorphisms, such as SNPs, between individuals, it becomes possible to seek associations between this genetic variation and variation in an agronomic trait of interest. One widely used method for the detection of such associations is Quantitative Trait Loci (QTL) analysis. This method relies on the coinheritance of markers and QTL (genomic region containing a causative gene) due to linkage (through lack of recombination between them) (Kearsey, 1998). This method has been utilised in studies to uncover QTL contributing to complex traits, including stem strength. Hai et al (2005) carried out QTL analysis using a DH population generated from a cross between two winter wheat accessions, known to differ in stem strength properties, as well as traits thought to be linked with mechanical strength such as stem diameter and stem wall thickness. This study resulted in the detection of QTL for stem strength and related traits (Hai et al., 2005). QTL contributing to the inheritance of stem diameter and stem wall thickness were also detected in a QTL analysis study in soybean. This latter analysis was based on a mapping population derived from the crossing of two parental genotypes known to vary for lodging resistance. This study identified two QTLs common for stem strength, stem diameter and number of nodes (Chen et al., 2011). Similar studies have also been conducted in rice (Ping Mu, 2004).

1.5.3 Associative Transcriptomics as an alternative approach

In recent years, the use of association mapping in uncovering the genetic control of complex traits has been considered as an alternative approach, or indeed a complimentary approach, to QTL analysis. To date, this method has been used widely only in medical genetics, but has more recently been used to study complex traits in plant populations (Donnelly, 2008). There are two main strategies used in association studies to identify trait-controlling loci, 1) genome wide association studies (GWAS) and 2) the candidate gene approach. GWAS, involves the scanning of the entire genome in search of polymorphic genetic loci which are associated with a trait of interest. This approach requires no previous knowledge of what may contribute to the trait. Candidate gene association analysis however, is a more hypothesis driven approach, which specifically targets genetic markers surrounding loci which, based on existing knowledge, may be important in contributing to the trait of interest (Long et al., 2004). For example, in a candidate gene approach

looking for genes contributing to stem strength, the researcher may target markers known to be associated with lignin biosynthesis, due to the existence of evidence which suggests that lignin may be a key contributor in determining plant stem strength.

Unlike QTL analysis, association studies make use of panels made up of existing germplasm. This has several advantages (Skøt et al., 2005). Not only does the use of natural germplasm eliminate the need to establish artificial mapping populations, which in QTL analysis can be a laborious and expensive task (Myles et al., 2009; Hall et al., 2010), but it also provides a rich source of variation to explore (Zhu et al., 2008). This is in contrast to QTL analysis, where the variation available is limited to that exhibited between the parents chosen for the bi-parental cross, and which is unlikely to be representative of the variation available in the species under study (Zhu et al., 2008; Myles et al., 2009).

The high level of variation existing within an association mapping panel maximises the opportunity for analysing multiple traits which may be of interest. For example, when looking at a complex trait such as stem mechanical properties, it would be advantageous to not only look into mechanical strength, as determined through mechanical testing, but to also explore factors such as chemical composition and stem structural properties. In this way, it is possible to build a much bigger picture surrounding the genetic control of the trait of interest. It may also be advantageous to look at the way in which the genetic control of the trait of interest is related to genetic components of other important traits, such as crop yield or disease resistance for example. In this way, it becomes possible to build an entire crop ideotype which can be tailored to meet the needs of the breeder.

One particularly promising advantage of association mapping in comparison to the QTL analysis approach, is the much higher resolution with which trait-controlling loci can be located (Yu and Buckler, 2006). Natural populations have undergone many rounds of historic recombination, a process which reduces the level of non-random association, or linkage disequilibrium (LD), between alleles (Myles et al., 2009). This results in a genome made up of many small blocks, containing only highly associated alleles. These are said to be in high LD. This means that when a marker shows an association with a phenotype of interest, it is more likely that the

marker is located within close proximity to the causative allele (Mandel et al., 2013). With a QTL mapping population however, a very limited number of recombination events would have taken place since the establishment of the population. For this reason, when an association is found between a marker allele and the trait of interest, the causative allele cannot be accurately pin-pointed, and is typically localised to a large genomic region of 10-20 cM (Zhu et al., 2008).

A number of different measures of LD have been developed. Two commonly used methods of measuring LD can be seen with R^2 and D which are calculated according to the following equations (Soto-Cerda and Cloutier, 2012).

$$r^2 = \frac{D^2}{P_A P_a P_B \, P_b} \qquad D = P_{AB} - P_A P_B$$

where:

P_{AB} = the probability of observing the A and B segregating together

P_A P_B = probability of seeing A and B segregating independently

$P_A P_a P_B P_b$ = Combined probability of all allele states segregating independently

As can be seen from the above equations, both D and R^2 calculate LD by assessing the difference between the rate at which you see alleles at two loci segregating together and the frequency at which they are observed independently. If two alleles are seen to segregate together more often that would be expected by chance, these loci are said to be in LD. Of these two measures of LD, R^2 is considered to be the most appropriate for use in association studies. This is because this measure of LD also takes into account the frequency at which the alleles are present within the population, thus eliminating bias caused by unequal allele frequencies at each locus (Gupta et al., 2005).

To allow for successful association mapping, it is important that many markers are identified across the genome. This will improve the chances of a marker closely located to the causative allele, being detected, and therefore show a high association with the phenotype of interest. The resolution, with which causative alleles can be located using association mapping, will depend largely on the extent of LD between marker alleles (Myles et al., 2009; Reif et al., 2010). If LD is very extensive,

association mapping may not be very successful, and in this case, a QTL analysis approach may be more suitable (Myles et al., 2009). It is therefore important that the extent of genome-wide LD is assessed before association mapping is initiated (Myles et al., 2009).

1.5.4 Factors influencing LD

We have already seen that LD is broken down through recombination. For this reason, it is important that a diverse set of individuals are selected for association mapping, where LD decay is positively related to the amount of time since the most recent common ancestor. The size of the region over which LD extends therefore also depends on the breeding history of the individuals. Self-pollinating species for example, are more likely to exhibit more extensive LD, due to high relatedness (resulting in a decrease in allelic variation) (Zhu et al., 2008; Myles et al., 2009). The random fluctuation of allele frequencies, genetic drift, is also known to have an effect. This is most prominent in small populations, where it is more likely that rare allele variants will be lost from the population (Ohta, 1982). Another factor to consider is selection. If the inheritance of a particular allelic variant gives a selective advantage, the frequency of the allele will increase in the population and may, over time, become fixed. Through a so-called 'hitch-hiking' effect, marker alleles closely located to the allele which is being selected for, will be pulled along, and as a result also become fixed, therefore increasing LD across that region (Gaut and Long, 2003; Pfaffelhuber et al., 2008). LD which may more commonly exist between closely located alleles can also exist between those distantly located or even alleles on completely different chromosomes. This situation may indicate epistasis, where the interaction of two or more loci is required for the inheritance of a particular phenotype. Of all factors known to influence LD, population structure and relatedness have received most attention with regards to association studies. This is due to the realisation that these factors can lead to false-positive associations between markers and the trait of interest (Hall et al., 2010). If in an association mapping panel, there are closely related and/or non-random mating individuals, more common alleles will be found between these individuals. In situations where the sharing of common alleles is correlated with trait variation, associations may be found. However, although in some cases these alleles may indeed represent true associations with the trait of interest, it is likely that many will instead represent the

shift in allele frequency caused by relatedness and non-random mating of the individuals in question (Myles et al., 2009).

Both population structure and relatedness are very likely to exist in plant populations, especially when looking at crop species which in some cases have undergone extensive selection and local adaptation (Yu and Buckler, 2006). A number of different statistical approaches have been developed which aim to reduce the confounding effect of population structure and relatedness within association studies, a few examples of which will now be discussed (Zhang et al., 2008).

1.5.5 A General Linear Model approach

One method used in association studies is the General Linear Model. This model allows the detection of trait-marker associations while also taking into account broad-scale population structure between genotypes. This is achieved through the implementation of Bayesian models employed by programs such as STRUCTURE or principle components analysis (PCA) (Bradbury et al., 2007) which are able to apportion a set of genotypes into different groups describing shared genome ancestry. STRUCTURE, is a computer program which has been designed specifically to deal with statistics related to genetic studies (Lee et al., 2009). This method, based on allele frequencies calculated from marker data, estimates the likelihood that the data can be sub-divided into a given number of populations (K). The researcher can then pick the K value which gives the highest likelihood score. This information is then be used to generate a Q-matrix, which describes what proportion of ancestry can be assigned to each of the K clusters (Pritchard et al., 2000).

The use of PCA has in a number of cases, been considered a preferable choice for inferring population structure. This is due mainly to the ease and speed with which the analysis can be performed (Lee et al., 2009). This method reduces the data down to a few principle components, which explain most of the variation within the complete data set. However, this method cannot take into account real prior-information of population dynamics such as admixture and the interpretation of the number of clusters can be heavily subjective, which may be considered as a downfall of this approach (Lee et al., 2009).

1.5.6 A Mixed Linear Model approach

More recently, it has been suggested that a mixed-model approach may be more effective in accounting for false positives caused by population structure and relatedness. A mixed-model approach not only incorporates population structure information (in form of the previously described Q-matrix), but also takes into account pairwise relatedness (fitted into the model as a random effect) (Bradbury et al., 2007). Relatedness may be estimated through the calculation of inbreeding coefficients from pedigree information. The success of this approach will however depend largely on the accuracy of pedigree information, which in some cases may be difficult to predict. Alternatively, pair-wise relatedness between individuals, based on information taken from markers throughout the genome, may be used. This measure of kinship (K), together with the population structure information (Q), provides a much more detailed description of the relationships between individuals within a given population, allowing much improved correction for false-positive associations (Bradbury et al., 2007). Both the Q and K matrices, along with trait and SNP data, can be fed into a mixed-linear model using the computer program, TASSEL, to allow for association modelling. The interpretation of the results obtained from TASSEL, allows the researcher to identify any associations between marker alleles and the trait of interest (corrected for false-positives) (Bradbury et al., 2007).

1.5.7 Associative Transcriptomics-combining the power of SNP and gene expression marker (GEM) variation

Recent work carried out by Harper et al (2012) describes a novel association mapping method name Associative Transcriptomics. This method, although in many ways similar to the previously described association mapping methods, incorporates some unique features which have been found to promote increased mapping power, especially in crops where genome complexity may prove a hindrance. Associative Transcriptomics utilises variation at the transcriptome sequence level. The use of transcriptome sequencing in place of genome sequencing, although unable to provide information regarding non-coding regions, is advantageous for several reasons. As previously discussed, the genomes of many crop species are large and complex which can make whole-genome sequencing logistically very challenging and expensive (Trick et al., 2009). The use of transcriptome sequence in place of genomic provides a way through which the scale of the sequencing project may be

reduced while still capturing a high level of what may be considered as the most functionally important regions, the coding sequence (Trick et al., 2009). One particularly striking feature of Associative Transcriptomics, is the combined use of both SNP marker variation and variation in gene expression. By taking the transcriptome sequencing approach, it is possible to utilise sequence read depth as a measure of gene expression. This approach, named the Gene Expression Marker (GEM) analysis, allows for the detection of associations between gene expression and the trait of interest (Harper et al., 2012). This method was adapted from the microarray-based Associative Transcriptomics study which explored the use of expression markers for predicting hybrid vigour in plants (Stokes et al., 2010). The combination of both SNP variation and gene expression variation, provides a powerful approach for the detection of marker-trait associations, and allows a more in depth understanding of the nature of the marker-trait associations detected. This method was recently utilised in a proof of concept study carried out by Harper et al (2012) in *B. napus.* This study involved use of a panel of 53 *B. napus* accessions on which transcriptome sequencing was performed. Using the methods described by Trick et al (2009) for Tapidor and Ningyou7, 101,644 SNPs were identified between these accessions. In addition, quantification of the sequence read depth, allowed for the identification of gene expression markers (GEMs). Together, these two marker types were successfully utilised in Associative Transcriptomics to identify genes controlling seed erucic acid content. This method was also found to be robust when dissecting the genetic control of a more environmentally sensitive trait, seed glucosinolate content (Harper et al., 2012). This work very elegantly illustrates the power of including both SNP and GEM markers in a single analysis. In the work carried out for glucosinolate content for example, genomic regions were identified where numerous closely linked GEMs were seen to be differentially expressed between the *B. napus* accessions, forming a peak of GEMs associating to the trait of interest. Given the many GEMs making up this association peak, it was hypothesised that the signal observed may be caused by a deletion (covering multiple genes) in a subset of the accessions included. Further experimental work confirmed this hypothesis (Harper et al., 2012). Of course (although not the case for the example presented here), such a signal may also be indicative of a duplication, or other genomic features known to influence the expression of many genes together such as DNA methylation for example.

More recently a pipeline has also been put in place, allowing a similar approach to be taken in hexaploid wheat. The successful employment of this method in wheat can be seen with the recent work carried out by Harper et al (Under review). This study utilised the previously described reference transcriptome (developed from Chinese Spring and Paragon ESTs), against which the transcriptome sequence for 100 wheat accessions were aligned and SNP and GEM markers identified. The use of the identified markers in Associative Transcriptomics allowed for the identification of associating loci for plant biomass traits, including plant height and straw weight (Harper et al., Under review).

1.5.7 Utilising GEMs for the identification of *cis* and *trans*-acting loci

In addition to the use of gene expression markers allowing for the identification of genomic features such as gene duplication or deletions, these data may also be used in identifying *cis* and trans-acting regulation of gene expression in a way analogous that seen in E-QTL analyses (Kabakchiev and Silverberg, 2013). Through Associative Transcriptomics, a number of GEM associations may be detected, suggesting that the differential expression of the associating GEM is somehow related to the trait of interest. To gain more information about such associating GEMs, it is possible to make further use of transcript abundance (measured as RPKM), as a trait in SNP association analyses. In doing this, it may be possible to reveal additional loci related to the expression of the gene to which the GEM corresponds. The SNP association signals obtained in doing this will vary depending on whether the GEM relates to *cis*-acting or trans-acting regulation. In the case of *cis*-acting, we would expect to identify a SNP association in the same genomic position as that to which the GEM corresponds. This suggests that some variation in the sequence of the corresponding gene, is influencing its expression profile. However, in some cases, it may also be possible to identify additional, unlinked loci associating to the transcript abundance values. This may be indicative of *trans*-acting regulation, such as that seen by transcription factors. In this way, the addition of this analysis to the very effective Associative Transcriptomics methods described has the potential to provide more information regarding the biological mechanisms behind the GEM associations detected.

1.6 CONCLUSION

It is clear that despite efforts, there still remains a great need for developing an improved understanding of what contributes to stem mechanical strength in important crop species. A number of studies have been carried out to determine which strength components may be of greatest importance in plant stems. However, it seems that what may be important in determining stem strength in one species, may have little contribution in another. It is therefore important that each species of agronomic importance be assessed individually for this important agronomic trait. Furthermore, many of the studied carried out to date have focused on likely candidate genes assessed through loss-of-function or over expression. Although this may be an effective way of assessing the biological function of these genes, it does not necessarily represent effects that may be seen through the selection of allelic variation for these loci through breeding. Given this, it seems more valuable to assess phenotypic variation at the allelic level through genetic mapping. Although, as discussed, some studies have been successful in detecting genomic regions which co-segregate with stem mechanical and related traits in crop species, these studies, through the utilisation of QTL analysis, may be limited by low allelic variation and low mapping resolution. Given the many genomic advances which have been seen in recent years, it seems that lack of resolution and variation needs no longer hinder scientific efforts aimed at the detection of trait-controlling loci. Association mapping is a powerful approach which aims to reduce the limitations seen in other mapping methods, and has in recent years been implemented successfully in a number of plant species. In addition, novel methods developed by Harper et al (2012), hold great promise for the high resolution mapping of complex traits in species for which genomic resources are limited and in projects where high costs incurred by sequencing may hinder scientific progress.

This thesis will describe the successful application of Associative Transcriptomics for the dissection of the genetic control of stem mechanical strength in both *B. napus* and hexaploid wheat. Stem mechanical strength is a complex trait which is likely to result from the combined effect of several factors. In the present study, the complexity of stem mechanical strength in these important crops is reduced by studying the genetic control of several potentially key component traits separately. Further experimental work will also be described which allowed for the validation of

the detected markers and proposed candidate genes for these traits. In combination, the methods employed will allow for an improved understanding of the genetic control of stem mechanical strength in *B. napus* and wheat.

Chapter 2. General methods

The following methods are those which are of relevance to the work carried out in both wheat and *B. napus*. Where relevant, crop-specific information will be described in Chapters 3 and 4 for *B. napus* and wheat respectively.

2.1 ACCOUNTING FOR VARIATION IN STEM MOISTURE CONTENT

Moisture content is known to have a considerable effect on the mechanical properties of plant stems (Esehaghbeygi et al., 2009). To account for this variation, all stem samples included in the study were first stored for an appropriate length of time within a silica gel tank, set at 55% relative humidity (RH) at 23°C. To determine the length of time required for the moisture content of the stem samples to equilibrate with that of the silica tank, pilot studies were carried out for *B. napus* and wheat separately. Ten stem samples of varying thickness for each crop species were selected for inclusion within these pilot studies. These stem samples were weighed and then place within the silica tank. Each day, the stem samples were re-weighed. This process continued until no further changes in stem weight were observed. The point at which no further weight decrease was seen, was selected as the amount of time required for the stem samples of each species to reach a common moisture content. All stem samples were then stored for the appropriate number of days determined for each species prior to any mechanical testing.

2.2 SCREENING CROP DIVERSITY PANELS FOR MECHANICAL STRENGTH TRAITS

2.2.1 Mechanical testing using a three-point bend test

To screen for stem mechanical strength in *B. napus* and wheat, a three-point bend test method was used. This screen was carried out on the lower portion of the stems of each species and was performed using the Texture Analyser (TA-XT2®- Stable Microsystems, Godalming, UK) facility available at The Institute of Food Research (IFR). The texture analyser was used with a three-point bend test setup. Figure 2.1A provides an illustration of the mechanical testing apparatus used. This method of mechanical testing provides a highly reproducible way to determine the flexure and fracture properties of the sample being tested. Prior to mechanical testing, the texture analyser must be fitted with a load cell. The load cell selected will determine the force capacity which can be achieved during mechanical testing. This will vary

depending on the sample being tested. Following this, the Texture Analyser must be fitted with a deformation probe. For both *B. napus* and wheat, a Texture analyser Break Probe was used. Once fitted, the program Exponent (Stable Microsystems Ltd., Goldalming, UK) (run on a computer linked to the Texture Analyser equipment) was used to set specific test parameters. Firstly, the Exponent program was set up for the desired mode of mechanical testing. For the present study "compression mode" was used. Following the setting of the mechanical testing mode, the speed at which the deformation probe descends towards the test sample was set. In the present study the probe was set (for both *B. napus* and wheat), to descend at a constant speed of 2 mm/sec. To avoid the descending probe damaging the support base (Figure 2.1A), the probe was programmed to return to its original position after a given distance travelled following contact with the sample specimen (This will vary depending of the sample thickness). Following the mechanical testing of each sample, the probe was set at a return speed (the speed at which the probe returns to its initial position) of 15 mm/sec. The stem sample being tested, was placed horizontally (at a previously marked position) across two metal supports. Mechanical testing was then initiated.

During mechanical testing, the deformation probe descends towards the sample. On contact with the sample, an increasing force is applied until mechanical failure. Following mechanical failure, resistance of the stem sample is reduced and as a result, the amount of force applied by the deformation probe decreases. For each sample tested in this way, a real-time graphical output is produced, where the amount of force applied is recorded every 0.005 seconds. On completion of mechanical testing, these data are presented as a plot of stress (the amount of force applied (kg)) against strain (the deformation level observed (mm/sec)). All raw data making up these curves were exported from Exponent into a single Excel spreadsheet, to allow for all further data analyses.

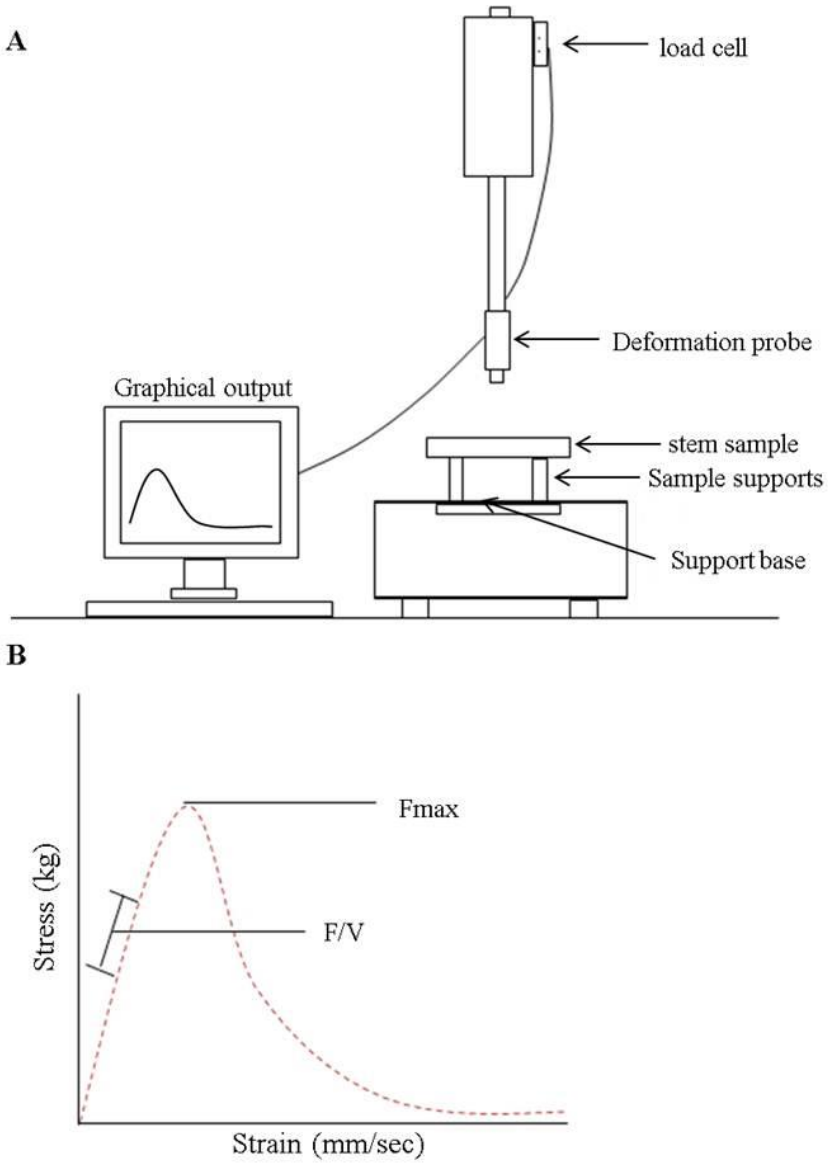

Figure 2. 1. Three-point bend test setup used (A) in determining two absolute strength traits, Fmax and F/V (B) across a panel of wheat and *B. napus* accessions. Fmax, the absolute resistance of the stem sample to break under-load, was determined as the highest point of the stress-strain curve. F/V (force/vertical displacement), which describes the absolute resistance of the stem sample to bend elastically, was determined by assessing the slope of the linear portion of the stress-strain curve prior to mechanical failure (Kern et al., 2005).

2.2.2 Calculating stem mechanical strength from the three-point bend test outputs

Using the stress-strain information obtained from the three-point bend test, two measures of stem strength can be determined. These measures are illustrated in Figure 2.1B where an example of the stress-strain curves generated can be seen. The first of these measures is Fmax, which may be described as the maximum force the stem sample can withstand before mechanical failure. The second measure, F/V (force/vertical displacement) describes the resistance of the stem sample to bend elastically. These measures describe the overall, or absolute, strength of the stem samples, which is thought to be the result of both structural and material strength combined (Kern et al., 2005).

To allow for the compartmentalisation of these absolute strength measures into their structural and material strength constituent parts, several stem structural measurements were taken. This allowed for the later calculation of a commonly used measure of structural strength, the second moment of area (I) (equation 1). In addition, other stem structural traits, which may be of relevance to stem mechanical strength, were obtained. To acquire these data, the transverse cross-section of each stem sample was photographed prior to mechanical testing. Each sample was photographed alongside a ruler, to allow each image acquired to be scaled individually. To obtain stem structural measurements, all photographs were analysed using the program, Sigma Scan (Stystat Software Inc.). Using the trace function, the following measurements were taken:

1- Stem area (used in the later calculation of stem diameter (D2))
2- Stem hollow area (used in the later calculation of stem hollow diameter (D1a))
3- Stem parenchyma + hollow area
4- Outer cortex (stem wall) thickness

These various stem components are illustrated in Figure 2.2. The subtraction of measurement 2 from measurement 3 described above, allowed for a measure of parenchyma tissue area to be obtained. The calculation of stem diameter and stem hollow diameter from the area measurements (1 and 2 above) allowed for any irregularities in the shape of stem cross-section to be accounted for (as opposed to a single measure of diameter which may vary around the stem perimeter).

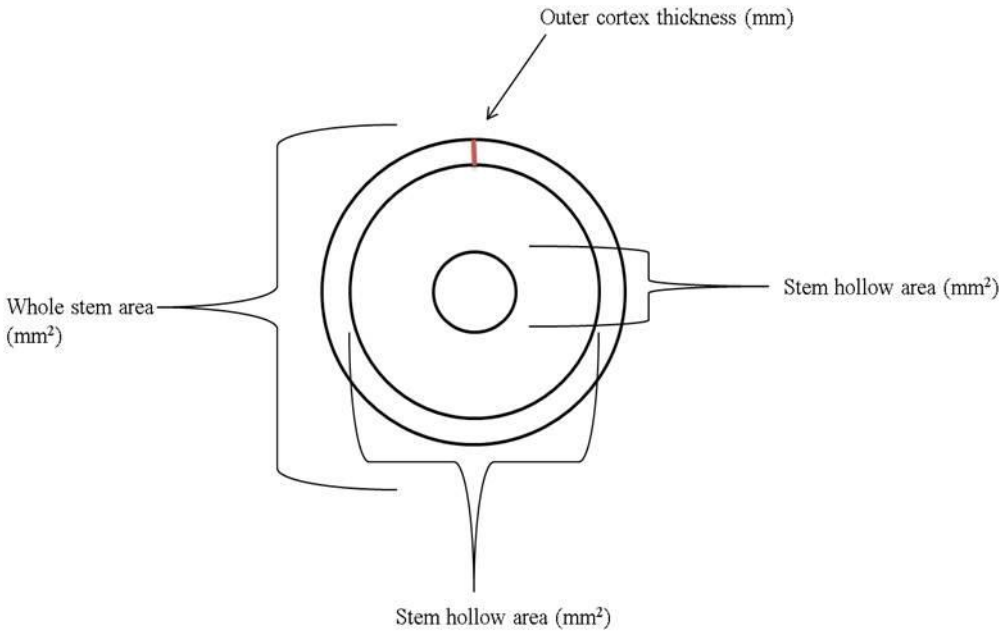

Figure 2. 2. Stem cross-sectional measures recorded in both what and *B. napus*. Hollow area was used in calculating stem hollow diameter (D1a). Whole stem area was used in calculating stem diameter (D2). D1a and D2 were used together in calculating stem second moment of area for each sample.

Equation 1

$$\mathbf{I} = \pi (D2^4 - D1a^4)/64$$

Where:

D2 = diameter of whole stem calculated from stem cross-sectional area

D1a = diameter of stem hollow calculated from stem hollow area

The structural strength trait, I, is a traditional measure of structural strength often applied in engineering practices. Given the often hollow nature of plant stems, a measure of second moment of area which models the geometry of a hollow cylinder was used. In this model, the parenchyma tissue and the outer cortex are modelled together as the cylinder wall. This measure of structural strength was used, together with the absolute strength measures, Fmax and F/V, to obtain measures of stem material strength, the Modulus of Rupture (MOR) and the Modulus of Elasticity (MOE). MOR may be seen as the material equivalent of the absolutes strength trait Fmax, and describes the resistance of the stem material to breakage. MOE, which

may be seen as a material equivalent of F/V, describes the resistance of the stem material to bend elastically. These measures of material strength were calculated according to the following equations:

Equation 2

MOR= (Fmax*a*D2)/I

Equation 3

MOE = (F/V)*(a^2/12)*(3L-4a)/I

Where:

L= the length of the stem sample between the two supports

a=L/2

The above methods used for mechanical testing and trait calculation were adapted from the methods described by Kern et al (2005).

To minimise calculation errors, a script, written by myself and Colin Morgan, was used. Using the stress-strain data obtained from the Texture Analyser, along with the stem structural measures obtained through Sigma Scan, this script was design to perform the calculation of the absolute strength, structural strength and material strength measures according to the equations described above. The script was designed to identify Fmax as the highest force value in the stress-strain data file for each sample. F/V was calculated from the slope of the linear portion of the stress-strain curve prior to reaching the point of Fmax. This script was written in Genstat (Payne et al., 2006) 15th edition (VSN International, Hemel Hempstead, UK) . A copy of the script used can be found in Supplementary Data file 2.1.

2.3 Associative Transcriptomics

2.3.1 Mixed Linear Modelling-SNP analysis

Associative Transcriptomics was carried out according to the methods described by Harper et al (2012). The program used to perform this analysis was TASSEL (Bradbury et al., 2007). Following the loading of SNP data into TASSEL, minor alleles were removed. Those SNPs with minor alleles at <5% were omitted from

further analysis. This filtering process is important in removing SNP variation caused, for example, by sequencing errors.

To reduce the risk of false-positive associations, TASSEL was used to construct a Kinship matrix which describes pairwise relatedness between accessions. This was carried out using the SNP dataset filtered for minor alleles. This information was used as a random effect in the subsequent Associative Transcriptomics analysis. In addition, broad-scale population structure was assessed using the Bayesian clustering method, STRUCTURE (Pritchard et al., 2000). Based on the SNP data, this method was used to identify ancestral groups to which the different accessions could be apportioned. The SNP data was processed using an admixture model with independent allele frequencies. To allow for likelihood estimates of a range of ancestral populations to be made, the model was set to run with hypothetical population (K) estimates of 1 to 5. The SNP data was processed for each value of K three times with a burn-in length of 100,000. This was followed by 100,000 iterations of the Monte Carlo Markov Chain algorithm. To allow for a more accurate estimate of K, the results obtained from STRUCTURE were further analysed using the methods described by Evanno et al (2005). This method assesses the level of variance observed between the replicated runs for each proposed K value. Once the optimal K value had been determined using this approach, a mean of the output from STRUCTURE for the determined K value was used to construct a Q-matrix. This matrix provides information describing the proportion of SNP information within each accession that can be apportioned to each of the determined populations. K-1 Q-matrix scores for each accession were used as a fixed effect in the subsequent Associative Transcriptomics analysis. STRUCTURE analysis for wheat and *B. napus* was carried out by Dr A. Harper.

Together the K-1 Q-matrix and the Kinship matrix were fitted, along with SNP and trait data, in a Mixed Linear Model in TASSEL. This analysis provides an output of P values which describe the strength of the relationship between variation at each marker locus and trait variation. To allow for the identification of clear SNP association peaks, the $-\log_{10}$ of these P values were plotted in Pseudomolecule order and visualised using scripts developed by Dr A. Harper in R (http://www.bioconductor.org/).

2.3.2 Linear regression-GEM analysis

In addition to the SNP analysis, Associative Transcriptomics utilises sequence read depth as a measure of gene expression in an analysis which has been termed Gene Expression Marker (GEM) analysis. Following the quantification and normalisation of the transcript levels as reads per kb per million aligned reads (RPKM), and the filtering of transcripts with an RPKM of less than 0.4 across accessions (our definition of non-zero transcription), linear regression was performed. By fitting the RPKM values for each unigene as the dependant variable and the trait data as the independent variable, it was possible to assess the relationship between this measure of gene expression and the stem mechanical strength and related traits. This analysis was carried out in R, using scripts developed by Dr A. Harper. To allow for the detection of highly associating GEMs, the P values obtained through the linear regression analyses were again plotted (as $-Log_{10}P$) in pseudomolecule order in R.

2.3.3 Utilising the GEM analysis results for the identification of interactions between loci.

The transcript abundance values of associating GEMs were further analysed in a SNP association analysis. This was carried out as described for the SNP analysis for the stem mechanical strength and related traits, where transcript abundance values across accessions is treated as a trait. This analysis can reveal additional loci, which may either directly or indirectly relate to the variation in transcript abundance of the corresponding GEM. This analysis was carried out as previously described, using TASSEL.

NB. General descriptions of candidate genes proposed based on the results obtained from the Associative Transcriptomics analyses, were taken from The Arabidopsis Information Resource (TAIR) (Rhee et al., 2003) and The Rice Genome Annotation project (Kawahara et al., 2013) websites.

2.4 GENOTYPING THROUGH PCR AND SEQUENCING

2.4.1 DNA extraction

96 well plates of leaf material were submitted to the JIC DNA extraction facility where the following protocol (adapted from Palotta et al (2003) by group members of the Professor Graham Moore group, JIC (Pallotta et al., 2003)) was used:

1 µl of RNaseA was firstly added to each tube, within the 96 well racks (each tube containing a single leaf sample). 333 µl warm extraction buffer (pre-heated to 65 ^{0}C) was then added to each tube. The tubes were then sealed with Qiagen caps and the racks held tightly between adaptor plates prior to shaking for 2 minutes at 27 Hz. The racks were then unclamped, turned around and re-clamped and shaken for a further 2 minutes at 27 Hz. Following this, all racks were pulse spun up to 3000 rpm. All racks were then incubated at 65 ^{0}C for 30-60 minutes and were then allowed to cool down to room temperature (approximately 15 minutes). To each tube, 167 µl of 6 M ammonium acetate was added (previously cooled at 4 ^{0}C. All tubes were then again sealed with Qiagen caps and were shaken vigorously for 15 seconds. All samples were then left in a freezer for 10-15 minutes. Samples were then centrifuged for 15 minutes at 5000 rpm to precipitate proteins/plant tissue. Following centrifugation, 400 µl supernatant was added to 240 µl isopropanol (pre-cooled in fridge) into fresh tubes. All tubes were then capped and the samples shaken vigorously for a further 15 seconds. Following a further pulse spin, the caps were replaced with a plastic, adhesive seal and the samples placed into a freezer for 10-15 minutes to allow the DNA to precipitate. Samples were then centrifuged for 15 minutes at 5000 rpm to pellet the DNA and the supernatant discarded. The remaining DNA pellets were then washed in 70% ethanol and centrifuged for 15 minutes at 5000 rpm. Again, the supernatant was discarded. The remaining pellets were allowed to dry over night at room temperature. Following drying, the pellets were re-suspended in 200 µl TE buffer and the tubes sealed with a PCR film and briefly vortexed. The DNA was left to dissolve at room temperature for an hour prior to a final vortex and spin at 5000 rpm for 20 minutes.

2.4.2 Gradient and standard PCR

All primers used in the present study were initially tested through gradient PCR to determination of the optimal conditions for each primer set. All gradient PCRs were conducted using high quality DNA taken from DNA stocks available for the relevant species which have proven robust in previous experiments. The gradient PCRs were carried out according to the following protocol:

For a single 20 µl reaction:

-10x PCR buffer

-2 µl forward primer (10 mM)

-2 µl reverse primer (10 mM)

-1.3 µl DNTPs (2 mM)

-0.2 µl AMPLITAQ Gold (250 u – Life Technologies Ltd (Invitrogen Division, Paisley, UK))

-11.5 µl distilled water

-1 µl DNA

Gradient PCR was carried out using the following cycles on a G-Storm GS1 thermal cycler (Somerton, UK). The elongation time was adjusted to best suit the expected product size (as determined through sequence alignment), allowing 30seconds for every 500 bp of sequence. The gradient temperature range was adjusted depending on the specific melting temperature (Tm °) values calculated by Sigma Genosys for each primer. In most cases it was effective to select the lowest temperature at 3 °C below the lowest calculated Tm° for each primer pair (plus 12 °C for remaining gradient steps). For example, for a primer pair where the lowest Tm ° was calculated to be 61°C with an expected PCR product size of 1 kb, the following gradient PCR cycles would be appropriate:

-Start 40 cycles

-Denaturation at 94 °C for 30 sec

-Gradient at 58 °C-70 °C for 30 sec

-Elongation at 72 °C for 60 sec

-End cycles

-72 °C for 7 min

-Store at 10 °C for infinite

Following the optimisation of PCR conditions, standard PCR was carried out on the relevant genotypes (using the optimal annealing temperature and elongation time as assessed through gradient PCR). The cycles used were the same as those used in gradient PCR, with the gradient step replaced with a single temperature step at the determined annealing temperature

Products were run on 1% agarose to confirm amplification before sequencing

2.4.3 Preparation of samples for capillary sequencing

Prior to preparing sequencing reactions, all PCR products were cleaned-up using the ExoSAP protocol (Etchevers, 2007). This procedure allows for the removal of any remaining primers (through the action of Exonuclease I) and DNTPs (through the action of the Shrimp Alkaline Phosphotase enzyme) from the PCR product. This PCR clean-up step was carried out according to the following protocol:

For a single reaction:

-10 µl PCR product
-1µl Alkaline phosphatase (Roche Diagnostics Ltd, Burgess Hill, UK)
-0.5µl Exonuclease 1(New England Biolabs Ltd, Hitchin, UK)

The SAP-clean was carried out using a G-Storm GS1 thermal cycler with the following cycles:

-Heated lid 110 °C
-Start cycles 1x
-Temperature step 1: 37.0 ^{0}C for 30 min
-Temperature step 2:80 °C for 10 min
-End cycle
-Store at 10 °C infinitely

Following the ExoSAP-clean step sequencing reactions were set up in 0.2 ml tubes according to a revised protocol from BigDye V3.1 terminator cycle sequencing kit (Applied.Biosystems, 2002).

For a single reaction:

-1.5µl 5x sequencing buffer
-1µl primer of interest
-1µl BigDye V3.1
-4.5µl distilled water
-2µl SAP cleaned PCR product

The sequencing reaction was carried out using a G-Storm GS1 thermal cycler machine with the following cycles:

-Heated lid 110.0°C
-Automatic hotstart: 96 °C for 1 min
-Start cycles 25x
-Denaturation: 96.0 °C for 10 sec
-Annealing: 50 °C for 5 sec
-Elongation: 60 °C for 4 min
-End cycles
-Temperature step: 4.0 °C for 10 min
-Store at 4 °C infinitely

Following the sequencing reaction, all samples were sent to GATC (http://www.gatc-biotech.com/en/index.html), Germany, for single read capillary sequencing.

2.4.4 Assessment of sequencing results

All sequencing trace files obtained from GATC (ABI files) were analysed using Contig Express (Vector NTI advance® 11.5.2, Paisley, UK). This allowed for the quality of the sequence reads to be assessed and for the screening of the sequences for the targeted variation.

Chapter 3. The genetic control of stem mechanical strength in *Brassica napus*

CHAPTER OVERVIEW

This Chapter describes the work carried out to explore the genetic control of stem mechanical strength in the polyploid crop species, *Brassica napus (B. napus)* through the utilisation of a novel genetic mapping approach, Associative Transcriptomics. Section 3.1 will describe the variation in stem mechanical strength and related traits across a panel of 79 *B. napus* accessions. In addition, this section will explore the relationships between stem mechanical strength and stem morphological and structural traits. Understanding the level of genetic variation available and exploring trait relationships will provide important information for the consequent genetic analyses. Section 3.2 will describe the Associative Transcriptomics analysis carried out to explore the genetic control of stem strength and that of additional traits found to explain a significant proportion of variation in stem mechanical strength in *B. napus*. Using information regarding gene function from the closely related species Arabidopsis, potential candidate genes for these traits will be discussed. Section 3.3 will describe a marker validation study carried out to assess the true value of the markers detected through Associative Transcriptomics for breeding purposes. Section 3.4 will focus on an Arabidopsis T-DNA insertion line screen, carried out to assess the role of the proposed candidate genes in contributing to stem mechanical strength. Each of these sections will be followed with a discussion specific to each section. Finally, Section 3.5 will provide an overall discussion of the progress made in understanding the genetic control of stem mechanical properties in *B. napus*.

3.1 EXPLORING VARIATION IN STEM MECHANICAL STRENGTH AND THE RELATIONSHIP BETWEEN STEM TRAITS ACROSS A PANEL OF 79 *B. NAPUS* ACCESSIONS

The key objective of this study was to gain an improved understanding of the genetic control of stem mechanical strength in *B. napus*, through the utilisation of genetic mapping to identify marker-trait associations. An important prerequisite for such mapping studies is that sufficient genotypic variation exists for the traits of interest

in the mapping panel used. The following section describes the statistical analyses carried out to determine the level of variation in stem mechanical strength and related traits across a panel of 79 *B. napus* accessions. This section will also identify which variation is of greatest importance when looking to breed for improved stem mechanical strength in this species. This will elucidate which traits should be explored further through genetic mapping. Unless otherwise stated, results presented are those obtained from the material grown as part of the JIC field trial in 2012. Where relevant, additional data will be discussed regarding the results obtained from the additional field trials conducted.

3.1.1 Methods

3.1.1.1 Plant material and harvest

Plant material was harvested across three years. The first two years (grown in 2010 and 2011) consisted of small panels of *B. napus* accessions grown at KWS, Thriplow (with field location varying between years). At both sites, accessions were grown in single (non-replicated) one meter-square, open-field plots (with each accession grown in a single plot). 46 *B. napus* accessions were grown in the 2010 and 44 accessions in 2011. In each year, where possible, ten plants were harvested for each accession. In 2012, a larger trial consisting of 79 *B. napus* accessions was conducted at JIC. This trial consisted of widely-spaced plants grown in three blocks/rows. For each accession one plant was grown in each block. These plants were grown within a net-enclosed cage. Table 3.1 provides an overview of the accessions included in each trial. The majority of accessions grown within these trials were winter oilseed rape accessions. However, a number of spring oilseed rape, swede, Kale and Chinese semi-winter types were also included. For all trials, harvesting was carried out by hand using heavy-duty secateurs. Plants were harvested at the base of the stem. All plants harvested were labelled with a unique identifier and including information regarding the accession name; plant replicate number; harvest site and year of harvest. Following harvest, in some cases, damage to the stem was observed. All stems for which such damage was noticed were omitted from all further analyses. All harvested plants were dried at room temperature prior to any further processing.

Table 3. 1. *B. napus* accessions grown at three field sites between 2010 and 2012. The inclusion of an accession within a given trial is indicated with an "X".

Sequence name	Accession name	Crop Type	KWS 2010	KWS 2011	JIC 2012
-	Aberdeenshire Prize	Swede		x	x
IB02_AbuN	Abukuma Natane	Winter OSR	x	x	x
D3_ALT	Altasweet	Swede			x
D4_AMBxCOM	Amber x Commanche	Winter OSR			x
D6_APE	Apex	Winter OSR			x
D7_APExGIN	Apex-93_5 X Ginyou_3 line	Winter OSR	x		
D10_BAL	Baltia	Winter OSR			x
A11_Bie	Bienvenu	Winter OSR		x	x
D14_Bol	Bolko	Winter OSR			x
IB15_BraS	Brauner Schnittkohl	Siberian Kale	x	x	x
IB16_Bra	Bravour	Winter OSR	x		x
A18_Bro	Bronowski	Spring OSR			x
IB21_Cab	Cabernet	Winter OSR			x
IB22_Cab	Cabriolet	Winter OSR			x
IB23_Can	Canard	Winter Fodder Rape	x	x	x
IB24_CanxCou	Canberra x Courage	Winter OSR	x	x	x
A25_Cap	Capitol	Winter OSR			x
IB28_Cas	Castille	Winter OSR			x
D31_CES	Ceska	Spring OSR			x
A34_Chu	Chuanyou 2	Chinese OSR (semi-winter)			x
IB37_ColxNic	Columbus X Nickel	Winter OSR	x	x	
D39_COR	Coriander	Winter OSR			x
R_IB40_CouN	Couve nabica	Winter Fodder Rape		x	x
A42_Dar	Darmor	Winter OSR		x	x
-	Devon Champion		x	x	
IB45_Dim	Dimension	Winter OSR			x
D46_DIP	Dippes	Winter OSR			x
R_A48_Dra	Drakkar	Spring OSR		x	x
IB49_Dup	Duplo	Spring OSR	x	x	x
IB50_DwaE	Dwarf Essex	Winter Fodder Rape	x	x	x
IB52_EngG	English Giant	Winter Fodder Rape		x	x
IB53_Erg	Erglu	Spring OSR		x	x
IB57_Exc	Excalibur	Winter OSR			x
IB58_Exp	Expert	Winter OSR			x

Table 3.1b.

Sequence name	Accession name	Crop Type	KWS 2012	KWS 2011	JIC 2012
IB60_Fla	Flash	Winter OSR			x
-	Fortin Family	Swede	x	x	
-	Grizzly	-	x		
IB65_GroGS	Groene Groninger Snijmoes	Winter Fodder Rape	x	x	x
-	Guelzower	Spring OSR		x	
D69_HAN	Hanna	Spring OSR			x
IB70_HanxGas	Hansen x Gaspard	Winter OSR	x	x	x
IB73_Hug	Huguenot	Swede	x	x	x
D74_HURxNAV	Huron X Navajo	Winter OSR			x
R_IB78_JauC	Jaune a Collet Vert	Swede		x	x
D79_JETN	Jet Neuf	Winter OSR			x
-	Judzae	Swede	x	x	x
IB86_LemM	Lembkes Malchower	Winter OSR			x
A87_Les	Lesira	Winter OSR	x		
IB91_LicxExp	Licrown x Express	Winter OSR	x	x	x
IB99_MadxRec	Madrigal x Recital	Winter OSR	x	x	x
IB100_Maj	Major	Winter OSR	x	x	x
-	Mestnij	Winter OSR	x		
-	Michinoko	Winter OSR		x	
IB106_MoaR	Moana, Moana Rape	Winter Fodder Rape			x
D108_MON	Monty-028	Spring OSR			x
IB110_N01D_1330	N01D-1330	Spring OSR	x	x	x
IB111_N02D_1952	N02D-1952	Spring OSR	x	x	x
NIN	Ningyou 7	Chinese OSR (semi-winter)	x	x	x
D117_NOR	Norin	Winter OSR			x
IB120_Palm	Palmedor	Winter OSR			x
IB125_POH285B	POH 285, Bolko	Winter OSR	x	x	
D127_PRI	Primor	Winter OSR			x
D128_Q10	Q100	Synthetic (Winter OSR)			x
IB130_Qui	Quinta	Winter OSR	x	x	x
D131_Raf	Rafal	Winter OSR	x		x
D132_RAGJ	Ragged Jack	Siberian Kale			x
IB133_Ram	Ramses	Winter OSR	x	x	x
D134_RAPCR	Rapid Cycling Rape	Spring OSR			x

Table 3.1c

Sequence name	Accession name	Crop Type	KWS 2012	KWS 2011	JIC 2012
A135_Reg	Regent	Spring OSR			x
A136_Roc	Rocket	Winter OSR			x
D137_ROCxLIZ	Rocket (pst) x Lizard	Winter OSR			x
IB139_Sam	Samourai	Winter OSR			x
-	Sarepta	Winter OSR	x	x	
IB141_SenNZ	Sensation NZ	Swede	x	x	x
-	Shang-you	Winter OSR	x	x	
A144_She	Shengliyoucai	Chinese OSR (semi-winter)			x
-	Shen-Li Jutsaj	Winter OSR	x		
D146_SIBB	Siberische Boerenkool	Siberian Kale			x
IB150_SlaS	Slapska Slapy	Winter OSR	x	x	x
IB151_SloK	Slovenska Krajova	Winter OSR	x	x	x
A155_Ste	Stellar	Spring OSR	x		x
IB164_Tai	Taisetsu	Japanese Kale	x	x	x
TAP	Tapidor	Winter OSR	x	x	x
A168_TEM	Temple	Winter OSR	x		x
IB169_TeqxA	Tequila x Aragon	Swede	x	x	x
D170_TIN	Tina	Swede			x
-	TN145		x		
-	TN145		x		
IB175_Tri	Tribune	Spring OSR	x		
R_IB177_Vic	Victor	Winter OSR	x	x	
IB178_Vig	Vige	Swede			x
D181_WESD	Westar 10	Spring OSR			x
D182_WILR	Wilhelmsburger; Reform	Swede			x
A187_Xia	Xiangyou 15	Chinese OSR (semi-winter)	x	x	x
IB184_Yor	York	Swede			x
-	Yudal	Spring OSR	x	x	x
A188_Zho	Zhongshuang II	Chinese OSR (semi-winter)			x

3.1.1.2. Screening *B. napus* accessions for stem lodging susceptibility in the field

Lodging is a complex trait which is thought to be the result of a complex suite of factors including environmental inputs and plant structural and morphological traits (Keller et al., 1999). When looking at a specific aspect of lodging, in this case lodging due to stem breakage, it is therefore very difficult to screen due to the noise resulting from other inputs. Furthermore, in each of the field trials conducted, very little stem lodging was observed, meaning that it was not possible to obtain an accurate score of this trait occurring naturally in the field. To allow for a standardised measure of stem lodging risk (SLR) in field conditions, a pulley system connected to a digital force gauge was used. Given the availability of just three plants per accession (which needed to remain mechanically intact for later experiments), and the often high mechanical strength of the plant stems, it was not possible to induce stem breakage. It was however possible to obtain measures of stem bending stiffness in the field. By attaching the pulley cord to the upper portion of the plant stem (just above the 4th branch), this experiment was able to simulate mechanical perturbation similar to that which would be induced by wind or heavy rainfall. This experiment was carried out on all accessions included in the JIC trial. Each plant was pulled 40 cm from its upright position through a reproducible arc (aided by the pulley design and measured using a meter rule built in as part of the testing apparatus) (Figure 3.1). With the pulley chord running via the digital force gauge, it was possible to obtain a measure of the amount of force (N) required to flex the *B. napus* stems. For each accession, three plants were tested, allowing for a mean stem lodging risk score to be determined per genotype. This experiment was carried out just prior to harvest on green material.

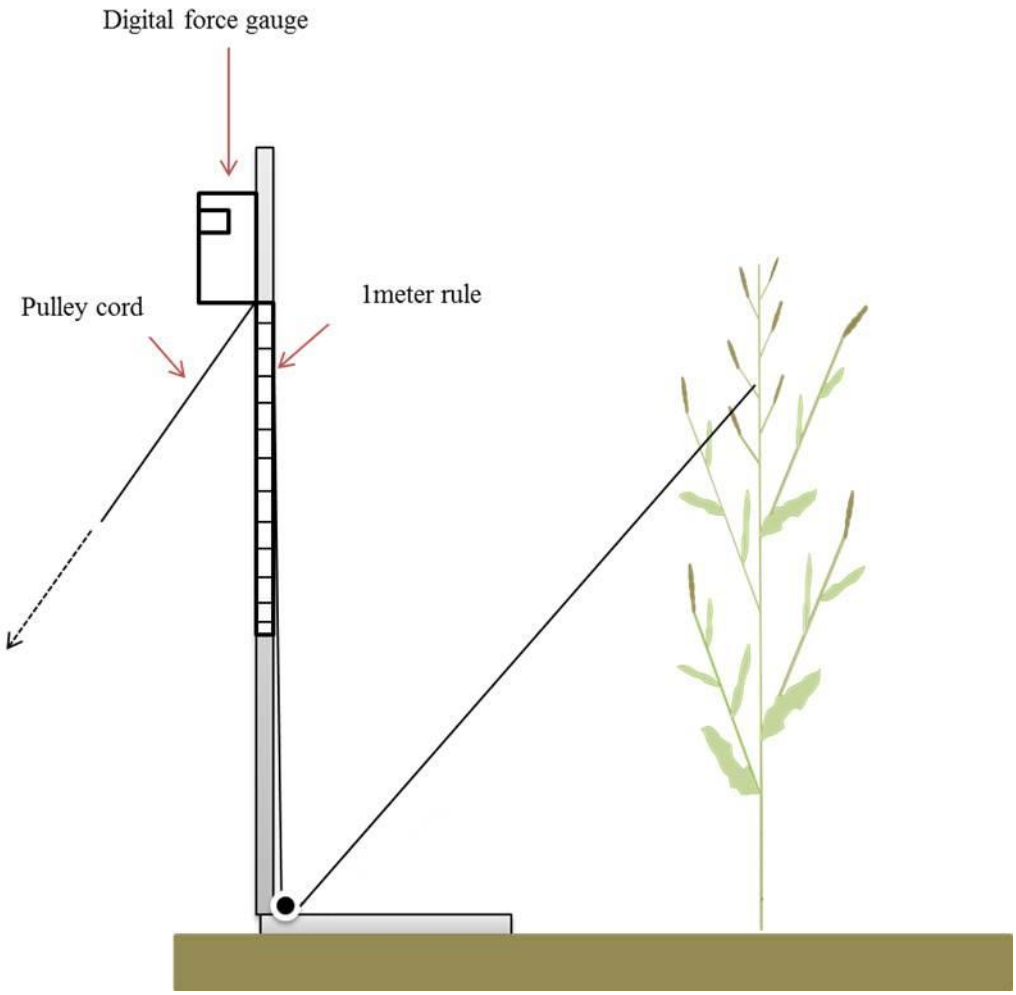

Figure 3. 1. Apparatus used in screening 79 *B. napus* accessions for stem lodging risk under field conditions.

3.1.1.3 Screening *B. napus* accessions for stem morphological and structural traits and preparation of stems for mechanical testing

For all plants harvested, plant height and stem weight measurements were collected. Following this, a 20 cm section was taken from the base of the stem. As described in Section 2.2.2, these sections were photographed to allow collection of stem cross-sectional data. To compensate for variation in moisture content, all samples were further dried in a silica chamber. To assess the length of time required for the moisture content of the stem samples to equilibrate with that of the silica chamber, a pilot study was carried out as described in Section 2.1. This study showed that after six days no further change in stem weight was observed (Figure 3.2). For this reason, prior to mechanical testing, all stem samples were dried for a minimum of six days at 55 %RH and 23 °C in the silica chamber.

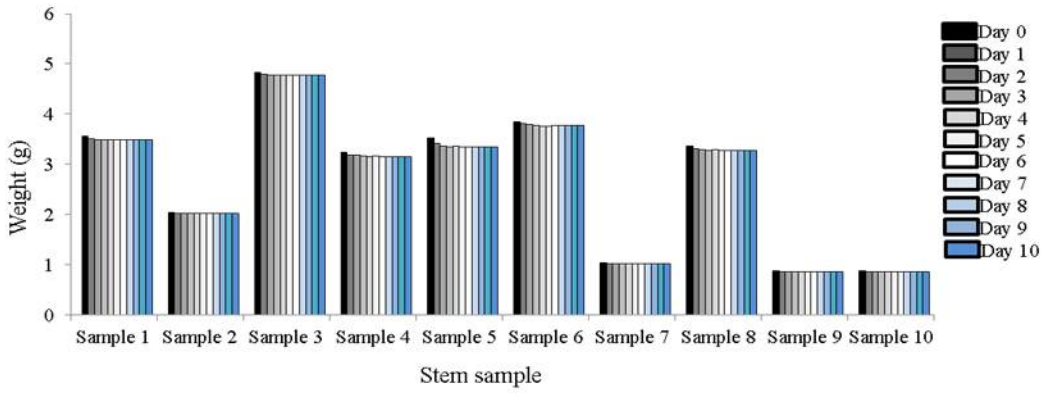

Figure 3. 2. Results obtained from a pilot study carried out on ten *B. napus* stem samples to determine the amount of time required for their moisture content to equilibrate with that of the silica tank (55%RH).

3.1.1.4 Mechanical testing

The three-point bend test was carried out as described in Section 2.2.1. The Texture Analyser was fitted with a 30 kg load cell and the sample support stands were set at 7 cm apart (L). The deformation probe was set at a start height of 7 cm from the support bed and was set to descend 3 cm following initial contact with the stem sample.

3.1.1.5 Data analysis

To assess trait normality and trait variance components, Analysis Of variance (ANOVA) was performed using Genstat 15th edition. Where non-normality was observed, the raw trait values were transformed using a $\log_{10}$ transformation and the ANOVA repeated using transformed data values. To determine whether the accessions varied significantly for the various traits, a minimum significance level of $P<0.05$ was implemented. Following the calculation of trait means, Genstat was used to perform a correlation analysis.

3.1.2 Results

3.1.2.1 Trait normality testing and ANOVA

Using ANOVA, the trait data obtained from the JIC trial was assessed for normality. In all cases approximately normally distributed residuals were observed and therefore no transformations were implemented. An example of this is given in Figure 3.3A for F/V. Through ANOVA, significant differences of $P<0.001$were detected for the majority of traits included in the study. Outer cortex thickness

however, although also varying significantly between accessions, obtained a P value of 0.007. All trait means can be seen plotted in Figures 3.4-3.9.

In addition to the JIC trial, two smaller trials were conducted at KWS, Thriplow. In the majority of cases, accessions were replicated across both trials. However a number of accessions were represented in just a single year (see Table 3.1). To add power to the analysis, it was considered desirable to combine the trait data across years into a single analysis. This can be achieved using Restricted Maximum Likelihood (REML). Following preliminary data analysis, a significant interaction between years was seen for all traits measured across these trials. To account for this, a model was fitted in REML with genotype + year as fixed effects and the interaction term (genotype.year) as a random effect. However, following the fitting of this model, no significant differences were observed between accessions, suggesting that the between year variation is having a large effect. For this reason, these datasets were analysed separately. Although in some cases, the traits measured in these two trials were found to have approximately normally distributed residuals (an example of which is given in Figure 3.3B for stem diameter in the KWS 2010 trial), a number of traits were found to have skewed distributions. In 2010, non-normally distributed residuals were seen for MOE, MOR, second moment of area and stem parenchyma area. In 2011, the same traits were seen to exhibit skewed distributions with the addition of stem outer cortex thickness. In all cases where such non-normality was observed, a $\log_{10}$ transformation was applied which proved successful in achieving greater normality. An example of the skewed residual plots obtained is given in Figure 3.3C for MOR (KWS 2010 data analysis). Figure 3.3D illustrates the increased normality achieved following a $\log_{10}$ transformation. Through ANOVA, with the exception of stem hollow area in the 2010 KWS analysis ($P=0.618$), significant differences were seen between accessions for each of the traits included ($P<0.05$). Data summaries and all ANOVA outputs for each of these field trials can be found in Supplementary Data file 3.1a and b.

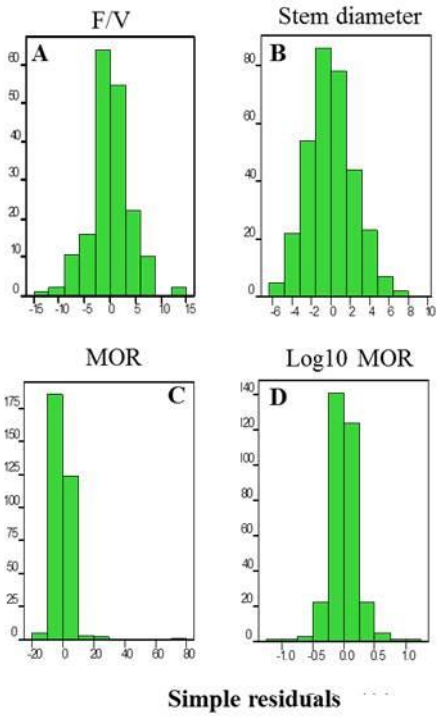

Figure 3. 3. Example trait variance normality plots for *B. napus* grown at JIC and KWS. The residuals for F/V in the JIC-trial analysis can be seen here with a normal distribution (A). Stem diameter also exhibits a normal distribution in the 2010 KWS trial (B). MOR for the same field trial displayed right-skewed residuals (C). Following a log_{10} transformation, greater normality was achieved (D).

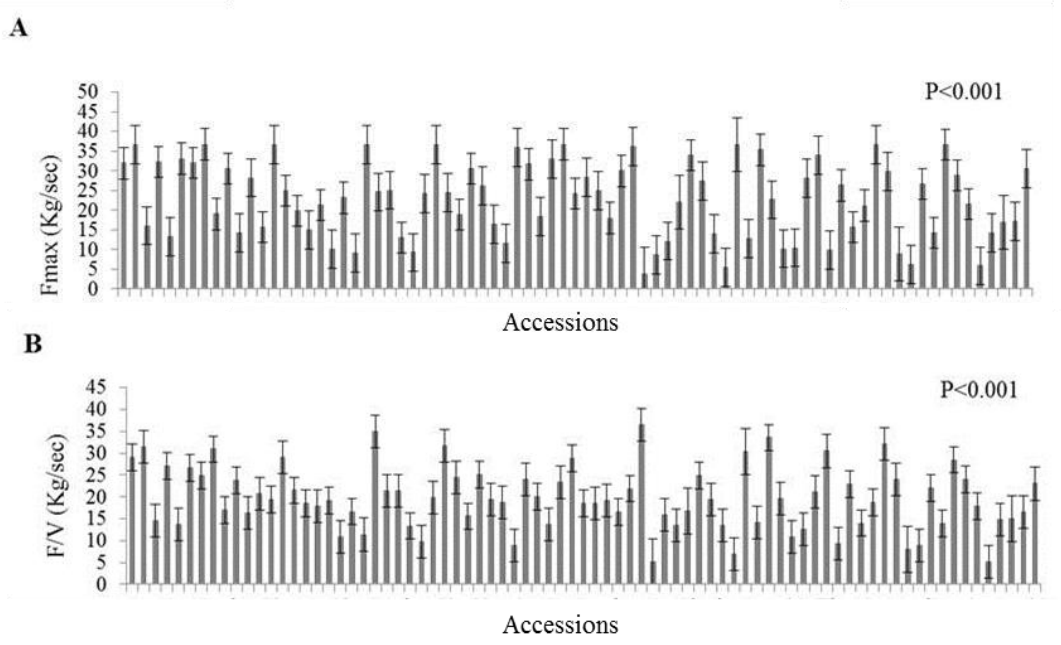

Figure 3. 4. Variation observed for Fmax (A) and F/V (B) across 79 *B. napus* accessions grown in the JIC trial. Error bars represent standard errors calculated through ANOVA.

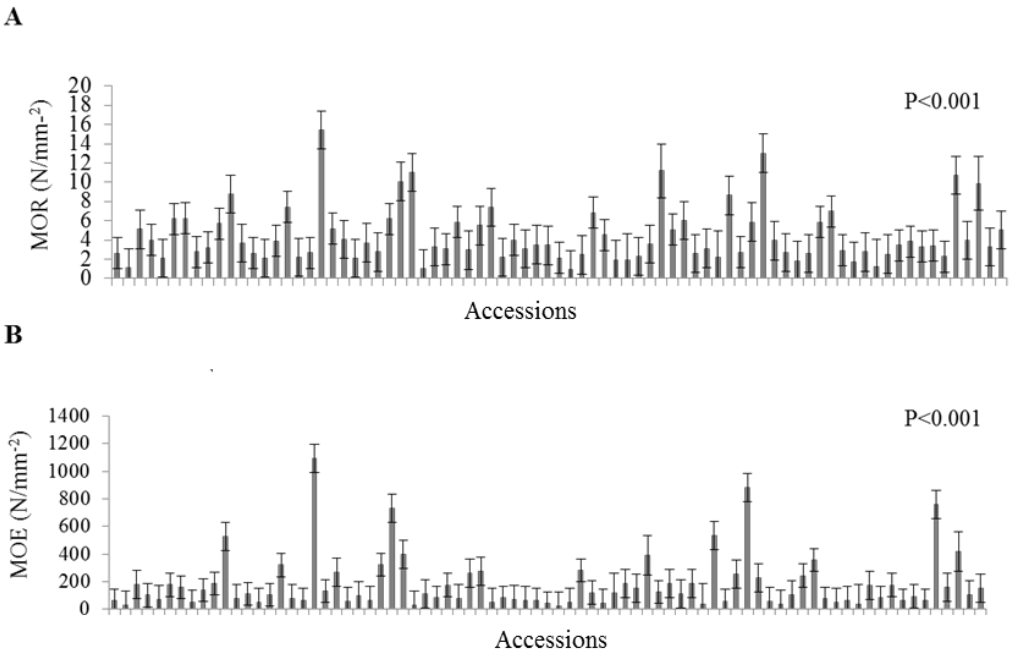

Figure 3. 5. Variation observed for MOR (A) and MOE (B) across 79 *B. napus* accessions grown in the JIC trial. Error bars represent standard errors calculated through ANOVA.

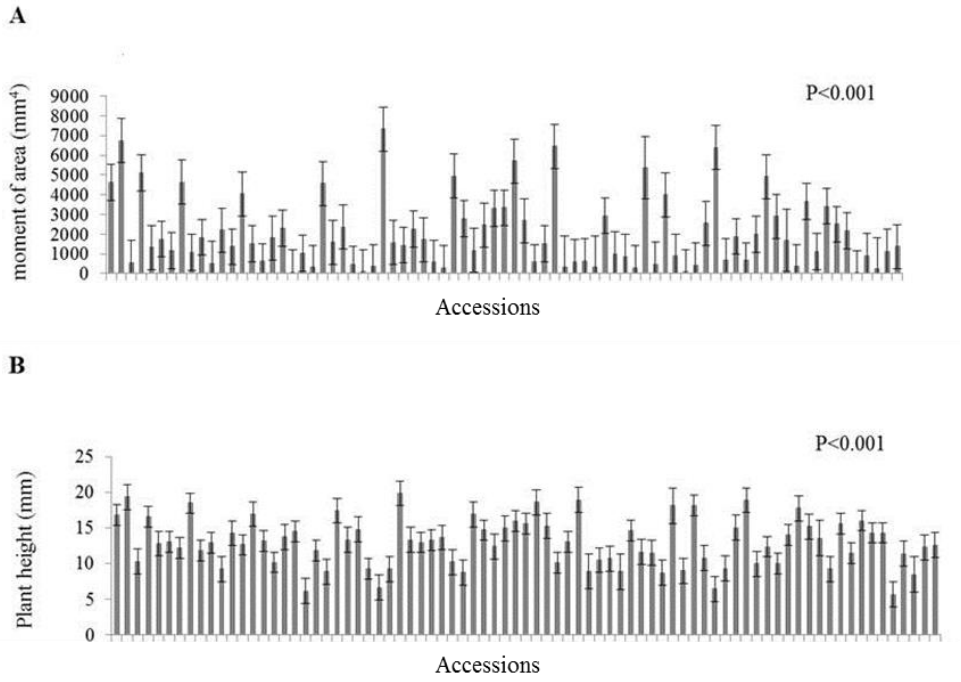

Figure 3. 6. Variation observed for second moment of area (A) and stem diameter (B) across 79 *B. napus* accessions grown in the JIC trial. Error bars represent standard errors calculated through ANOVA.

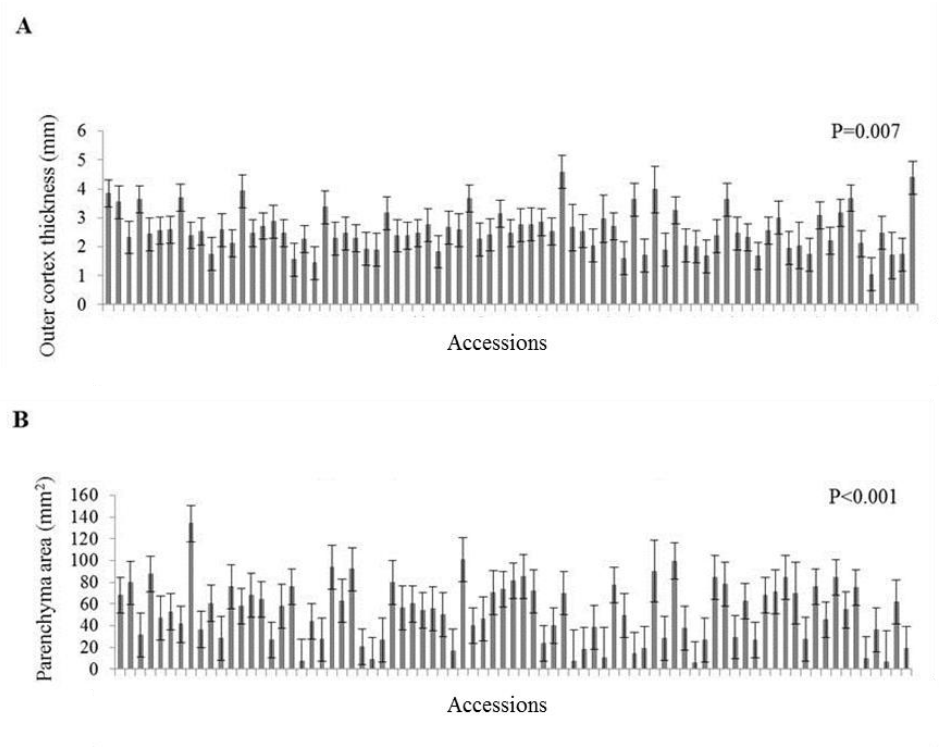

Figure 3. 7. Variation observed for stem outer cortex thickness (A) and parenchyma area (B) across 79 *B. napus* accessions grown in the JIC trial. Error bars represent standard errors calculated through ANOVA.

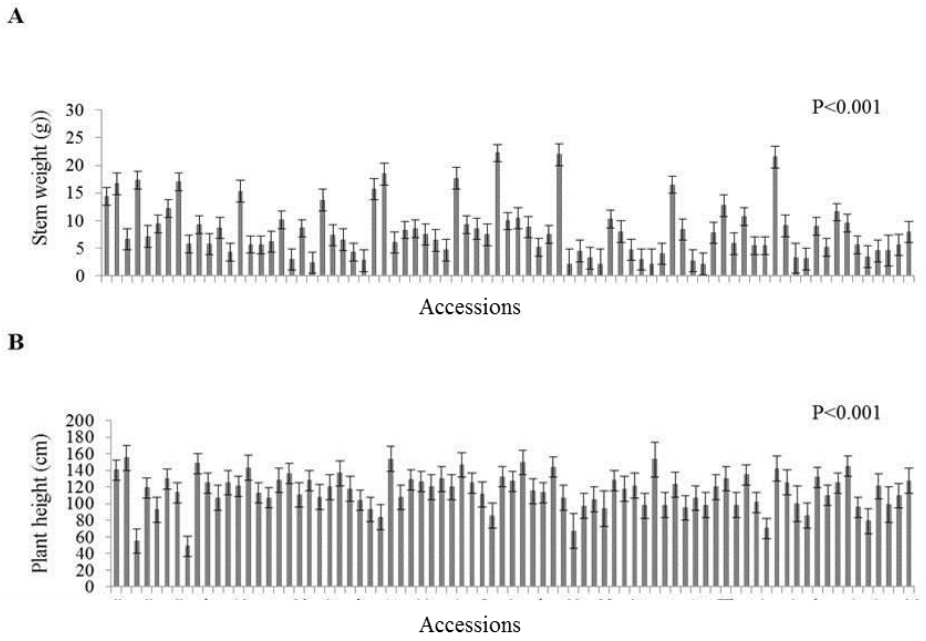

Figure 3. 8. Variation observed for stem weight (A) and plant height (B) across 79 *B. napus* accessions grown in the JIC trial. Error bars represent standard errors calculated through ANOVA.

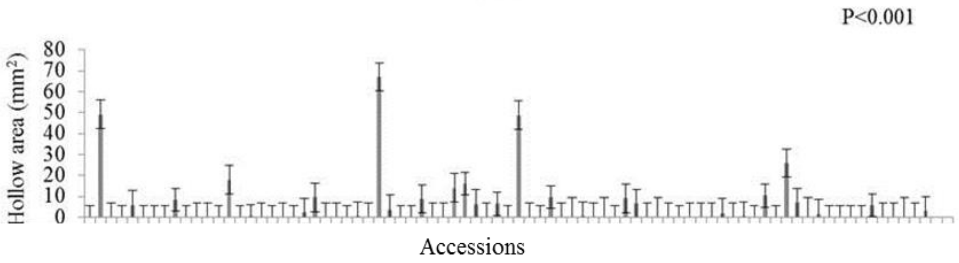

Figure 3. 9. Variation observed for stem hollow area across 79 *B. napus* accessions grown in the JIC trial. Error bars represent standard errors calculated through ANOVA.

3.1.2.2 Relationships between stem mechanical strength and stem morphological and structural traits in *B. napus*

To assess the importance of stem morphological and structural traits on stem mechanical strength, a correlation analysis was conducted using Genstat 15th edition. A summary of the results obtained from this analysis is given in Table 3.2. Many significant correlations were observed. Firstly, the absolute strength traits Fmax and F/V correlate positively with an R^2 of 0.88 ($P<0.001$). This suggests that these traits are highly related i.e. stems resistant to bending are also more resistant to breaking. These absolute strength traits also correlate significantly with a number of stem morphological and structural traits. The structural strength trait, second moment of area, correlates positively with both Fmax and F/V with respective R^2 values of 0.63 ($P<0.001$) and 0.66 ($P<0.001$). Stem diameter was found to explain a greater level of variation than second moment of area in these absolute strength traits with an R^2 value of 0.70 ($P<0.001$) for both Fmax and F/V. Stem outer cortex thickness, and Parenchyma area were also found to correlate positively with Fmax and F/V. These traits are also highly related to stem diameter and second moment of area. It is therefore difficult to assess whether these tissues themselves are contributing to strength, or whether the correlations are due to their relationship with a more general increase in stem thickness.

Weak positive correlations were observed between stem hollow area and Fmax and F/V with respective R^2 values of 0.12 ($P<0.001$) and 0.09 ($P=0.007$). As seen in Figure 3.9, a hollow stem was not a common trait seen across accessions in the plants grown as part of the JIC trial. However, plants with a hollow stem tended to have a large stem diameter. Given this, it is likely that the weak positive correlations observed here are artefacts of the relationship between these stem structural traits, rather than being indicative of higher strength due to larger hollow area. A strong relationship was detected between the absolute strength traits and stem weight where R^2 values of 0.62 ($P<0.001$) and 0.64 ($P<0.001$) were seen for Fmax and F/V respectively. When looking at the relationships between this biomass trait and the stem structural and morphological traits, it can be seen that the correlations with Fmax and F/V are stronger. Similar correlations were also identified between stem weight and stem diameter/second moment of area. It is therefore likely that the relationship observed between stem weight and absolute strength reflects this, but

may also suggest that stem density is important. Stem absolute strength also correlates positively with plant height with respective R^2 values of 0.16 ($P<0.001$) and 0.13 ($P<0.001$) for Fmax and F/V. Plant height was found to show similar correlations with several stem structural traits, including stem diameter with an R^2 of 0.17 ($P<0.001$).

Interestingly, when looking at the relationship between stem absolute and stem material strength traits, a negative relationship can be seen. As described in Section 2.2.2, Fmax and F/V are used in calculating MOR and MOE respectively. Based on the equations used, it would be expected that an increase in these absolute strength traits would result in increased material strength values. However, when looking more closely at the raw data, it can be seen that in the majority of cases, high absolute strength values can be largely explained by high structural strength (large diameter/high second moment of area) and in these accessions low material strength is often observed. While this suggests that stem structural strength is the most important contributor to absolute stem strength in these accessions, it does not mean that the material strength properties do not have a role to play. When looking at the mean trait values, there are several examples which illustrate a certain importance of material strength in determining overall stem strength. An example can be seen when comparing the trait data for the *B. napus* accessions, Dwarf Essex and Taisetu. The absolute strength values for Dwarf Essex are 24.2 kg/sec and 19.9 kg/sec for Fmax and F/V respectively. The mean values for stem diameter and second moment of area for this accession are 9.2 mm and 362 mm^4 respectively. When comparing these data to those obtained for Taisetu, it can be seen that despite displaying very similar values for stem diameter and second moment of area (9.288 mm and 365.8 mm^4 respectively), this accession shows much lower absolute strength values of 6.259 kg/sec and 8.79 kg/sec for Fmax and F/V respectively. This may be explained by the much lower material strength seen for this accession in comparison to that of Dwarf Essex. For Taisetu MOR and MOE value of 2.52 N/mm^{-2} and 173.4 N/mm^{-2} were seen. In comparison, Dwarf Essex was found to have a mean MOR of 11.03 N/mm^{-2} and a mean MOE of 39.1 N/mm^{-2}. This clearly shows the effect that material strength can have in the absence of variation in stem structural strength.

As seen in the correlation analysis results carried out for the JIC trial material, in the analyses carried out on the KWS grown material, the absolute strength traits Fmax

and F/V were found to correlate most significantly with stem diameter; second moment of area; stem weight; parenchyma area and outer cortex thickness (see Supplementary Data file 3.1b. for these correlation analyses)

	Fmax	F/V	MOR	MOE	Second moment of area	Stem diameter	Parenchyma area	Outer cortex thickness	Stem weight	Plant height	Hollow area
Fmax	-										
F/V	***0.88	-									
MOR	**-0.11	***-0.14	-								
MOE	***-0.26	***-0.24	***0.84	-							
Second moment of area	***0.63	***0.66	***-0.32	***-0.3	-						
Stem diameter	***0.7	***0.7	***-0.52	***-0.57	***0.87	-					
Parenchyma area	***0.55	***0.53	***-0.39	***-0.44	***0.63	***0.79	-				
Outer cortex thickness	***0.44	***0.47	***-0.194	***-0.25	***0.48	***0.49	***0.19	-			
Stem weight	***0.62	***0.64	**-0.11	***-0.18	***0.59	***0.55	***0.39	***0.32	-		
Plant height	***0.27	***0.23	**-0.09	**-0.1	***0.25	***0.26	**0.13	***0.14	***0.13	-	
Hollow area	**0.12	**0.09	**-0.09	*-0.05	***0.31	***0.2	*0.06	0.05	***0.17	***0.19	-

Table 3. 2. Results obtained from a correlation analysis for the mechanical strength traits obtained through the three-point bend test and stem morphological and structural traits assessed in *B. napus*. Inverse relationships can be seen marked with "-". * is indicative of a P value of <0.05; ** is indicative of a P value of <0.01 and * is indicative of a P value of <0.001.**

3.1.2.3 The relationship between stem lodging risk and stem mechanical strength across 97 JIC-grown *B. napus* accessions

Using the methods described in Section 3.1.1.2, all *B. napus* accessions grown in the JIC trial were assessed for stem lodging risk (SLR) in field conditions. The variation detected for this trait can be seen plotted graphically in Figure 3.10. Through ANOVA a significant difference was observed between accessions for SLR (P=0.043).

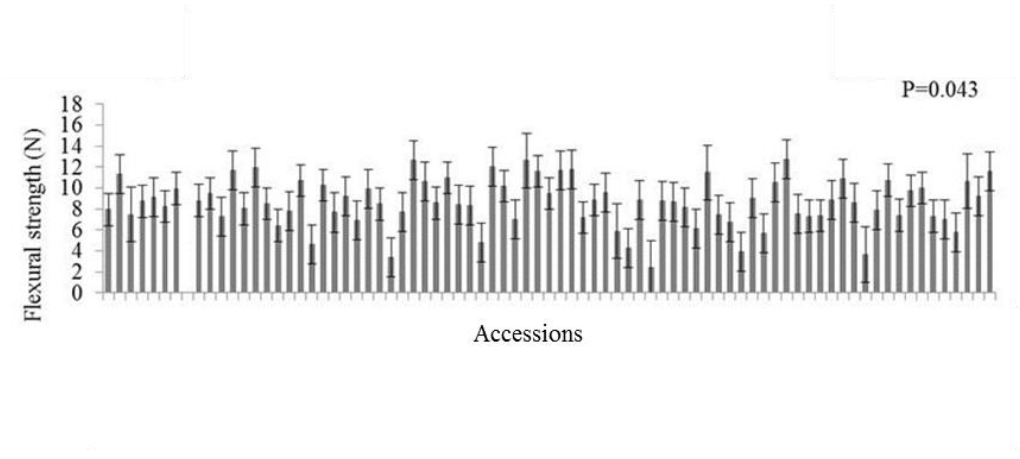

Figure 3.10. Variation in stem lodging risk observed across 79 *B. napus* JIC-grown accessions. Error bars represent standard errors as calculated through ANOVA.

Through a further correlation analysis, the relationship between this field-based scoring and the mechanical strength traits obtained from the three-point bend test were explored. Some of the key relationships observed are plotted in Figures 3.11 and 3.12. Significant positive correlations were observed between the absolute strength traits and the SLR scores measured in the field with R^2 values of 0.42 (P<0.001) and 0.33 (P<0.001) for Fmax and F/V respectively. These lodging risk scores were however found to correlate negatively with the material strength measures with R^2 values of 0.16 (P<0.001) for MOR and 0.21 (P<0.001) for MOE respectively. This again illustrates the importance of stem structural strength in determining stem strength in *B. napus*. Indeed, very similar correlations were observed between the stem structural traits stem diameter and second moment of area and the SLR score as those seen for absolute strength. For both of these structural traits, positive correlations were observed with R^2 values of 0.42 (P<0.001) and 0.37(P=0.001) for stem diameter and second moment of area respectively. Stem

height was also found to correlate positively with the SLR scores with an R^2 value of 0.25 ($P<0.001$). However, given that this correlation is weaker than those seen between SLR and the absolute strength measures it seems that the variation in plant height has not had a confounding effect on SLR. The detection of these significant correlations is very promising. These findings suggest that the mechanical strength measures obtained through the three point bend test are capturing important variation relevant to stem mechanical strength under field conditions (see Supplementary Data file 3.1c. for data summary and ANOVA output for this experiment).

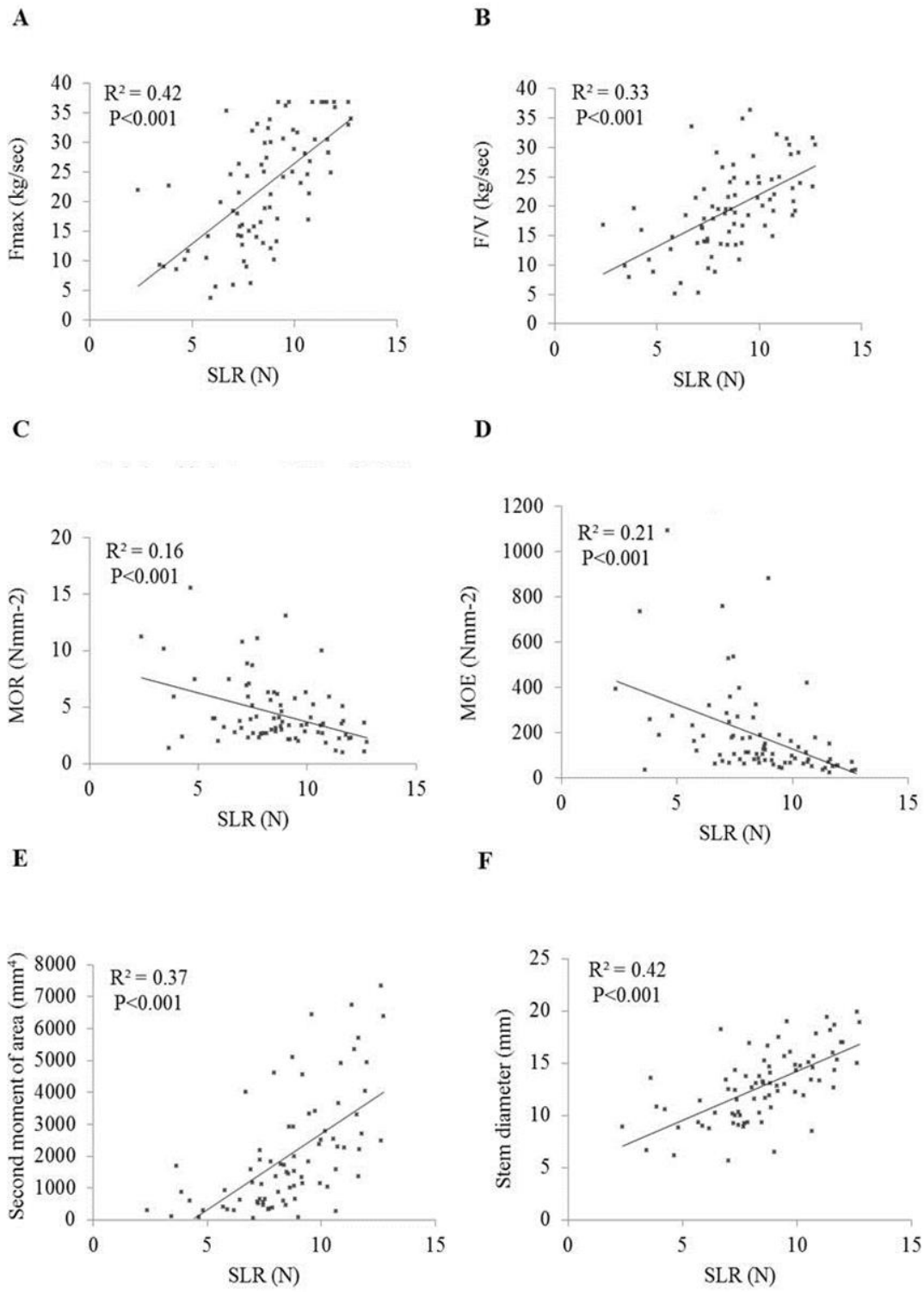

Figure 3. 11. Relationships observed between the field-based stem lodging risk (SLR) score and the stem mechanical strength traits (as determined through the three-point bend test) across 79 *B. napus* accessions (Fmax (A), F/V (B), MOR (C), MOE (D), Second moment of area (E) and stem diameter (F)).

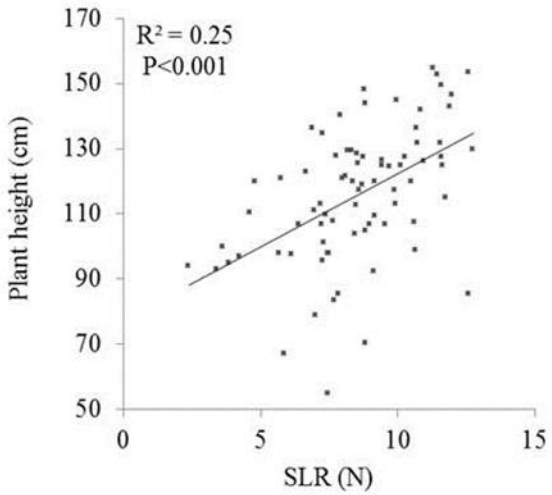

Figure 3. 12. Relationships observed between the field-based stem lodging risk (SLR) score and plant height across 79 *B. napus* accessions.

3.1.3 Discussion

Section 3.1 has focused on exploring the variation in stem mechanical strength across a panel of 79 *B. napus* accessions. The results presented here show that the mechanical strength traits explored are often highly variable within a single genotype. Through calculating the variation coefficient (CV) for Fmax, it was revealed that this trait has a much lower CV in the results obtained from the JIC trial (45.4 %) in comparison to the results obtained from the 2010 (91.6 %) and 2011 (74.4 %) trials. This suggests that despite a much lower level of replication, the JIC trial had greater power to express the genetic potential of this trait. This is likely to be due to the very different plot design between the JIC trial and the two KWS trials. Plants grown in the JIC trial were widely spaced, in comparison to those grown at KWS which were grown in drilled plots. Widely spaced plants are able to grow freely, with reduced competition for space and resources with neighbouring plants. As a result (and particularly given the large contribution of stem diameter to stem strength in *B. napus*), it is much easier to detect variation for these traits in a more controlled environment where a more accurate measure of genetic potential can be realised. However, despite being able to more accurately assess variation in stem strength in widely spaced plants, increased replication would have been desirable if

greater field space had been available in the more comprehensive JIC trial. This would allow for more accurate estimates of the population mean to be obtained.

In the majority of cases, a high level of genetic variation was uncovered. This suggests that there is potential to improve these traits through breeding. The results obtained for the KWS field trials do however suggest that these traits are significantly influenced by the environment with significant interactions between genotype and year being seen for all traits included.

For all field trials, very similar trait correlations were observed. These analyses suggest that the stem structural strength traits, second moment of area and stem diameter, are key contributors to stem absolute strength in this species. In addition, stem outer cortex and stem weight also seem to be important. These structural components may therefore be key targets in breeding *B. napus* accessions with improved stem mechanical strength.

As previously discussed in Section 3.1.2.2, an inverse relationship was seen between stem absolute strength and stem material strength. While it seems clear from the analyses conducted that structural strength is the main contributor to stem strength across this panel of *B. napus* accessions, examples where stem material strength appears to be important (although the contributions are rather modest in comparison to the stem structural traits) were identified. It is possible that there is a compensatory effect, where a lack of strength due to structure can to some extent be improved with high material strength. The finding that in general the absolute and material strength measures are inversely related may have several biological explanations. It may be that the biological process (or processes), that leads to high structural and material strength are in some way antagonistic and that it is therefore not biologically possible to achieve large contributions from both strength components. It may also simply be that having both high material and high structural strength requires too much energy and resources and that stem structural strength is achieved in a more energy efficient way. As a final possible explanation for the trend observed, it may suggest that there has been a greater selection pressure for structural strength and/or against material strength throughout the breeding history, or indeed, evolutionary history of this *B. napus*. It may for example be that high material strength has been selected against for the improvement of combining ability.

The results from the stem lodging risk field experiment are very promising. Under field conditions, forces, applied by wind and rainfall, perturb the entire plant body, not just the base of the stem. The field-based stem lodging risk experiments were designed to more closely represent this. That the results obtained from this field-based experiment correlate positively with the absolute strength measures obtained from the three-point bend test, suggests that important variation relevant to the mechanical perturbations induced in field conditions has been uncovered.

It would be interesting to assess the relationship between stem mechanical strength and stem chemical composition. Although work is currently underway (in the lab of Professor Keith Waldron, IFR) to assess the straw chemical composition of these *B. napus* accessions, these data were unfortunately not available to allow potential relationships to be explored in this present study. Given the potential link between chemical composition, feedstock processibility and stem mechanical strength, it would also be of interest to assess how variation in stem mechanical strength relates the processibility of *B. napus* straw in lignocellulosic ethanol processing. These would be very interesting areas for future research.

3.2 UNCOVERING THE GENETIC CONTROL OF STEM MECHANICAL STRENGTH AND RELATED TRAITS IN *B. NAPUS* THROUGH ASSOCIATIVE TRANSCRIPTOMICS

In Section 3.1, it was found that a high level of variation in mechanical strength and related traits exists between the *B .napus* accessions included in this study. This suggests that there may be good breeding potential for these traits. As described in Section 1.5, the detection of marker variation which segregates with a trait of interest, may be utilised in the marker-assisted breeding of that trait. Very few studies have focused on stem mechanical strength in *B. napus*. Grundas (2008) carried out a small survey to determine the effects of stem mechanical strength across five winter oilseed rape accessions on lodging resistance. This work revealed a close correlation between these traits. (Grundas and Skubisz, 2008). Mahdavian et al (2012) assessed variation in shear strength in three rapeseed accessions. This study was carried out to improve the efficiency of mechanical harvesting in *B. napus* (Mahdavian et al., 2012) . QTL analyses have been conducted to assess the genetic control of lodging resistance and related traits in a number of crop species including

rice (Kashiwagi et al., 2008), wheat (Hai et al., 2005) and maize (Flint-Garcia et al., 2003). This method has also been used to explore the genetic control of lodging resistance in *B. napus* (Gu and Qi, 2009). However, as described in Section 1.5.2, this method suffers from low allelic variation and resolution. More recently, association mapping, which aims reduce these limitations, has been employed to uncover the genetic control of agronomically important traits in plants. (Zhu et al., 2008). The following section describes the Associative Transcriptomics analyses carried out to uncover genetic markers which may be utilised in marker assisted breeding to improve stem mechanical strength in *B. napus*. The majority of the results presented here, as in the previous section, will focus on those obtained from the JIC trial material. However, any common associations detected between the JIC and KWS trials will be highlighted and any strong associations detected in the KWS field trials only will also be described. This Associative Transcriptomics analysis was carried out for both the material and absolute strength traits. In addition, all morphological and structural traits found to correlate closely with stem absolute strength are also included. The compartmentalisation of mechanical strength in this way, will allow breeders to select for multiple traits relevant to stem mechanical strength, allowing for the development of an ideotype for lodging resistance in this species. Furthermore, the identification of potential candidate genes in close linkage with associating markers, will allow for any potential conflict between the improvement of stem mechanical strength and that of lignocellulose processibility to be identified. This will provide important information regarding the improvement of *B. napus* as a potential feedstock crop.

3.2.1 Methods

3.2.1.1 mRNAseq and marker detection in *B. napus*

All accessions included in this study were sequenced using Illumina transcriptome sequencing. To allow for the identification of SNPs between these accessions, the sequencing reads were aligned to a reference sequence comprised of a collection of 94,558 unigenes. These unigenes were assembled using publically available ESTs. Using a high density linkage map (Bancroft et al., 2011) together with *B. rapa* (Wang et al., 2011) and *B. Oleracea* genome scaffolds (Liu et al., 2014), these unigenes were assembled into a hypothetical gene order. The ordering of these

unigenes was also aided by information regarding conserved synteny between Arabidopsis and *B. napus* (Harper et al., 2012).

These assembled sequences, or pseudomolecule (see Supplementary Data file 3.2e for pseudomolecule V4-a revised version of the *B. napus* pseudomolecule published by Harper et al (2012)), were then be used in place of a full reference genome sequence. Alignment of the mRNAseq data to this reference sequence (using MAQ), identified 225,001 SNP markers. In addition, quantification of sequence read depths (as reads per kb per million aligned reads; RPKM) provided a measure of expression for each unigene. RPKM values therefore provide the information required for exploring the relationship between gene expression and the trait of interest in what has been termed a Gene Expression Marker analysis, or GEM analysis. A full description of the methods used for the detection of these markers can be found in Trick et al (2009) and Harper et al (2012).

3.2.1.2 Accounting for population structure and relatedness and performing Associative Transcriptomics in *B. napus*

The Associative Transcriptomics analysis was carried out as described in Section 2.3. To assess the level of broad-scale population ancestry, STRUCTURE analysis was used. This method allows the number of populations (K) between which the genome ancestry of the accessions can be apportioned to be determined. This analysis was carried out as described in Section 2.3.1. Following this analysis and further calculations as described by Evanno et al (2005), the best estimate of K was revealed to be 2. Figure 3.13A shows a summary of the output obtained from STRUCTURE. Figure 3.13B shows the detection of the highest Delta K value where K=2 (see Supplementary file 3.2a. for full calculation of K=2). This analysis was performed by Dr Andrea Harper (Harper et al., 2012). Based on these results, a Q matrix, describing the apportioning of the 79 *B. napus* accessions between the two detected populations, was used in performing the Mixed Linear Model (MLM) analysis in TASSEL as described in Section 2.3.1 (Harper et al., 2012).

The SNP data file available (see Supplementary Data file 3.2b) consisted of 225,001 SNP markers. Following the removal of minor alleles present at <5%, this data set was reduced to 144,131 SNPs. This filtered SNP data set was used in TASSEL to create a Kinship matrix. The GEM analysis, carried out as described in Section 2.3.2, utilised transcript abundance of 189,116 unigene sequences (the transcript abundance data used for the GEM analysis can be found in Supplementary Data file 3.2c) (Harper et al., 2012). All highly associating GEM markers were further analysed by mapping their respective RPKM values as a trait against the SNP data (this method is described in Section 2.3.3).

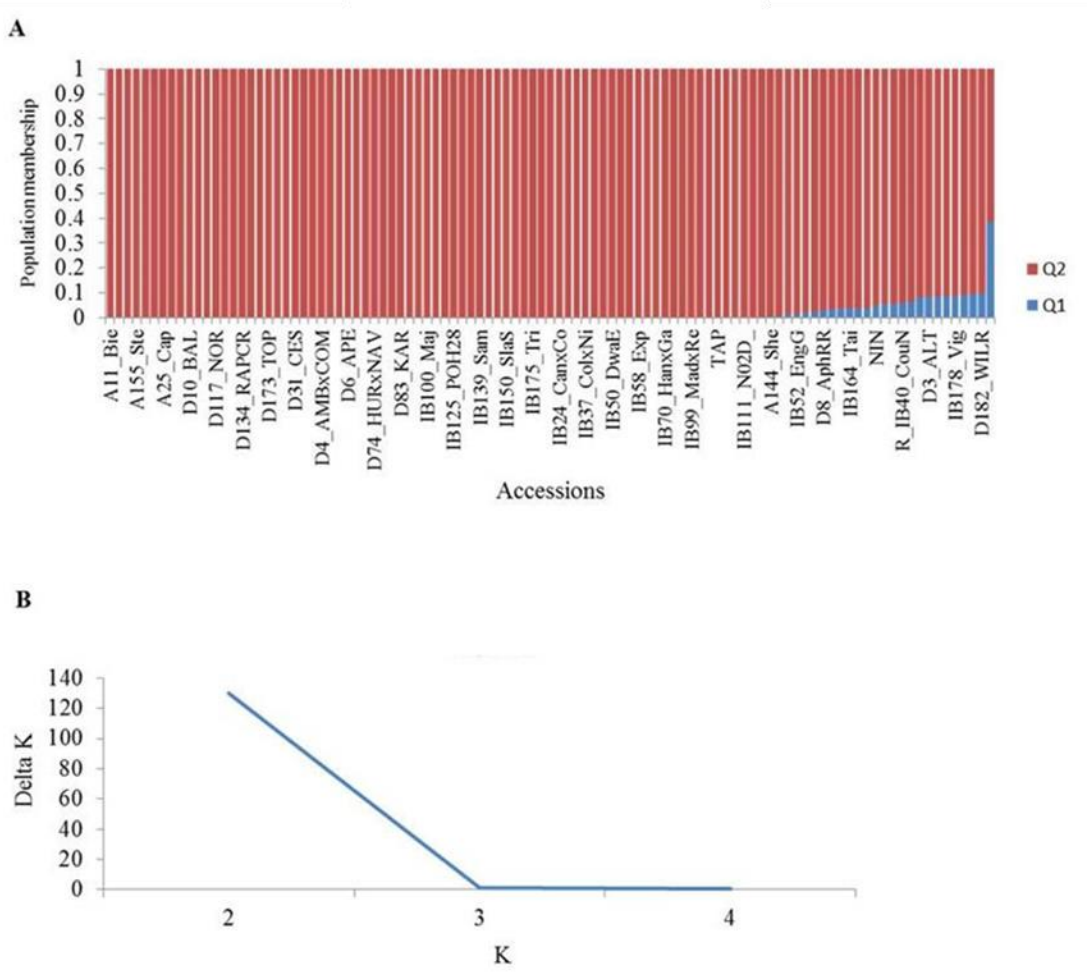

Figure 3. 13. Results of STRUCTURE analysis carried out in *B. napus*. The STRUCTURE output (A) shows two main populations (Q1 and Q2). Using methods described by Evanno et al (2006), a K value of 2 was confirmed to be the most likely estimate (B) (Harper et al., 2012).

3.2.2 Results

This section describes only the most promising marker associations detected for each of the mechanical strength and related traits included in the present study. The scripts used in plotting the results obtained from the SNP and GEM analyses and in performing Linear Regression can be found in Supplementary Data file 3.2d.The most promising SNP associations for each trait are summarised in Table 3.3. Table 3.4 summarises some of the most significant GEM associations. Full SNP and GEM association plots can be found in Supplementary Figure files 3.1-3.24.
Supplementary Data file 3.2f provides further information regarding candidate genes as described in TAIR (The Arabidopsis Information Resource).

Table 3. 3. SNP associations detected for absolute and material strength in the *B. napus* Associative Transcriptomics analyses.

Trait	marker ID	Linkage group	pseudomolecule position A/C homeologs (kb)	P value	increasing allele (No. observations)	decreasing allele (No. observations)	effect %	Candidate gene (Pseudomolecule position A/C (kb))
Fmax	JCVI_31359:1657*	A2/C2	10264.1/22158.4	4.45E-05	C (32)	S (37)	30.9	AT5G50030 (10168.1/21818.4)/ AT5G50060 (10221.6/22075.5)
Fmax	JCVI_39914:1128	A5/C5	20443.6/39615.3	4.07E-04	C (36)	S (30)	26.9	AT3G12880 (20419.9/39582.9)/ AT3G12830 (20435.5/39598.9)
Fmax	JCVI_639:522	A8/C3	10113.9/65259.7	5.90E-04	A (68)	M (8)	33.6	AT5G60690(10033.5/47883.9)
Fmax	JCVI_38663:382*	A3/C3	7381.8/9829.4	4.16E-05	C (43)	Y (35)	25.2	-
F/V	JCVI_39914:1143	A5/C5	20443.6/39615.3	4.41E-04	R (36)	G (31)	21.7	AT3G12880 (20419.9/39582.9)/ AT3G12830 (20435.5/39598.9)
F/V	JCVI_31359:1657*	A2/C2	10264.1/22158.4	5.73E-05	C (32)	S (37)	24.6	AT5G50030 (10168.1/21818.4)/ AT5G50060 (10221.6/22075.5)
F/V	JCVI_4281:518	A6/C5	5768.2/6694.9	1.53E-04	A (26)	R (52)	19.2	-
MOR	JCVI_19955:183*	A1/C1	1247.4/1677.1	7.45E-0 6	A (4)	W (66)	52.6	AT4G36160 (1177.863/1512.4)
MOR	JCVI_10685:430*	A6/C6	1645.9/7765.3	2.05E-05	R (6)	G (47)	44.5	AT1G51590 (1703.2/ 7662.2)/ AT1G51630 (1679.6/?) AT1G52150 (1480.5/8107.7)
MOE	JCVI_19955:183*	A1/C1	1247.4/1677.1	2.56E-08	A (4)	W (66)	61.9	AT4G36160 (1177.863/1512.4)
MOE	JCVI_22602:1182	A2/C2	23581.7/47550.6	8.38E-06	W (6)	T (71)	37.2	-
MOE	JCVI_10685:430*	A6/C6	1645.9/7765.3	2.62E-06	R (6)	G (47)	45.5	AT1G51590 (1703.2/ 7662.2)/ AT1G51630 (1679.6/?) AT1G52150 (1480.5/8107.7)
MOE	JCVI_19824:614*	A7/C6	20430.5/37273.1	5.29E-06	Y (5)	C (70)	47.6	-

3.3b

Trait	Marker ID	Linkage group	Pseudomolecule position A/C (kb)	P value	Increasing allele (No. observations)	decreasing allele (No. observations)	Effect %	Candidate gene (Pseudomolecule position A/C (kb))
Stem diameter	JCVI_31359:1651	A2/C2	10264.1/22158.4	1.01E-04	G (30)	R (33)	25.5	AT5G50030 (10168.1/21818.4)/ AT5G50060 (10221.6/22075.5)
Stem diameter	JCVI_38024:112	A4/C4	14821.7/47143	2.12E-04	Y (26)	T (45)	20.1	AT2G30575 (14754/47055.7)
Stem diameter	JCVI_18806:670	A5/C5	20469.1/39646.7	4.36E-04	M (10)	A (49)	26.2	AT3G12880 (20419.9/39582.9)/ AT3G12830 (20435.5/39598.9)
Parenchyma area	JCVI_36764:445	A1/C1	19485.6/35654.8	2.30E-04	G (49)	R(28)	21.2	AT2G30810 (19547.7/35739.2)
Parenchyma area	JCVI_31359:1651	A2/C2	10264.1/22158.4	1.15E-04	G (30)	R (33)	24.8	AT5G50030 (10168.1/21818.4)/ AT5G50060 (10221.6/22075.5)
Outer cortex thickness	JCVI_335:451	C6	16454.2/1543.5	2.70E-04	T (12)	W (57)	24.2	-
Stem weight	JCVI_33733:304*	A4/C4	20092.7/547.2	3.59E-05	G (14)	R (52)	43	-
Stem weight	JCVI_22620:181*	A9/C9	5004.3/7919.5	5.38E-05	C (5)	Y (69)	45.3	-
Plant height	JCVI_26003:352	A3/C7	27977.9/47879.2	2.86E-04	C (24)	T (39)	21.7	AT4G30610.1 (27923.5/27923.9)
Plant height	JCVI_19156:300	A9/C8	36427.9/33160.6	8.82E-04	G (25)	R (45)	16.8	-

Table 3. 4. GEM associations detected for absolute and material strength in the *B. napus* Associative Transcriptomics analyses. The trend column describes the relationship between the trait and the transcript abundance (measured as RPKM) of the associating GEM.

Trait	GEM ID	linkage group	P value	R^2	Trend	PSEUDO POS	Candidate gene
Fmax	JCVI_19323	A9	7.64E-08	0.32	+	5991.9	AT2G19130
Fmax	JCVI_26403	C7	4.34E-08	0.38	-	29074.8	AT4G17810
Fmax	JCVI_28286	C7	5.41E-08	0.34	-	35061.8	AT4G27710
Fmax	JCVI_7586	A5 (peak)	1.96E-06	0.30	-	24082.4	AT5G16110
F/V	JCVI_7586	A5 (peak)	2.02E-06	0.29	-	24082.4	AT5G16110
F/V	JCVI_28286	C7	1.45E-07	0.32	-	35061.8	AT4G27710
F/V	JCVI_26403	C7	3.19E-06	0.35	-	29074.8	AT4G17810
MOE	JCVI_38947	A1	5.17E-07	0.33	+	18665.1	-
MOE	ES266451	A1	9.00E-07	0.32	+	18667	-
MOE	JCVI_14486	A1	4.96E-07	0.30	+	25574.5	AT3G13690
MOE	JCVI_39585	A2	1.25E-07	0.36	+	6132.3	-
MOE	EX121877	A4	2.48E-08	0.31	+	17023.3	AT2G36810
MOE	AM062044	A6 (peak)	7.28E-07	0.3	+	22023.2	AT4G27710
MOE	CD841065	A8	2.54E-07	0.31	+	825.8	AT3G58670
MOE	JCVI_15213	A8	1.32E-06	0.28	+	826.9	AT3G58670
MOE	JCVI_30035	C1 (peak)	1.55E-06	0.27	+	50080.6	
MOE	JCVI_24465	C9	1.84E-07	0.34	+	55506.6	AT5G18030
MOE	JCVI_21226	C8	1.65E-06	0.29	+	36286.6	AT1G11580
MOE	JCVI_39585	C6	4.77E-07	0.36	+	23108.4	-
MOR	EX121877	A4	7.41E-06	0.26	+	17023.3	AT2G36810
MOR	AM062044	A6 (peak)	3.82E-06	0.24	+	22023.2	AT4G27710
MOR	CD841065	A8	1.10E-04	0.17	+	825.8	AT3G58670
MOR	JCVI_24465	C9	1.935E-05	0.29	+	55506.6	AT5G18030

Trait	GEM ID	linkage group	P value	R^2	Trend	PSEUDO POS	Candidate gene
Second moment of area	JCVI_21955	A5	2.42E-07	0.28	+	145.3	AT2G47600
Second moment of area	EX071451	C1	3.58E-07	0.21	+	47243.74	AT3G13080
Second moment of area	EX123229	C7	1.91E-06	0.22	+	43255.7	AT4G18630
Second moment of area	JCVI_10394	C7	3.21E-06	0.22	+	50855.1	AT5G61140
Stem diameter	JCVI_21955	A5	4.61E-07	0.26	+	145.3	AT2G47600
Stem diameter	JCVI_7586	A5 (peak)	9.49E-06	0.27	-	24082.4	
Stem diameter	JCVI_19323	A9	1.14E-06	0.25	+	5991.9	AT2G19130
Outer cortex thickness	JCVI_25901	A2	4.71E-06	0.26	+	23739.9	AT5G25610
Outer cortex thickness	EV133973	A10	1.35E-05	0.2	+	9862.9	AT5G17860
Outer cortex thickness	JCVI_26302	C1	2.97E-05	0.2	-	46708.9	AT3G13750
Outer cortex thickness	EV189273	C4	1.45E-05	0.19	+	25555.3	-
Outer cortex thickness	JCVI_26403	C7	1.12E-05	0.22	-	29074.8	AT4G17810
Plant height	EE440437	A5	2.98E-07	0.26	-	4809.8	AT2G35630
Stem weight	JCVI_40872	A2 (peak)	4.47E-07	0.23	+	8887.9	AT1G66880
Stem weight	JCVI_21955	A5	4.09E-06	0.24	+	145.3	AT2G47600
Stem weight	EX058924	A8	1.82E-06	0.18	+	9370.5	AT4G34390
Stem weight	JCVI_20079	A8	4.07E-06	0.17	+	9464.4	AT4G34138
Stem weight	jcvi_19323	A9	2.38E-08	0.35	+	5991.9	AT2G19130

3.4b

3.2.2.1 Fmax

For this absolute strength trait, two very clear SNP association peaks were identified. The first and most significant of these was found to be on chromosome A2/C2 (Figure 3.14A and B). The most significant marker within this peak was JCVI_31359:1657 (P=4.45E-05, trait effect=30.9 % (a percentage based on trait range across accessions)). This region was also detected (although to a much weaker extent) in the SNP association analysis carried out for the KWS 2010 material for the absolute strength trait F/V (Figure 3.14C and D). The most significant marker here was JCVI_31359:877 (P=1.76E-03). Using pseudomolecule V4, these markers were found to be in close proximity to two genes orthologous to Arabidopsis genes expected to have pectin methylesterase (PME)/pectin methylesterase inhibitor (PMEI) activity, AT5G50030 and AT5G50060. As described in Section 1.3.3, a number of studies have found that the methylesterification state of cell wall pectin has a key role to play in determining stem mechanical strength (Hongo et al., 2012). Pectin methylesterification is also known to effect cell expansion (Wolf et al., 2009), a process which likely contributes to variation in stem thickness (which as seen in Section 3.1, explains a high level of variation in stem absolute strength in *B. napus*).

The second SNP association for this absolute strength trait was detected on chromosome A5/C5, with the top marker being JCVI_39914:1128 (Figure 3.15A and B). This marker was assigned a P value of 4.07E-04 and a trait effect of 26.9%. This region was also detected in the results obtained for stem weight in the KWS 2011 data (Figure 3.15C and D). The most significant marker within this association peak was JCVI_14857:88 (P=8.52E-04). Within close proximity to these markers, are several very plausible candidate genes. One of these is again an orthologue of an Arabidopsis gene described as having PMEI/PME activity (AT3G12880). The second potential candidate gene is an orthologue of the Arabidopsis *SAUR-LIKE AUXIN RESPONSE 72* (*SAUR72)* (AT3G12830). SAURs have been implicated in regulating cell expansion and final organ size (Qiu et al., 2013). Furthermore, Sanchez et al (2012), reports an important role of auxin in contributing to lateral stem growth (Sanchez et al., 2012). Interestingly, there is also evidence to suggest that auxin signalling and the pectin methylesterification state of the plant cell wall may be related. For example, work carried out in Arabidopsis found that the removal of methyl ester groups (demethylesterification) from cell wall pectin may take place

in response to an upstream accumulation of auxin. This work suggests that it may be the combined effect of auxin and pectin demethylesterification that is required to promote cell expansion and organ growth (Braybrook and Peaucelle, 2013). Given this, it is possible that the association peak detected on chromosome A5/C5 is capturing variation at both the *SAUR72* and the PME/PMEI loci, which may be working together or independently to effect stem mechanical strength. Given the high level of variation in stem absolute strength which can be explained by stem diameter in *B. napus*, it certainly seems plausible to suggest that these genes, which are known to be involved in cell expansion (a key contributor to organ size (Breuninger and Lenhard, 2010)) may be good candidate genes.

In addition to these two very clear SNP association peaks, two minor peaks were also detected for Fmax. The first of these was seen on chromosome A8/C3 (Figure 3.16A and B). The most significant marker in this association peak was JCVI_639:522 (P=5.86E-04, trait effect 33.6%). This marker is also close to a very good candidate gene for mechanical strength. This is an orthologue of Arabidopsis *INTERFASCICULAR FIBERLESS 1* (*IFL1*) (AT5G60690), which is known to have a role in determining stem mechanical strength, at least in this model plant species. Mutants for this gene in Arabidopsis exhibit much reduced levels of interfascicular fiber cells in the stem, and are unable to maintain an upright growth habit (Zhong and Ye, 1999). An additional, minor SNP association peak was detected on chromosome A3/C3 (Figure 3.16C and D), with the top marker, JCVI_38663:382 (P=4.16E-05, trait effect 25.23 %). No candidate genes were identified in this region.

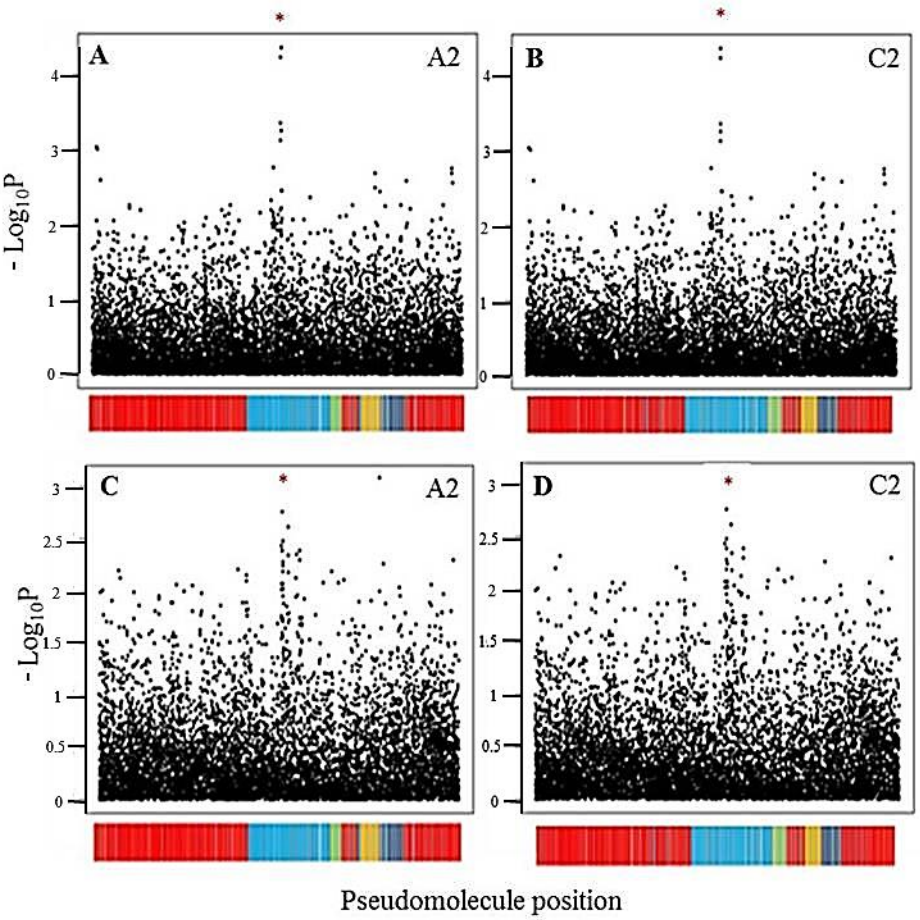

Figure 3. 14. SNP association peak detected for Fmax on chromosome A2/C2. Panels A and B show the detection of this association in the JIC trial analysis. This region was also seen in the KWS 2010 trial analysis (C and D). SNP associations can be seen marked with red asterisks. Marker P values are plotted at $-Log_{10}P$ in pseudomolecule position.

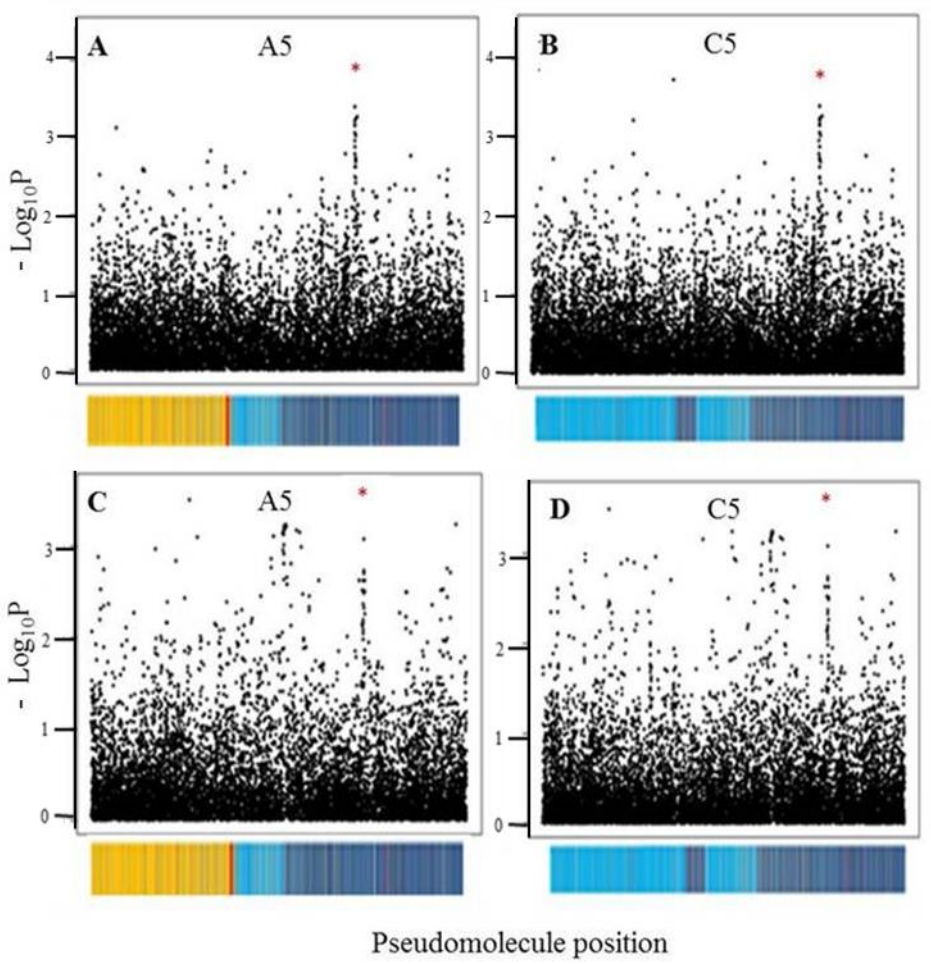

Figure 3. 15. The detection of a common SNP association for Fmax in the JIC trial analysis (Panels A and B) and stem weight in the KWS 2011 trial (Panels C and D) on chromosome A5/C5. SNP associations can be seen marked with red asterisks. Marker P values are plotted at $-Log_{10}P$ in pseudomolecule position.

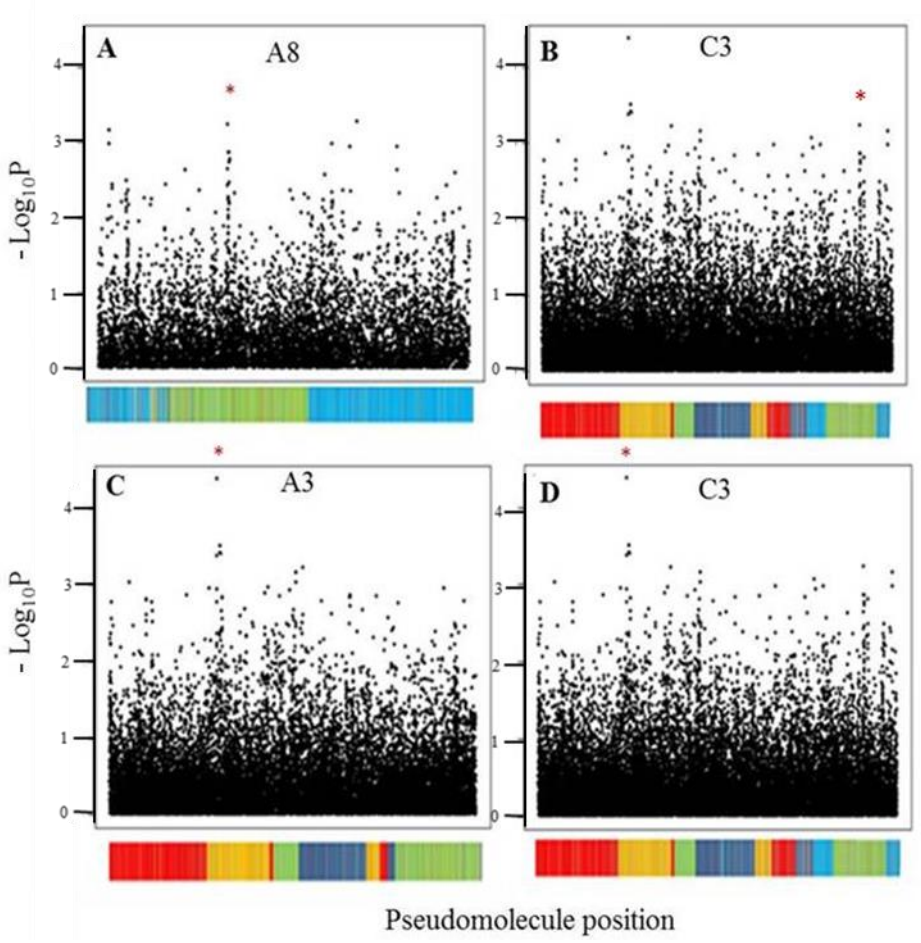

Figure 3. 16. Two SNP associations detected for Fmax in the analysis carried out on JIC trial data. One SNP association was detected on chromosome A8/C3 (A and B) and the other on chromosome A3/C3 (C and D). SNP associations can be seen marked with red asterisks. Marker P values are plotted at $-Log_{10}P$ in pseudomolecule position.

In addition to these very promising SNP associations, highly associating GEM markers were also detected for this trait. A summary of these associations is given in Table 3.4a. One of these GEMs, JCVI_28286, detected on chromosome C7, corresponds to an orthologue of an Arabidopsis *CYTOCHROME P450* (CYP450) (AT4G27710). The expression of this gene correlates negatively with Fmax (P=5.41E-08, R^2=0.34) (Figure 3.17A). In Arabidopsis, an increase in expression of this gene has been linked with increases in localised cell wall deposition (McCurdy et al., 2008). To further explore the role of this gene in contributing to stem mechanical strength in *B. napus*, the respective transcript abundance values for this marker were mapped as a trait against the SNP data. This analysis revealed very

clear SNP associations. The clearest of these associations were seen on chromosome A9/C8 (Figure 3.17D and E) and A6/C6 (Figure 3.17B and C). The most significant marker detected for the A6/C6 association was JCVI_11271:1070 (P=1.09E-05). The second association, seen on A9/C8, had a much stronger signal, with the most significant marker being JCVI_7691:268 (P=5.44E-07). In close proximity to the SNP association on A6/C6 there are several genes involved in cell wall biosynthesis. The first of these is an orthologue of Arabidopsis *INCURVATA 4* (AT1G52150). This gene is predicted to act as a transcriptional regulator, important for vascular development and cell differentiation. In addition, two genes known to contribute to cell wall mannan structure were identified. Firstly, there is a gene orthologous to Arabidopsis *MANNAN SYNTHESIS-RELATED 2* (AT1G51630). Cell wall mannans are thought to bind to cellulose, providing structural integrity to the cell wall (Rodríguez-Gacio et al., 2012). As a second example, an orthologue of Arabidopsis *ALPHA MANNOSIDASE 1* was identified. Mannosidases are a group of enzymes, involved in the cleavage of linkages between cell wall mannose and other cell wall polysaccharides (Mast and Moremen, 2006). No clear candidate genes were identified in close proximity to JCVI_7691:268. These associations suggest that there may be an interaction between these loci and the CYP450 gene detected in the GEM analysis. Based on the candidate genes identified in the A6/C6 region and the expected role of this CYP450 gene in contributing to cell wall biogenesis, these findings may be indicative of a cell wall biosynthesis pathway which contributes to stem mechanical strength in *B. napus.* Further experimental work would allow for the presence of this proposed interaction to be confirmed. It may for example be of interest to assess knock-out mutants for the proposed candidate genes identified in the A6/C6 region. Following loss of function, it would then be possible to assess any changes in gene expression of the *CYP450* gene identified in the GEM analysis. It may also be of interest to assess cell wall composition in a knock out mutant for this CYP450 gene.

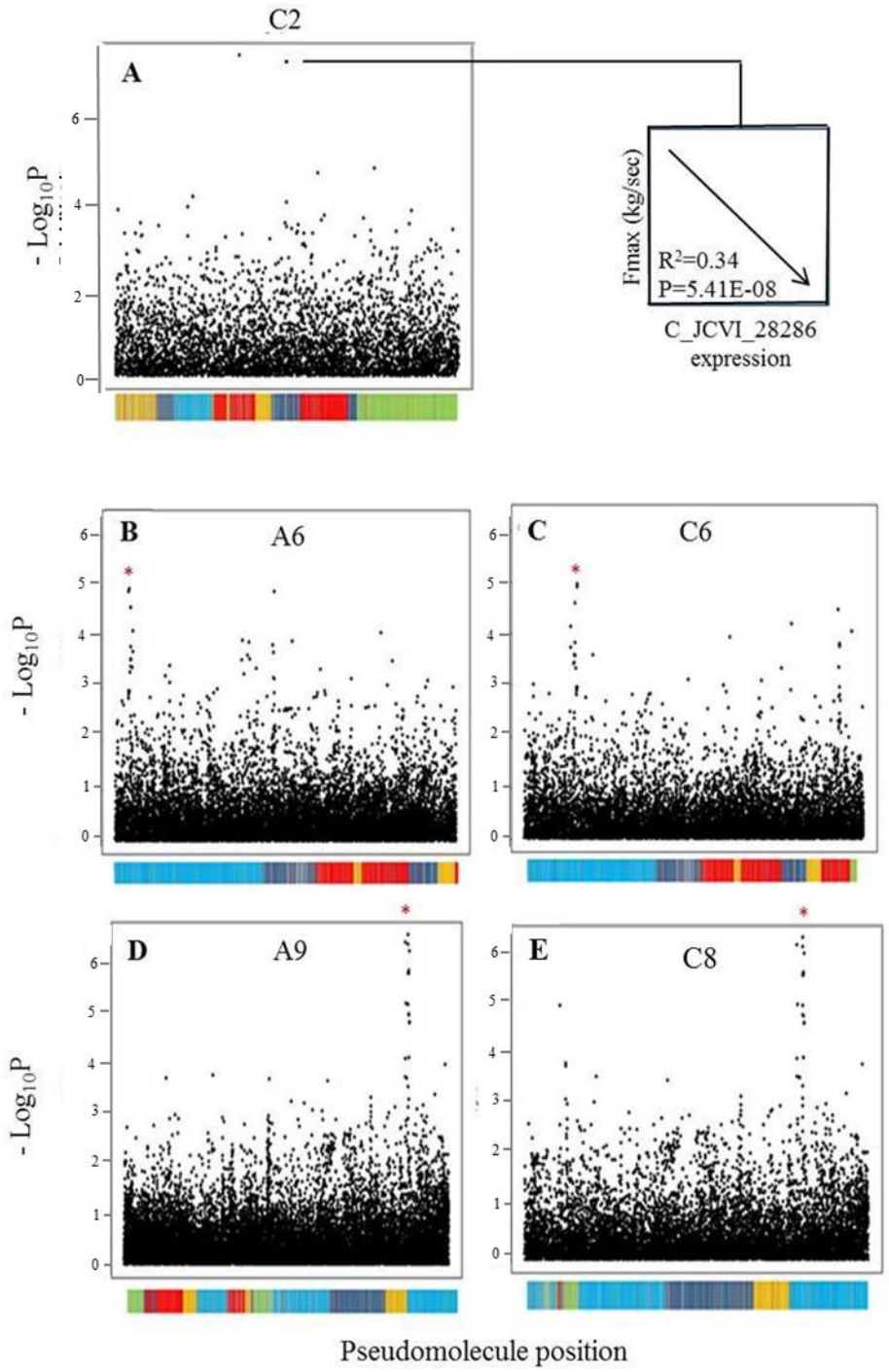

Figure 3. 17. SNP association mapping of transcript abundance (measured as RPKM) of the associating GEM, JCVI_28286, which correlates negatively with Fmax (A), as a trait. Two SNP associations were identified: one on chromosome A6/C6 (panels B and C) and the other on chromosome A9/C8 (panels D and E). SNP associations can be seen marked with red asterisks. Marker P values are plotted at $-Log_{10}P$ in pseudomolecule position.

3.2.2.2 F/V

Although not as clear, the previously described association on chromosome A2/C2 for Fmax was also seen for F/V in the JIC trial analysis (Figure 3.18A and B). The top marker for this association was again JCVI_31359:1657 (P=5.73E-05, trait effect 24.6 %). This region was also detected for F/V in the KWS 2010 analysis (Figure 3.18C and D). In addition, the previously described association on chromosome A5/C5 for Fmax was also seen for F/V (Figure 3.19A and B). As seen for Fmax, the most significant marker is JCVI_39914:1143 (P=4.41E-04, trait effect 21.7 %). This region was also detected for stem weight in the KWS analysis (Figure 3.19C and D).

A number of minor association peaks were also detected for F/V. An example of one of these is the weak association on A6/C5 where the top marker is JCVI_4281:518 (P=1.53E-04, trait effect 19.2 %) (Figure 3.20A and B). This region was also detected in the 2011 KWS analysis for Fmax, where the top marker was JCVI_1432:687 (P=1.86E-03) (Figure 3.20C and D). No potential candidate genes were detected in this region.

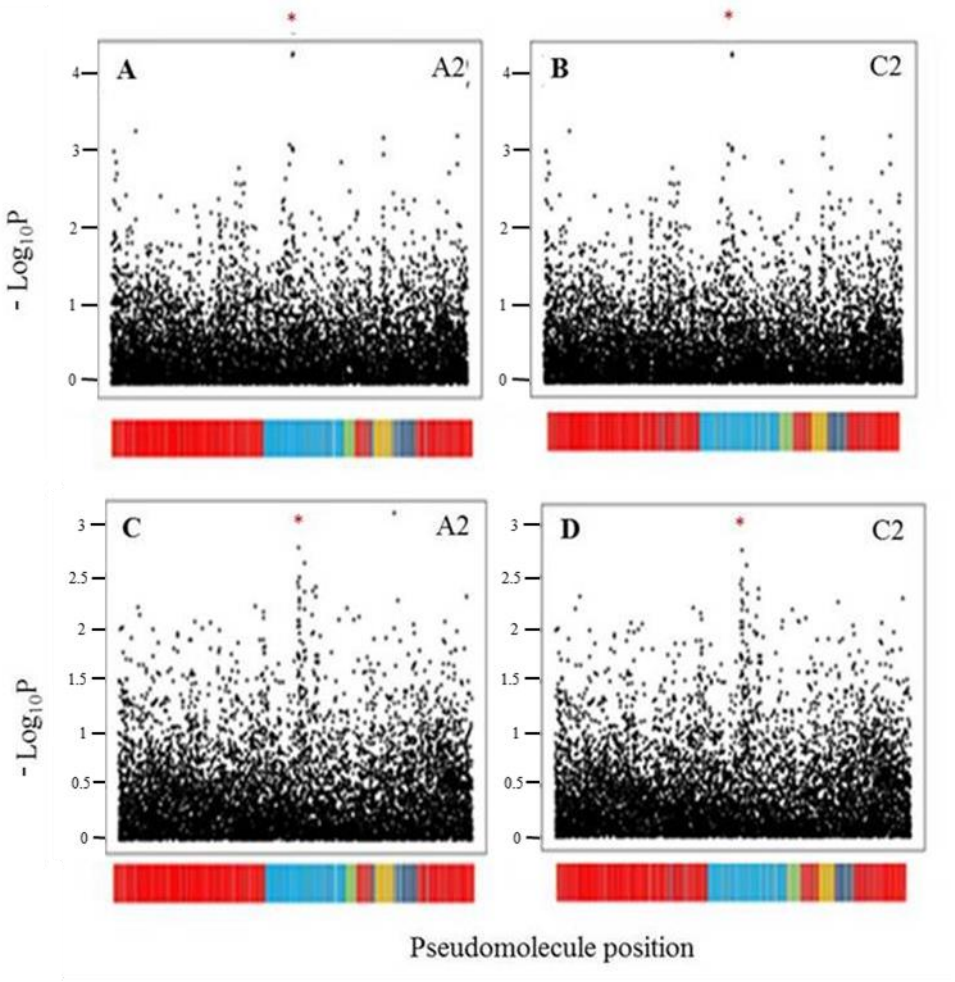

Figure 3. 18 The detection of a common SNP association peak between for the absolute strength trait F/V in the results obtained from both the JIC field. SNP associations can be seen marked with a red asterisk. Marker P values are plotted at $-Log_{10}P$ in pseudomolecule position.

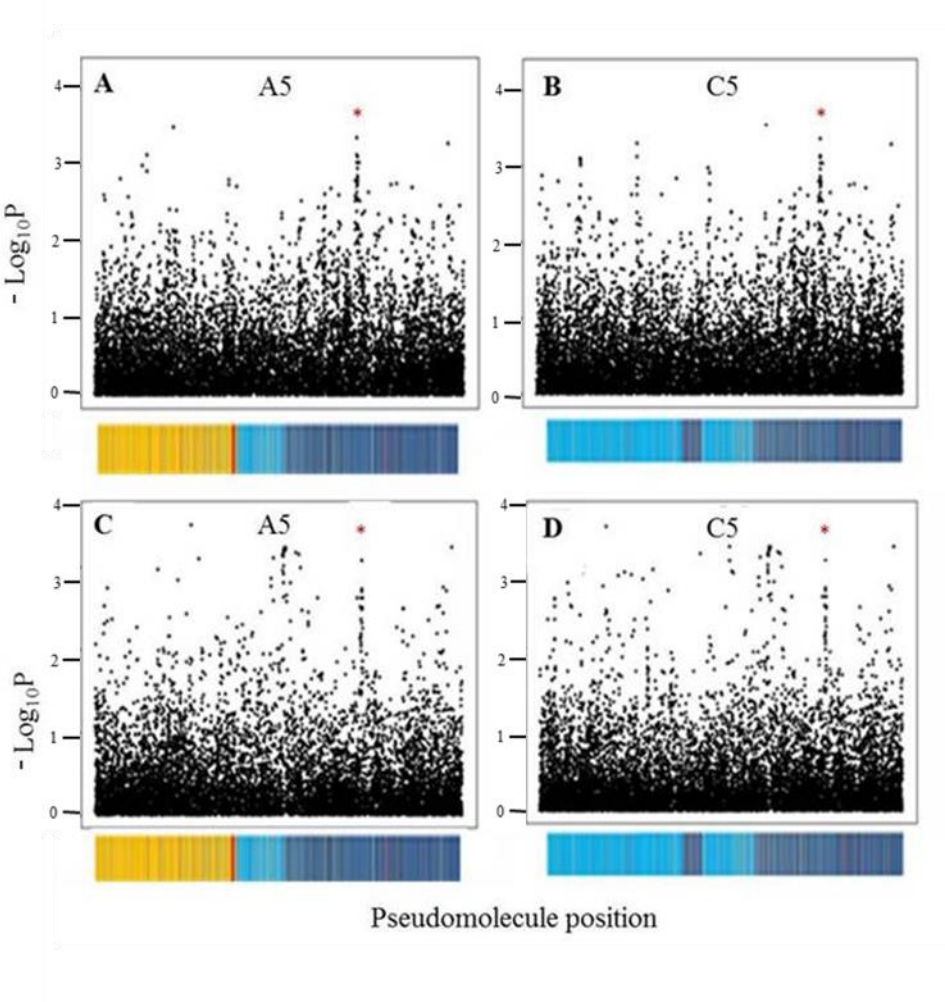

Figure 3. 19.A common association detected on chromosome A5/C5 for F/V in the JIC trial analysis (A and B) and stem weight (C and D) in the results obtained from the analyse carried out on the data obtained from the KWS 2011 field trial. SNP associations can be seen marked with red asterisks. Marker P values are plotted at $-Log_{10}P$ in pseudomolecule position.

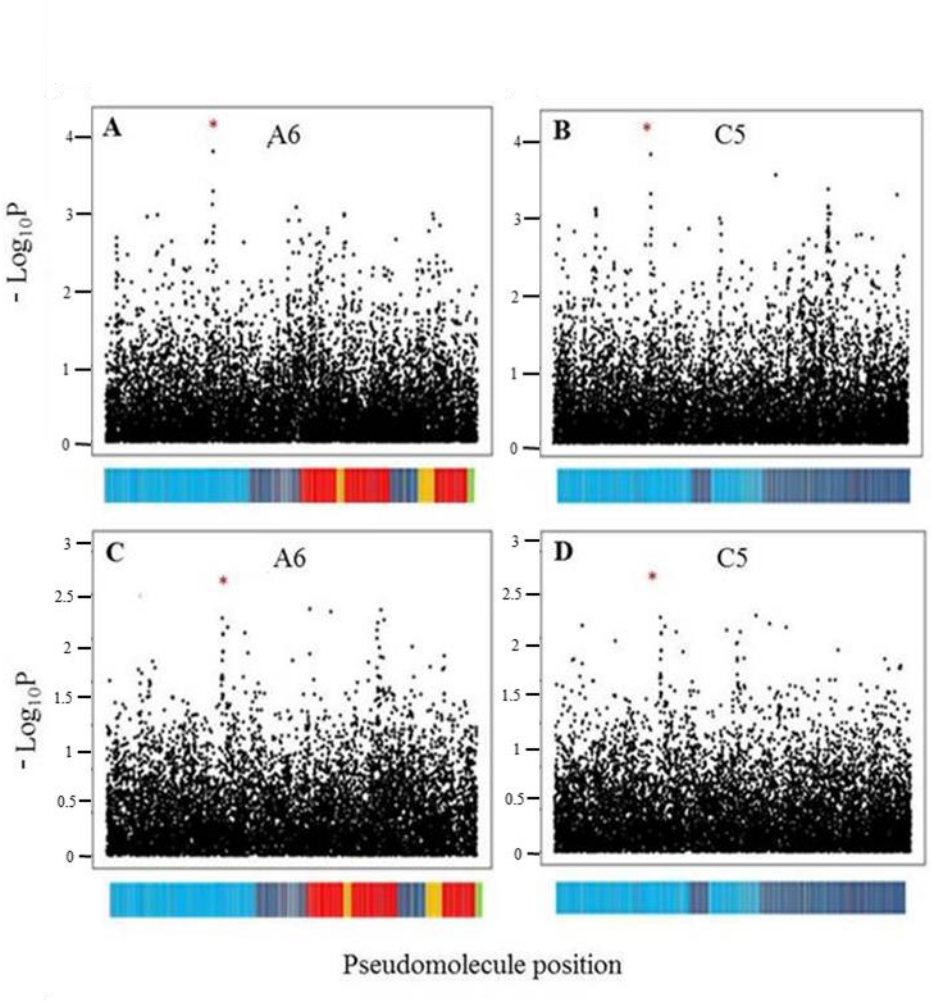

Figure 3. 20. A common SNP association detected on chromosome A6/C5 for F/V (A and B) in the JIC trial analysis and Fmax in the analysis carried out on the data obtained from the 2011 KWS field trial (C and D). SNP associations can be seen marked with a red asterisk. Marker P values are plotted at $-Log_{10}P$ in pseudomolecule position.

In addition to these SNP associations, GEM associations (common with those seen for Fmax) were detected for F/V. These are summarised in Table 3.4a.

3.2.2.3 MOE

The Associative Transcriptomics analysis for MOE, detected several interesting loci. The first of these associations was on A1/C1 (Figure 3.21C and D). The most significant marker within this peak is JCVI_19955:183 (P=2.56E-08, trait effect 61.9 %). In close proximity to this marker in the pseudomolecule is an orthologue of the Arabidopsis transcription factor, *VND2* (AT4G36160). *VND2* has been implicated in xylem differentiation and cell wall biosynthesis in Arabidopsis (Ko et al., 2007), making this a very plausible candidate gene for stem material strength in *B. napus.*

A second very strong association for MOE was detected on chromosome A6/C6 (Figure 3.21A and B). The most significant marker for this association peak is JCVI_10685:430 (P=2.63E-06, trait effect 45.5 %). Interestingly, this region was also detected when mapping the GEM, JCVI_28286, against the SNP data in the Associative Transcriptomics analysis described previously for Fmax and F/V. As previously mentioned, within this region are a number of genes with expected roles in cell wall biosynthesis.

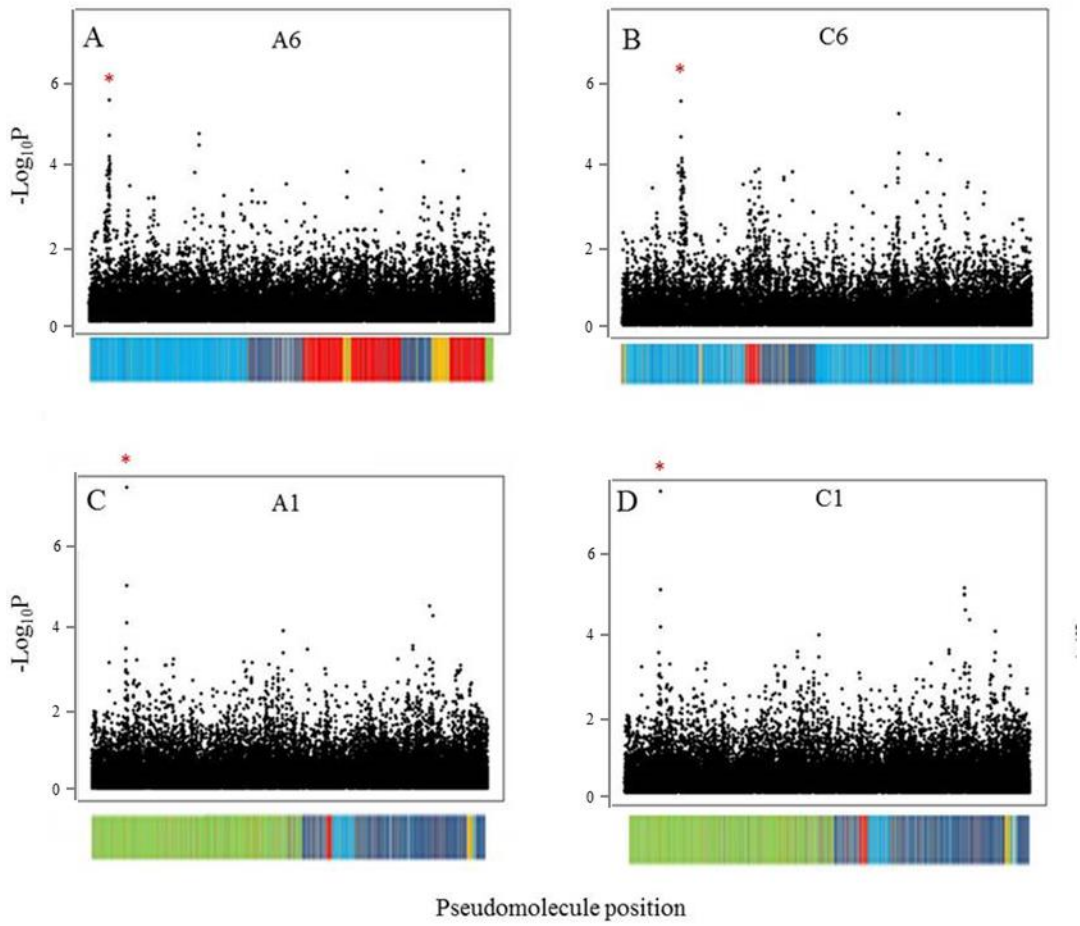

Figure 3. 21. SNP associations for the material strength trait, MOE, detected on chromosome A6C6 (A and B) and chromosome A1/C1 (C and D). SNP associations are marked with a red asterisk. Marker P values are plotted at –Log_{10}P in pseudomolecule position.

In addition to these very clear association peaks, a number of minor peaks were also detected for MOE. The first was on A2/C2, where the top marker is JCVI_22602:1182 (8.38E-06, trait effect 37.2 %) (Figure 3.22A and B). No candidate genes were detected in this region. The second of these minor associations was on chromosome A7/C6. The most significant marker for this SNP peak is JCVI_19824:614 (P=5.29E-06, trait effect 47.6 %) (Figure 3.22C and D). Again, no candidate genes were identified in this region.

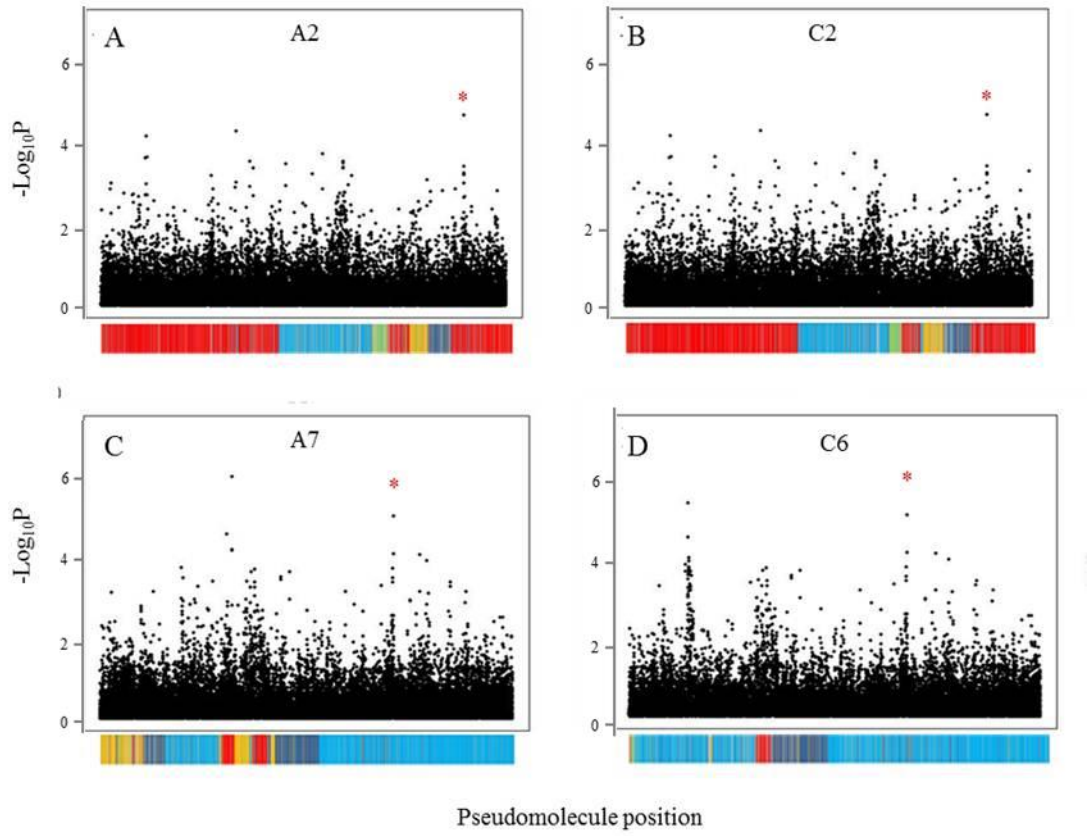

Figure 3. 22. Minor SNP associations detected for MOE on chromosomes A2/C2 (A and B) and A7/C6 (C and D). SNP associations can be seen marked with a red asterisk. Marker P values are plotted at $-Log_{10}P$ in pseudomolecule position.

The GEM analysis results for this trait identified a number of highly associated markers. The most promising of these GEM associations are summarised in Table 3.4a. Transcript abundance levels for each of the GEMs detected for MOE show a positive correlation with the MOE trait values. One associating GEM, JCVI_21226 (P= 1.65E-06, R^2=0.29), detected on chromosome C8, corresponds to an orthologue of a gene described in Arabidopsis as a *METHYLESTERASE PCR A* (AT1G11580). This gene has been described as having pectinesterase (possibly inhibitory) activity and is predicted to have a role in cellulose metabolism and pectin biosynthesis. A second single GEM association for MOE, on chromosome A1, was JCVI_14486 (P=4.95E-07, R^2=0.30) (Figure 3.23A). This marker corresponds to an orthologue of

Arabidopsis described as a Protein kinase gene (AT3G13690). Although this gene may not have any obvious role in determining stem material strength, through mapping the transcript abundance values (measured as RPKM) of this GEM as a trait against the SNP data, a small SNP association peak was identified on chromosome A7/C6, in a region previously detected for MOE in the SNP analysis (Figure 3.23C and D). The top marker for this association is JCVI_19824:257 (P=4.95E-07). Given that there are no clear candidate genes in the JCVI_19824 region and that the gene model to which the GEM corresponds has no clear role in contributing to the material strength of plant stems, it is not possible to form a hypothesis regarding the relationship between these loci. However, the results presented here suggest the presence of a genetic interaction.

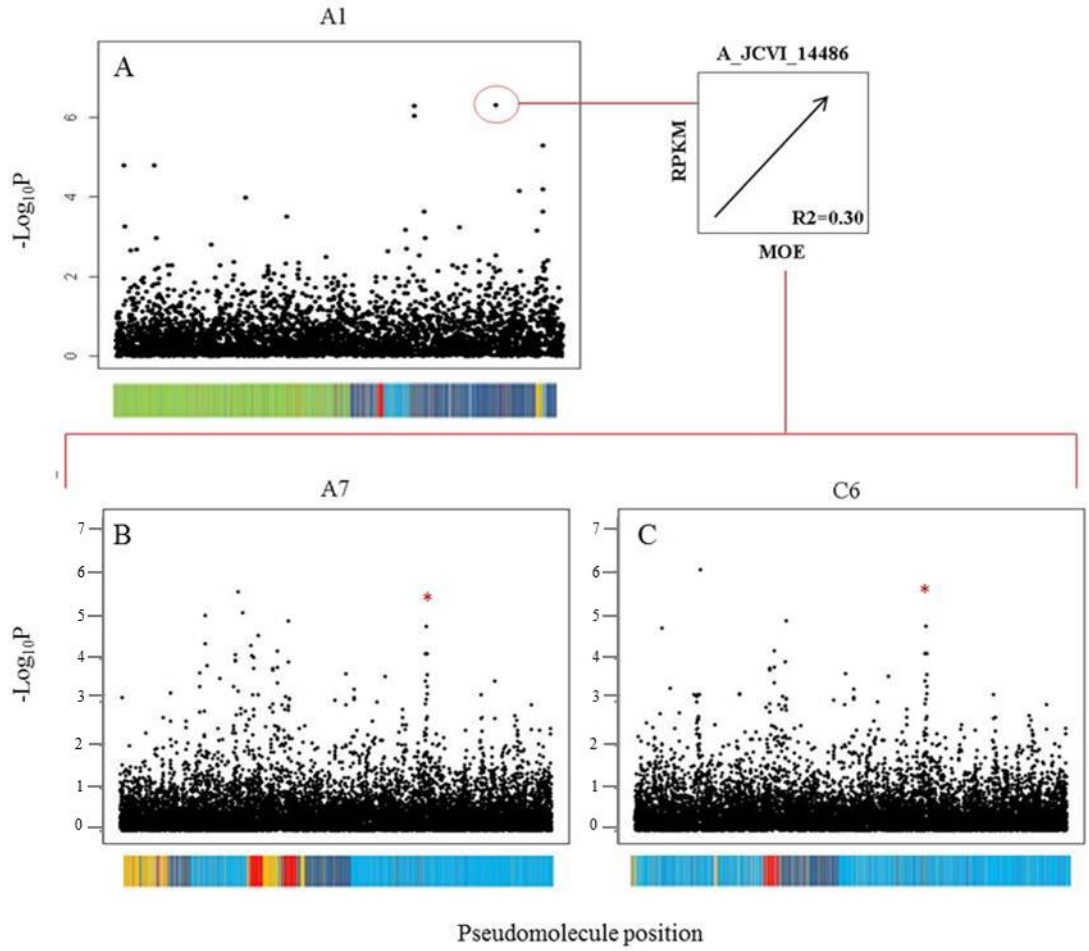

Figure 3. 23. The detection of a highly significant GEM marker on chromosome A1 which correlates positively with MOE. The RPKM values for this A_JCVI_14486 correlate positively with MOE (A). Panels B and C show SNP associations detected following the mapping of transcript abundance of A_EE440347 (measured as RPKM) as a trait. SNP associations can be seen marked with a red asterisk. Marker P values are plotted at –Log$_{10}$P in pseudomolecule position.

On chromosome A6, a small peak of GEMs was identified. Each of these GEMs correspond to a *CYTOCHROME P450* (709 family) in Arabidopsis (AT4G27710). A paralogue of this gene was also detected in the absolute stem strength GEM analysis. When mapping transcript abundance of the most significant of these A6 GEMs, AM062044 (P=7.28E-07, R^2=0.30) (Figure 3.24A)), as a trait against the SNP data, a

weak SNP association was identified on A1/C1 (3.24B and C), in the same region previously described in the SNP analysis for MOE. This region, as previously mentioned, contains an orthologue of the Arabidopsis transcription factor, *VND2*, which may again link this CYP450 gene with cell wall biosynthesis in *B. napus.* It is possible that *VND2* has a role in regulating the gene expression of this *CYP450* gene directly as part of a pathway which ultimately contributes to stem strength in *B. napus.*

For both paralogues of this CYP450 gene detected in the GEM analyses for Fmax and MOE, transcript abundances, when mapped as traits against the SNP data, revealed loci in close proximity to transcriptional regulators involved in plant vasculature development (*VND2* and *INCURVATA 4*). This may suggest that this CYP450 gene is contributing to stem mechanical strength in *B. napus*, as part of a pathway which contributes to the development of vascular tissues. This seems to be a very plausible hypothesis, given the many Arabidopsis mutants described in the literature which exhibit both xylem defects and reduced stem mechanical strength (Jones et al., 2001; Persson et al., 2007).

A further interesting GEM, A_CD841065, was identified in this analysis on chromosome A8 (P=2.54E-07, R^2=0.31) (Figure 3.25A). A second GEM, JCVI_15213, was also detected within the same region (P= 1.32E-06, R^2=0.28). These markers each correspond to an orthologue of a gene of unknown function (DUF1637) in Arabidopsis (AT3G58670). When mapped against the SNP data, the transcript abundance for the first of these GEMs revealed a number of very clear SNP associations. Firstly, two SNP associations can be seen on A1/C1 (Figure 3.25B and C). The first of these coincides with a SNP peak previously detected for MOE. As for MOE, the most significant marker within this association peak is JCVI_19955:183. As previously discussed, this A1/C1 locus contains a transcription factor, *VND2*, which has been implicated in xylem development and cell wall biosynthesis. The second association on A1/C1 was more significant, with the most highly associating marker being JCVI_ 1969:463. No clear candidate genes were identified within this region. The third strong association detected for this GEM, was on chromosome A6/C6 (Figure 3.25D and E). Again this region was previously detected in the MOE SNP analysis. In this region, a number of genes involved in cell wall biosynthesis were described. The observed interactions between SNP variation

and the differential expression of a GEM, corresponding to a gene of unknown function (DUF1637), may provide clues of a potential pathway contributing to stem material strength in *B. napus*. Given that the transcript abundance for this gene maps to two loci containing orthologues of genes involved in xylem development, it is possible that this DUF1637 gene (as proposed for the CYP450 gene previously described) is contributing to stem material strength as part of a pathway involving xylem development.

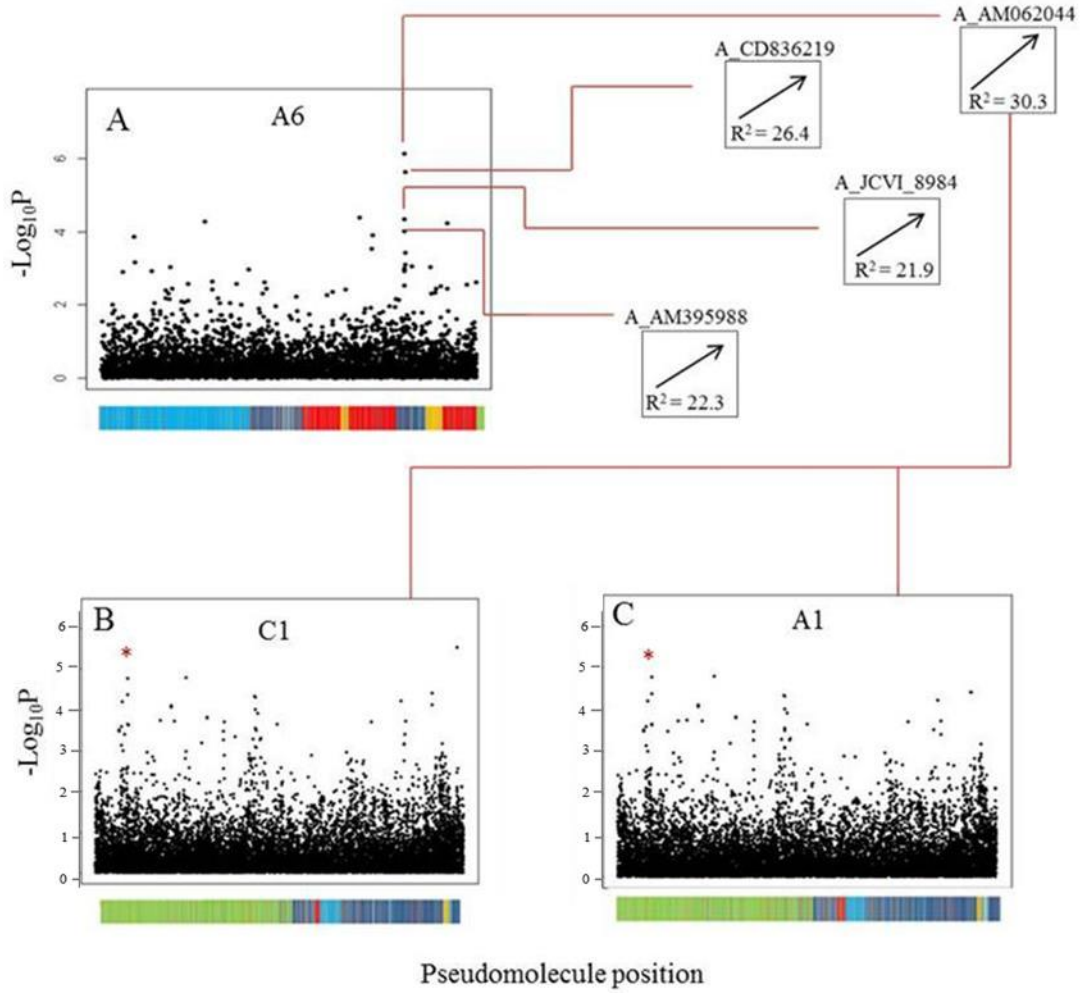

Figure 3. 24. The detection of a small peak of GEMs on chromosome A6 (A). SNP association mapping of transcript abundance of A_AM062044 (measured as RPKM) as a trait revealed an association peak on chromosome A1/C1 (B and C). This region was previously detected for MOE in the SNP analysis. SNP associations can be seen marked with a red asterisk. Marker P values are plotted at $-Log_{10}P$ in pseudomolecule position.

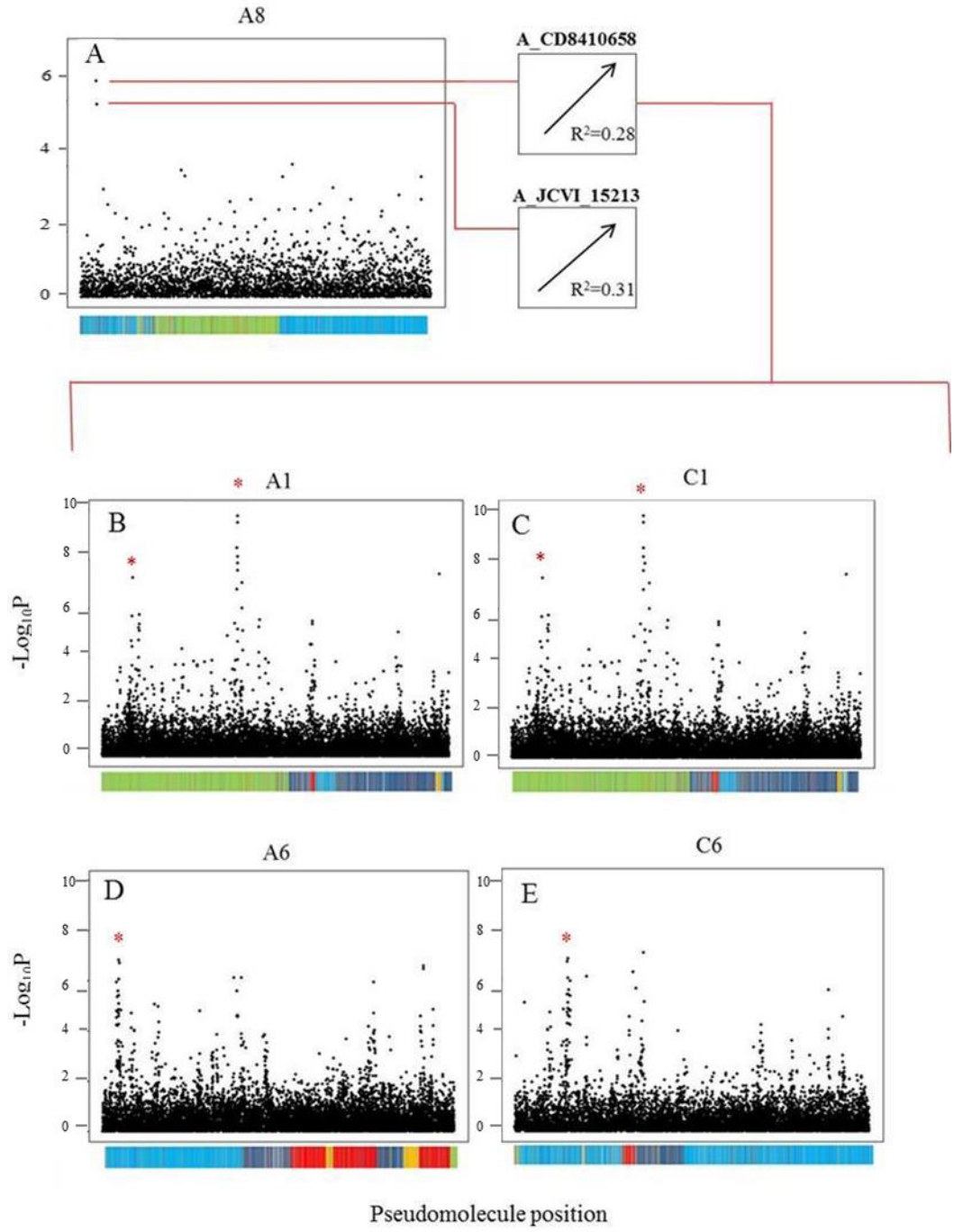

Figure 3. 25. Two highly associating GEM markers, A_CD8410658 and JCVI_15213, on chromosome A8(A). Both of these markers correspond to an Arabidopsis gene model of unknown function (AT3G58670). SNP associations following the mapping of transcript abundance of A_CD8410658 (measured as RPKM) as a trait revealed association peaks on A1/C1 (B and D) and A6/C6 (C and E). SNP associations can be seen marked with a red asterisk. Marker P values are plotted at $-Log_{10}P$ in pseudomolecule position.

3.2.2.4 MOR

For the second material strength trait, MOR, two clear SNP associations were detected. These associations, one on A1/C1 (Figure 3.26A and B) and one on A6/C6 (Figure 26C and D) were also detected for MOE and are described above. In both cases, the most significant markers within these association peaks were the same as those detected for MOE, with JCVI_19955:183 (P=7.46E-06, trait effect 52.6 %) the top marker for A1/C1 and JCVI_10685:430 (P=2.05E-05, trait effect 44.5 %) the most significant in the A6/C6 region. A summary of these results can be found in Table 3.3a.

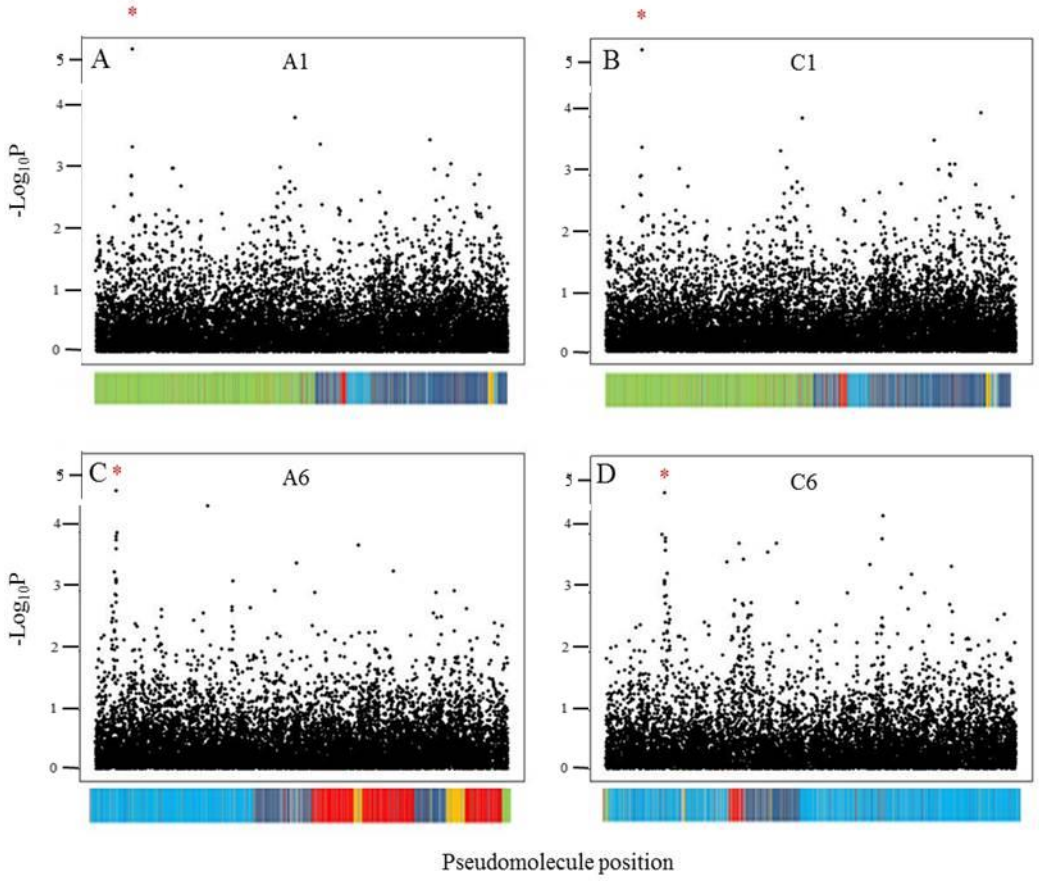

Figure 3. 26. SNP association for the material strength trait MOR on chromosome A1/C1 (A and B) and A6/C6 (C and D). SNP association peaks can be seen marked with a red asterisk. Marker P values are plotted at $-Log_{10}P$ in pseudomolecule position.

In addition, common GEM associations were also seen for MOR. These were also previously described for MOE and can be seen summarised in Table 3.4a.

3.2.2.5 Stem diameter

Only very weak associations were detected in the SNP analysis for stem diameter. Two of these were found to be common associations with Fmax and F/V. The first of these was on chromosome A2/C2. The most significant marker for this association peak was JCVI_31359:1651 (P=1.01E-04, trait effect 25.5 %) (Figure 3.27A and B). In addition, the A5/C5 region, previously described for Fmax and F/V, was also identified for stem diameter (Figure 3.27C and D). The most significant marker for this association was JCVI_18806:670 (P=4.36E-04, trait effect 26.2 %). As previously described, both of these loci are in close proximity to very plausible candidate genes for stem strength with roles in pectin biosynthesis and auxin signalling. Each of these candidates are more specifically plausible for stem structural strength, where as previously mentioned, both auxin-related and pectin methylesterase/inhibitor genes are known to be important for regulation of cell size and organ growth (Sanchez et al., 2012; Qiu et al., 2013). These associations were strongest for the absolute strength traits. This may suggest that the strength is not coming from stem diameter alone. It may also suggest that there is a greater level of human error related to the collection of data for this structural measure in comparison to the absolute strength traits.

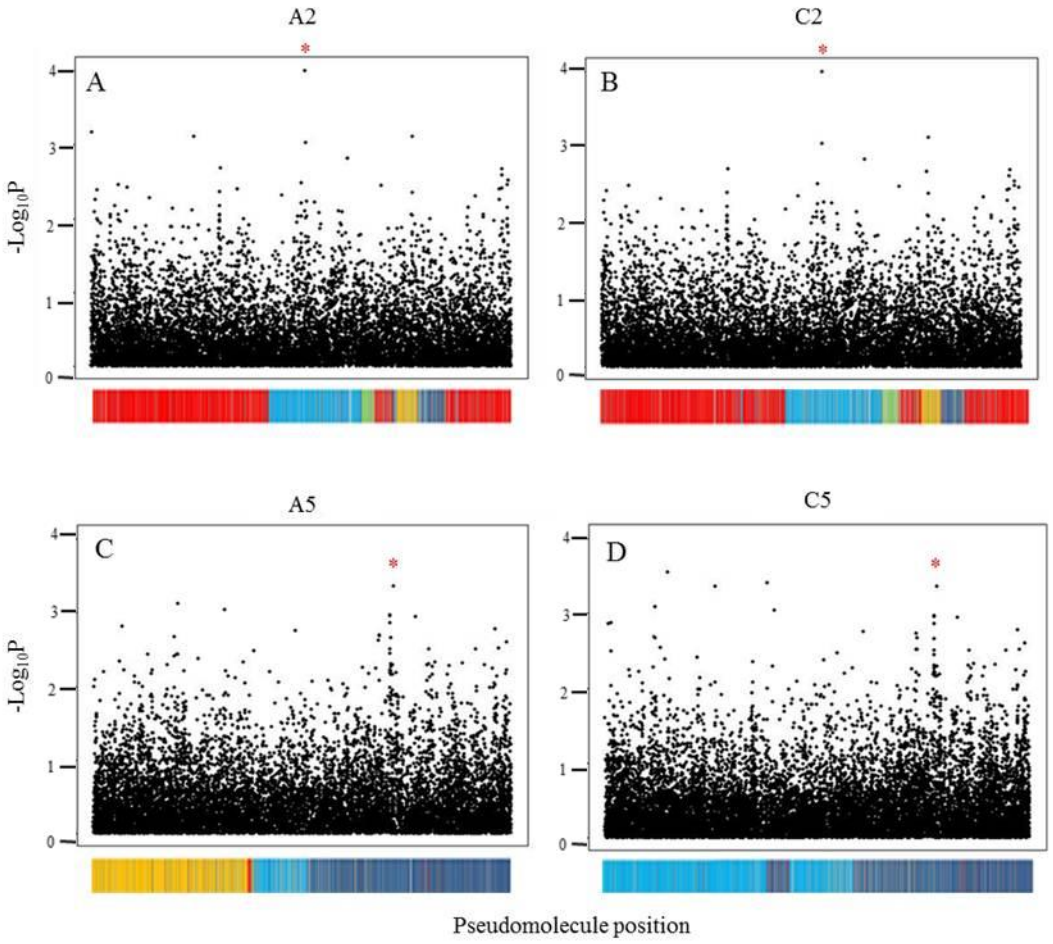

Figure 3. 27. The detection of two SNP associations for stem diameter on chromosome A2/C2 (A and B) and A5/C5 (C and D). These SNP associations were also detected for the stem absolute strength traits Fmax and F/V. SNP associations can be seen marked with a red asterisk. Marker P values are plotted at $-Log_{10}P$ in pseudomolecule position.

A third interesting locus was detected for stem diameter. This region was on A4/C4. The most significant marker was JCVI_38024:112 (P=2.12E-04, trait effect 20.1 %) (Figure 3.28A and B). Although not a strong signal, this association may be promising given that this region was also detected (although again weakly) in the results obtained for KWS 2010. In this analysis, this region was detected for both stem diameter and second moment of area. The signal was however clearer for second moment of area (shown in Figure 3.28C and D) where the most significant marker was JCVI_5344:486 (P=2.03E-03). This association was greatly diminished

in the SNP analysis carried out on the $\log_{10}$ transformed data for second moment of area, KWS 2010 data. However, the detection of a common association in the JIC trial, may suggest that the transformation of the KWS data, results in the masking of important variation. The much broader association peak identified in the KWS analysis in comparison to that of the JIC trial data, illustrates the increased resolution achieved with a greater number of accessions. Within close proximity to the most significant markers detected in the JIC and KWS analyses, is a very good candidate gene for stem diameter (and second moment of area). This gene is an orthologue of Arabidopsis *GALACTURONOSYLTRANSFERASE 5* (*GAUT 5*) (AT2G30575). This gene is thought to be involved in the biosynthesis of cell wall pectin. Arabidopsis mutants for other GAUT genes exhibit altered cell wall composition and reduced cell size and adhesion (Fabiana de et al., 2013).

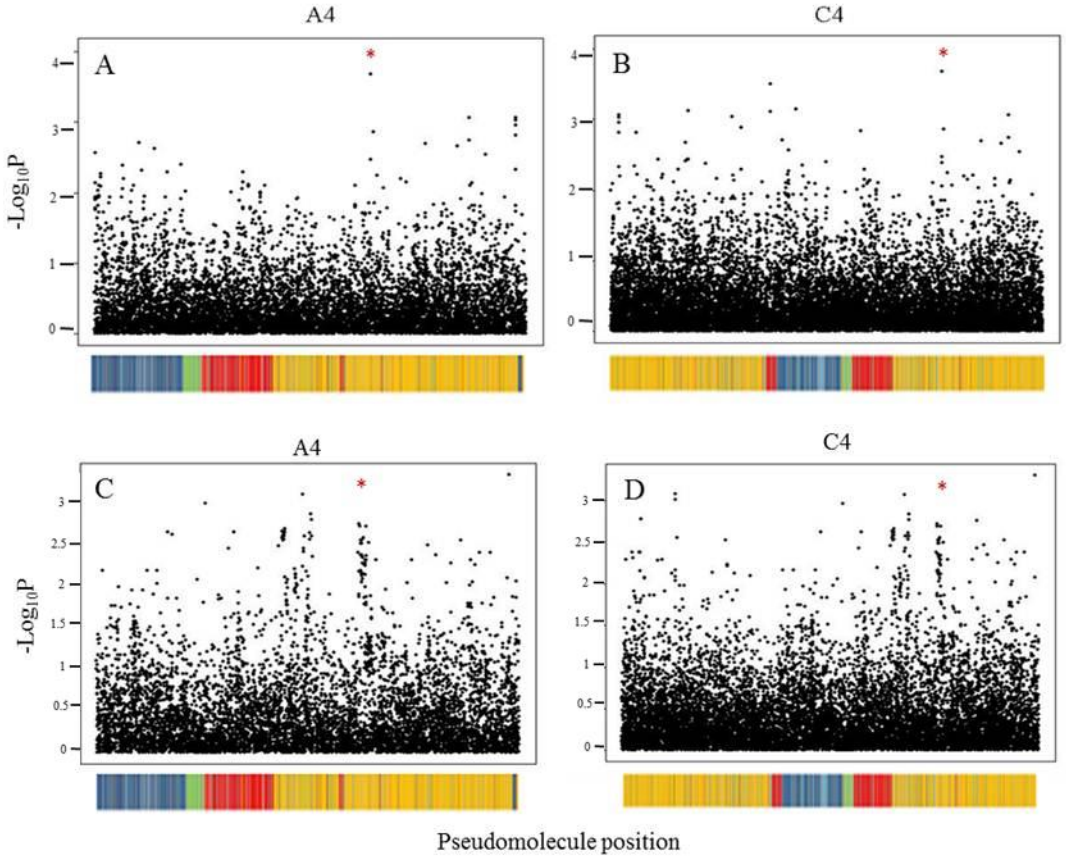

Figure 3. 28. SNP association detected on chromosome A4/C4 in both the JIC trial analysis (A and B) and the KWS 2010 trial analysis (C and D) for stem structural strength. SNP associations can be seen marked with a red asterisk. Marker P values are plotted at $-Log_{10}P$ in pseudomolecule position.

In addition to these SNP associations, a number of GEM associations were detected for stem diameter. Two of these GEM associations, JCVI_ 7586 (P= 9.49E-06, R^2=27) and JCVI_19323 (P=1.14E-06, R^2=0.25), were also detected for the absolute strength traits (Table 3.4b).

3.2.2.6 Second moment of area

No clear SNP associations were detected for this structural strength trait in the analysis for the JIC trial analysis. However, as described previously (when discussing the results obtained for stem diameter) a weak association was detected in the 2010 KWS analysis on chromosome A4/C4 (non-transformed data only) (Figure 3.28). A number of highly associating GEMs were also detected. These are summarised in Table 3.4b.

3.2.2.7 Plant height

One clear association signal was detected for plant height. This was seen on chromosome A3/C7 where the most significant marker is JCVI_26003:352 (P=2.86E-04, trait effect 21.7 %) (Figure 3.29A and B). This marker is in close proximity to a gene orthologous to Arabidopsis *BRI1* (*BRASSINOSTEROID-INSENSTIVE 1*) *SUPRESSOR 1* (AT4G30610). *BRI1* is known to be involved in regulation of plant height in Arabidopsis. Mutants defective for this gene have a very clear dwarfing phenotype (Noguchi et al., 1999). The detection of a suppressor of this gene for plant height in *B. napus* is therefore very encouraging.

A more minor peak was also detected on chromosome A9/C8. The most significant marker is JCVI_19156:300 (P= 8.82E-04, trait effect of 16.8 %) (Figure 3.29C and D). No clear candidate genes were detected in this region.

The GEM analysis identified two highly associating GEM markers on chromosome A5, both of which correspond to a gene orthologous to Arabidopsis *MICROTUBULE ORGANISATION 1* (*MOR1*) (AT2G35630). Arabidopsis mutants for this gene exhibit a temperature-dependant reduction in organ size. This was found to be due to the inability of the mutant plant to correctly organise cortical microtubules (Angela et al., 2001). To further explore the potential role of this gene in contributing to the genetic control of plant height in *B. napus*, transcript abundance levels for the most highly associating of these GEM markers, A_EE440437 (P=2.98E-07, R^2=0.26 (correlating negatively with plant height))(Figure 3.30A), were mapped as a trait against the SNP data. In doing this a very clear SNP association on chromosome A2/C2 was uncovered (Figure 3.30B and C). The most highly associating marker within this peak was JCVI_20133:246. No clear candidate genes were identified within this region. However, the results presented here, suggest that there may be a gene, which resides at this A2/C2 locus, which either directly or indirectly regulates the expression of *MOR1*, which contributes to variation in plant height in *B. napus.*

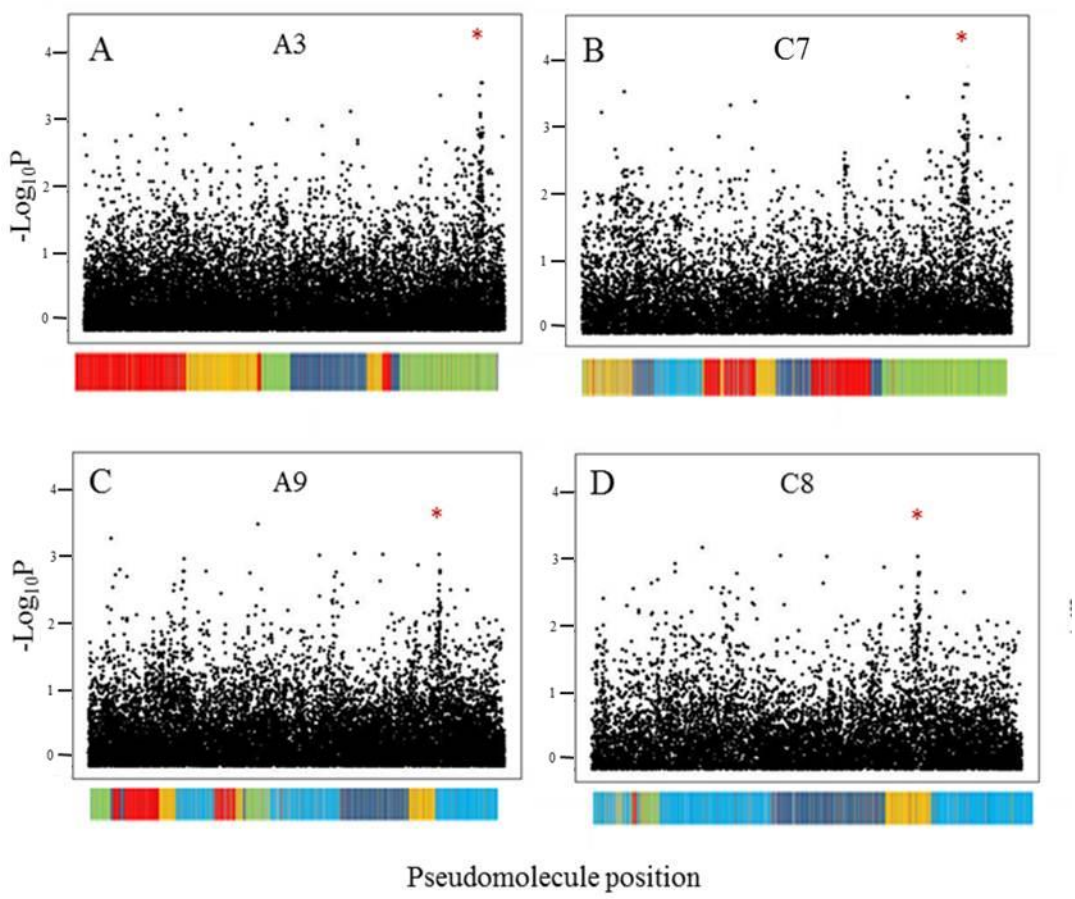

Figure 3. 29. Associative Transcriptomics SNP analysis results for plant height. A SNP association was detected on chromosome A3/C7 (A and B). A minor peak was also detected on A9/C8 (C and D). SNP associations can be seen marked with a red asterisk. Marker P values are plotted at $-Log_{10}P$ in pseudomolecule position.

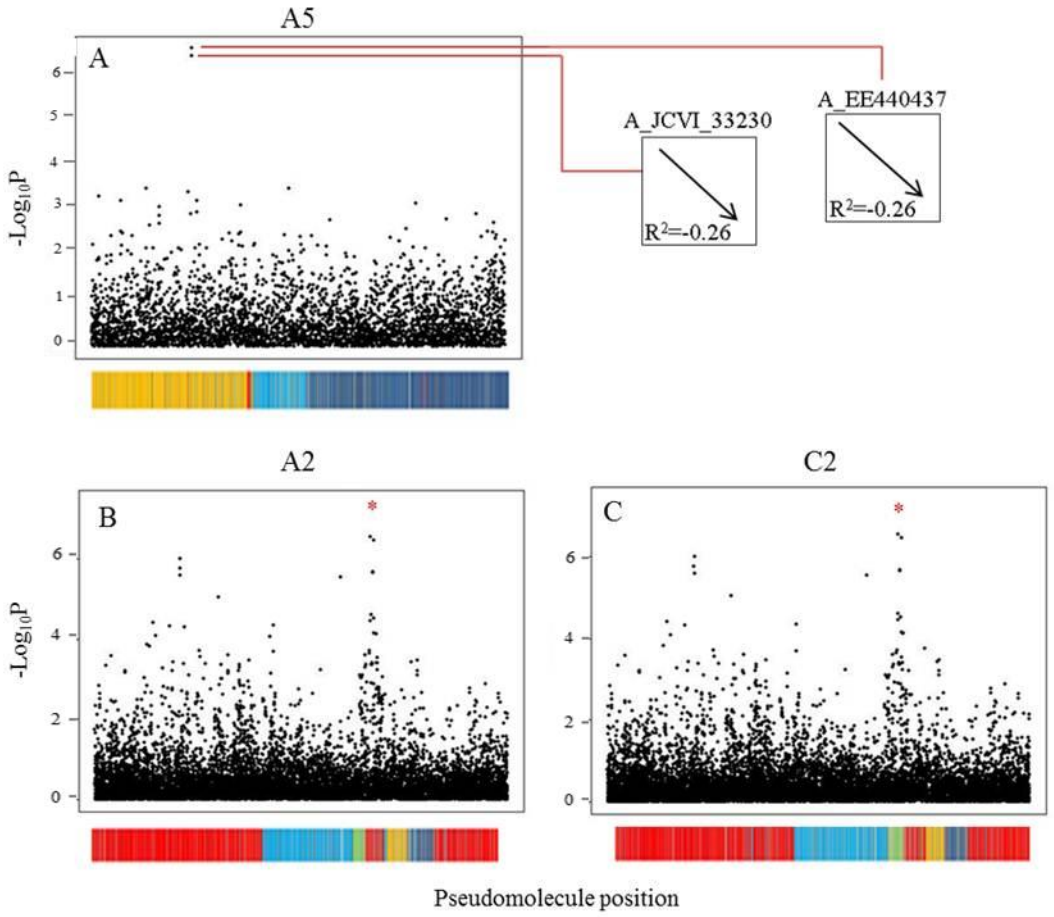

Figure 3. 30. Highly significant GEM associations detected for height on chromosome A5, the RPKM values of which correlate negatively with the trait. A_JCVI_33230 and A_EE440437 correspond to the same gene model which represents an orthologue of the Arabidopsis *MOR1* (A). Mapping A_EE440437 as a trait against the SNP data, identified a SNP association on chromosome A2/C2 (B and C). SNP associations can be seen marked with a red asterisk. Marker P values are plotted at $-Log_{10}P$ in pseudomolecule position.

3.2.2.8 Stem weight

Two main associating loci for stem weight were detected in the SNP analysis. The first of these is on chromosome A4/C4. The most significant marker at this locus was JCVI_33733:304 (P=3.59E-05, trait effect 43 %) (Figure 3.31A and B). The second SNP association is on A9/C9. The most significant marker is JCVI_22620:181 (P=5.38E-05, trait effect 45.3 %) (Figure 3.31C and D). No candidate genes were

identified for these SNP associations. In addition, an association for stem weight was also detected on A5/C5 in the KWS 2011 analysis (Figure 3.31E and F). This association was previously described in Section 3.2.2.2. Significant GEM associations were also detected for this trait. These are summarised in Table 3.4b.

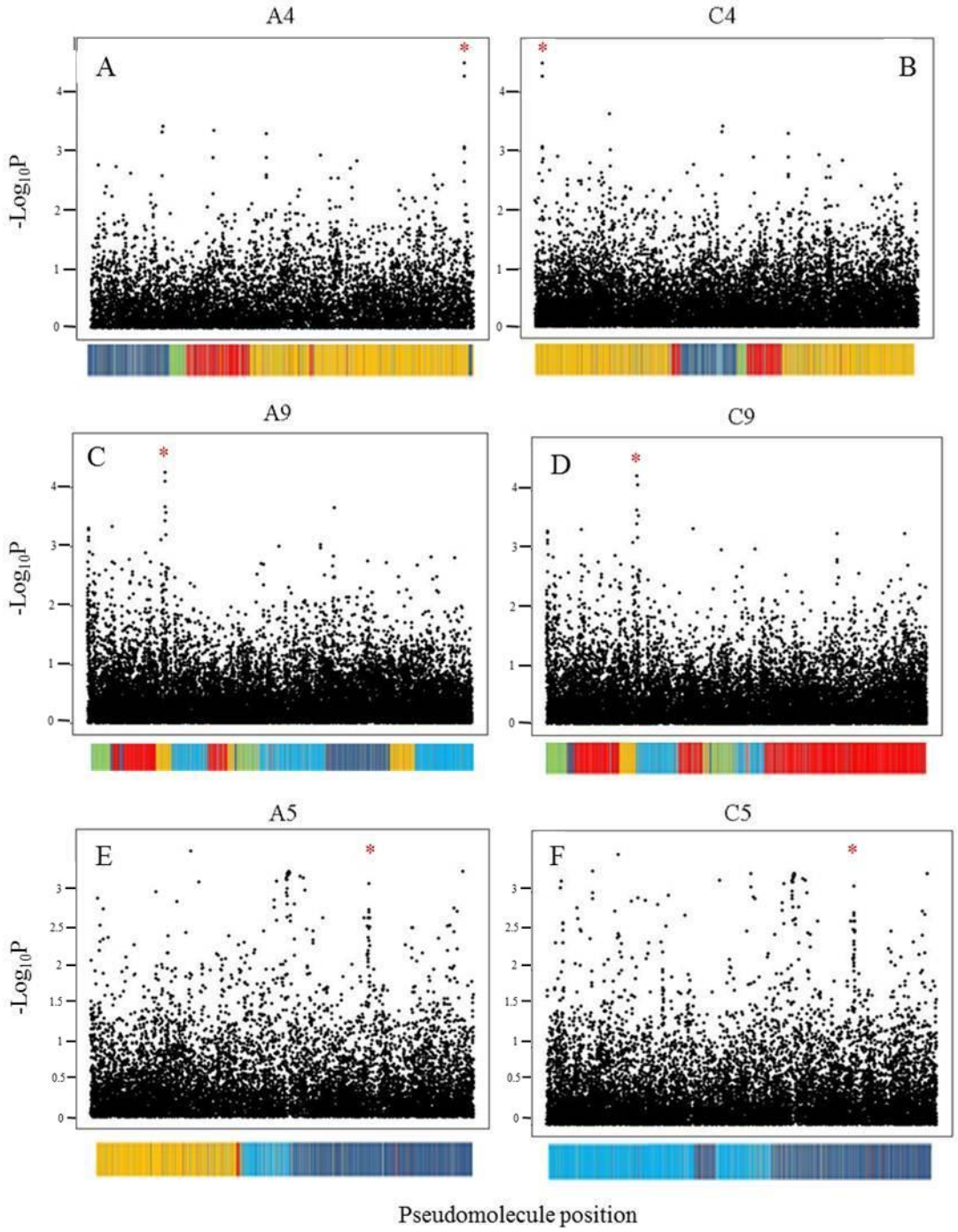

Figure 3. 31. Results obtained for stem weight in the Associative Transcriptomics SNP analysis. Two very clear associations were detected, one of chromosome A4/C4 (A and B) and the other on chromosome A9/C9 (C and D).In addition, an association peak was also detected on chromosome A/C5 for this trait in the KWS 2011 analysis. SNP associations can be seen marked with a red asterisk. Marker P values are plotted at $-Log_{10}P$ in pseudomolecule position.

3.2.2.9 Stem outer cortex thickness

A single SNP association was detected for stem outer cortex thickness. This association is on C6 (Figure 3.32). The homoeologous region for this locus is located on chromosome A2. However, this region does not show a clear association signal due to noise in the surrounding SNP data. The most significant marker for the C6 association is JCVI_335:451 (P=2.70E-04, trait effect 24.2 %). No candidate genes were identified in this region.

In addition to this SNP association highly associating GEM markers were also detected for this trait (Table 3.4b). An example of this is on chromosome C1where two significant markers, JCVI_26302 and JCVI_24162 were detected. The most significant of these markers is JCVI_26302 (P=2.97E-05). Transcript abundance (measured as RPKM) for this marker correlate negatively with stem outer cortex (R^2=0.20). Both of these GEMs correspond to an orthologue of Arabidopsis *BETA GALACTOSIDASE 1* (AT3G13750). Beta-galactosidases may be important in secondary cell wall thickening (Roach et al., 2011). Secondary cell wall thickening is an important characteristic of the outer cortex cells (Zhong et al., 2008) making an orthologue of this gene a good candidate for this trait.

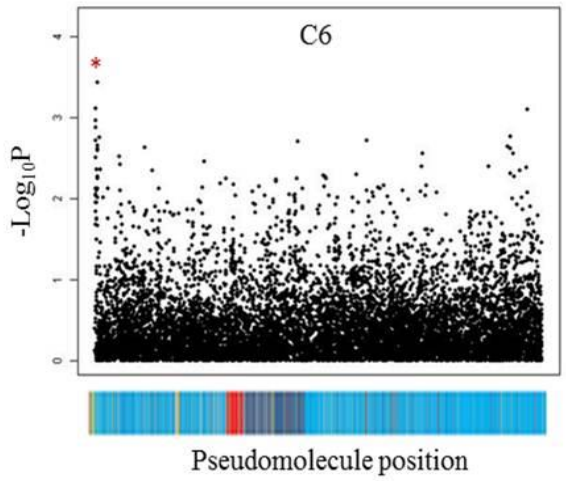

Figure 3. 32. A SNP association detected on chromosome C6 (marked with a red asterisk) for stem outer cortex thickness. Marker P values are plotted at – Log_{10}P in pseudomolecule position.

3.2.2.10 Stem parenchyma area

For stem parenchyma area, a SNP association peak was detected on chromosome A1/C1. The most significant marker was JCVI_36764:445 (P=2.30E-04, trait effect 21.2 %) (Figure 3.33A and B). No potential candidate genes were identified in this region. A second SNP association was detected on chromosome A2/C2 (Figure 3.33C and D). The most significant marker is JCVI_31359:1651 (P=1.15E-04, trait effect 24.8 %). This region was also detected Fmax, F/V and stem diameter. Finally, a very clear association was detected for this trait in the year KWS 2011 analysis on chromosome A3/C3 (Figure 3.33E and F). The most significant marker within this region was JCVI_29829:716 (P=6.01E-04). This region was also detected, although to a slightly weaker extent, in the transformed data for this trait and was detected (again, to a weaker extent) for stem diameter and stem absolute strength in the 2011 KWS analysis) No candidate genes were identified for these marker associations.

Unfortunately due to high levels of noise in the GEM data, it was not possible to confidently identify GEM marker associations for this trait.

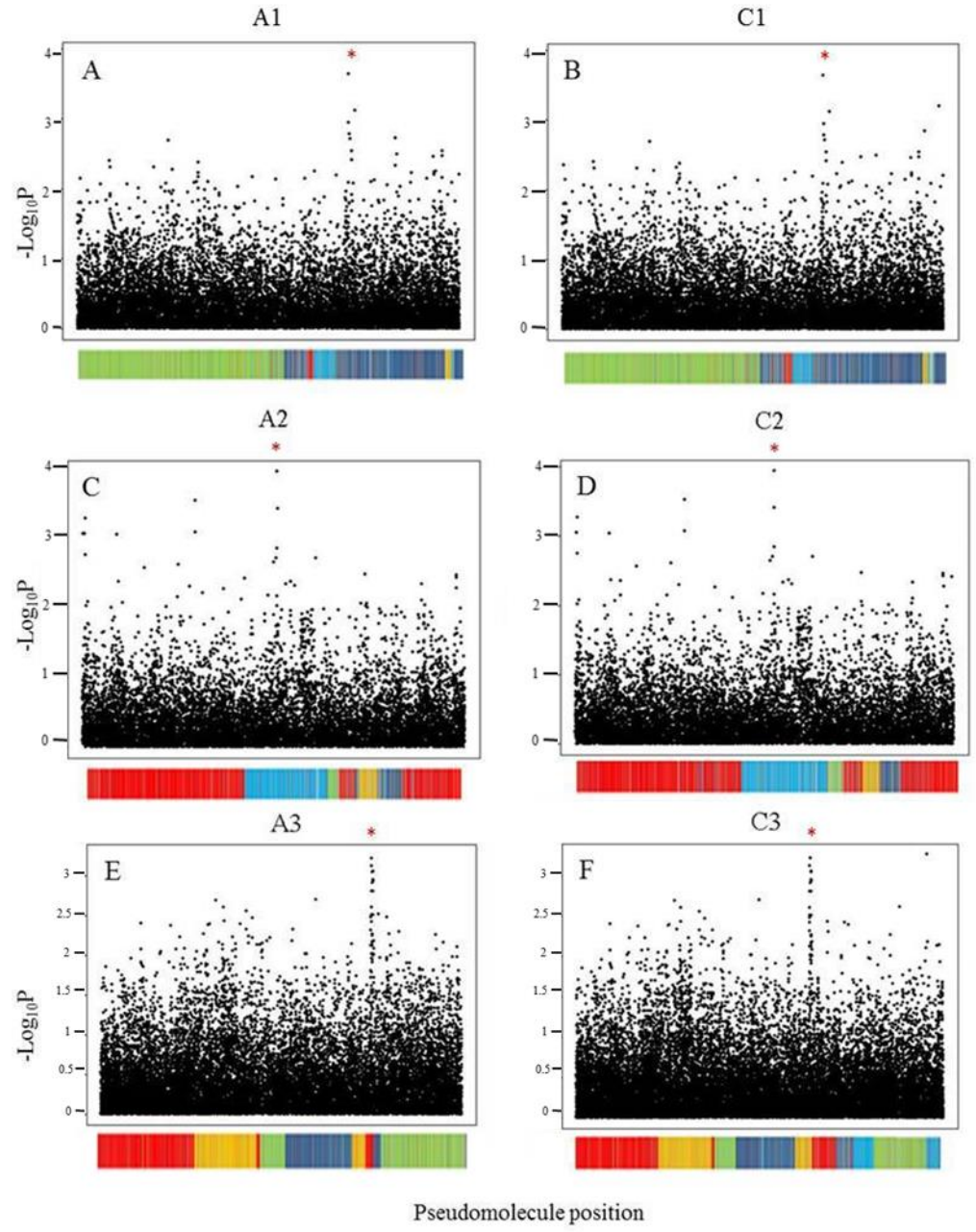

Figure 3. 33. SNP associations detected for stem parenchyma area in the results obtained from the JIC trial. The first of these SNP associations was on chromosome A1/C1 (A and B). The second SNP association, on chromosome A2/C2 (B and C), was previously described for stem absolute strength and stem diameter. A final SNP association was detected in the KWS 2011 analysis on chromosome A3/C3 (E and F). SNP associations can be seen marked with a red asterisk. Marker P values are plotted at $-Log_{10}P$ in pseudomolecule position.

3.2.3 Discussion

Through Associative Transcriptomics SNP associations for stem mechanical strength and related traits have been identified in *B. napus*. In the majority of cases, these associations were identified in the results obtained from the more comprehensive, JIC field trial. This illustrates the increased power achieved by including a greater number of accessions in an association study. In some instances, common associations were seen across different field trials for the same, or highly related traits. For example, on chromosome A2/C2 (where the most highly associating unigene was JCVI_31359), a SNP association for absolute strength was obtained from both the JIC and KWS field trial data. This suggests that these markers are durable across very different environments and may therefore be good targets for the selection of stem mechanical strength in *B. napus*. In many cases, associations were unique to a single field trial. There are a number of potential explanations for this. It is possible that these loci are more environmentally sensitive. This would not be surprising given the large interaction observed between years for the KWS-based field trials (see Section 3.1.2.1). However, given that the majority of associations were detected in the results obtained from the JIC trial, these unique associations may be indicative of the much greater level of variation within this panel in comparison to the much smaller panels grown at the two KWS sites. To assess this further, replication of the larger panel in an additional year/environment would be required.

For stem absolute strength and stem structural strength traits, genes involved in pectin methylesterification were identified. The pectin methylesterification state of the cell wall is thought to be important in cell wall extensibility during cell growth (Wolf et al., 2009). It is therefore possible that variation in stem thickness, which as previously described, explains a high level of variation in absolute strength in *B. napus*, is at least in part due to variation in cell size resulting from alterations in the pectin methylesterification state of the cell wall. The pectin methylesterification state of the plant cell wall has also been found to alter saccharification efficiency of lignocellulosic biomass. There is evidence to suggest that, in Arabidopsis for example, an increase in pectin methylesterification can improve the efficiency of glucose recovery (Francocci et al., 2013). Given this, it may be valuable to assess the saccharification efficiency of biomass taken from *B. napus* accessions showing

variation at marker loci in close linkage with these genes expected to influence cell wall pectin methylesterification state.

In addition genes regulating plant hormones were detected for the absolute strength, structural and morphological traits, including auxin-response genes and brassinosteroid-related genes. These plant hormones are known to be involved in many biological processes in plants (Vanneste and Friml; Sun et al., 2010). Given this, it would be important (following functional validation of these candidates) to assess any pleiotropic effects following the selection of these mechanical strength and related traits.

For stem material strength, genes involved in cell wall biosynthesis were identified as potential candidate genes. For example, genes related to mannan biosynthesis and a NAC transcription factor, *VND2*, which is thought to have a role in xylem development were found in close proximity to associating markers.

In many cases, marker associations were found to be unique to a single trait. For example, on chromosome A8/C3, a SNP association was detected for the absolute strength trait Fmax. This association was not detected for any of the material or structural strength traits (the two components proposed to make up absolute strength). This suggests that the detected association is the result of a combination of factors, possibly both structural and material strength. If this is the case, it may be that the individual contributions of material and structural strength are not sufficient when considered alone to allow for the detection of this association. When looking at the segregation of the alleles for the most significant marker in this association peak, JCVI_639:522, it was noticed that, on average, those individuals carrying the increasing allele for Fmax exhibited higher material and structural strength (second moment of area) in comparison to those carrying the decreasing allele. This observation agrees with the aforementioned hypothesis. The hypothesis is further supported when considering the candidate gene for this locus. This candidate is an orthologue of Arabidopsis *INTERFASCISCULAR FIBERLESS 1*. Arabidopsis mutants for this gene lack interfascicular fiber cell differentiation. These cells, when correctly formed, are highly lignified, which may contribute to stem material strength. In addition, these mutants are known to exhibit reduced cambial activity (Sanchez et al., 2012). An increase in plant stem diameter is promoted by the

differentiation of cells from the vascular cambium. These meristematic cells give rise to the specialised cells of both the phloem and the xylem, the development of which increases the girth of the plant stem (Sanchez et al., 2012). Given this, it is possible that this gene contributes to both material and structural strength.

For the material strength traits, very clear SNP associations were identified. These associations were not seen for any of the absolute or structural strength traits when mapped against the SNP data. When looking more closely at the raw SNP data, the majority of accessions were found to carry decreasing alleles for the associating loci for material strength (see Table 3.3a), with only a small number of accessions (4-6) carrying the increasing allele. However, these relatively rare increasing alleles co-segregate strongly with the few very high material strength values and may therefore be of key biological importance. Although these loci may contribute to absolute strength, given the low level of variation detected in comparison to those detected for absolute strength, it is not surprising that these association are not seen for Fmax and F/V where the associations (most probably relating to structural strength) for these absolute strength traits a) have a greater overall effect on stem strength and b) show a greater level of allelic variation and therefore show stronger segregation patterns. As described in Section 3.1.2.2, in general, the contributions to absolute strength made by material properties were seen to be modest. The lack of variation detected at these loci may help to explain this. In Section 3.1.3, the possible reasons for the weak, inverse relationship observed between stem absolute and stem material strength were discussed. One potential explanation for this relationship is that there may be genes which regulate material and structural strength antagonistically. If this were the case, we may expect to see common associations between the absolute/structural and material strength traits, where the increasing allele for absolute strength/structural strength acts as the decreasing allele for material strength, and vice versa. However, if this were the case, a stronger negative relationship would be expected between these traits, and therefore the relationship observed may simply be indicative of the rare contribution of material strength (due to low allelic diversity), rather than a negative effect on overall strength. However, there is one example where there may be evidence of conflict in the genetic control of these traits. When mapping the material strength traits MOR and MOE, a very clear SNP association was detected on chromosome A6/C6 (where JCVI_10685:430

was the most significant marker). This region was also detected when mapping the transcript abundance levels of a GEM (C_JCVI_28286, corresponding to a CYTOCHROME P450 gene) association detected for Fmax against the SNP data. The transcript abundance levels for this GEM correlate negatively with Fmax. This finding suggests that absolute strength and material strength are related genetically. To explore this relationship further, both SNP association peaks were analysed more closely to identify any markers common to both association peaks. One such marker is JCVI_38015:1581. When looking at the allele effects for these loci, it was seen that for both marker associations the increasing allele is "C" and the decreasing allele is "Y" (the latter being the ambiguity code indicative of a C/T difference in base call between homoeologues). The expression of this GEM correlates negatively with absolute strength, suggesting that an increase in MOR at this locus is coupled with a decrease in Fmax (resulting from an increase in C_JCVI_28286 expression). Therefore there may be antagonism between the genetic control of material and absolute strength. This CYP450 gene may be involved in cell wall biogenesis which may have inverse effects on material and absolute strength in *B. napus*. However, given that this GEM association was not detected for any of the stem structural traits, it is difficult to conclude in what way this gene contributes to absolute strength. It may be that, as for the SNP association peak on A8/C3, the A6/C6 association corresponds to a combined effect of structural strength and material strength. This finding does however suggest that there is a complex relationship between these different stem strength components in *B. napus*.

One particularly interesting aspect of the Associative Transcriptomics analysis described, is the detection of potential interactions between unlinked loci, following the mapping of transcript abundance levels of associating GEMs as traits against the SNP data. For example, in a number of cases, CYTOCHROME P450 genes (detected for absolute and material strength) were associated to loci which were in close proximity genomic regions expected to contain genes involved in cell wall biosynthesis. The additional loci detected in this way may also be good targets for the genetic improvement of these traits.

Through Associative Transcriptomics very promising associations have been detected for stem mechanical strength in *B. napus*. These markers may prove useful in breeding for *B. napus* accessions with improved (or indeed maintained) stem

mechanical strength through Marker Assisted Selection. However, it is important to assess the durability of markers detected through genetic mapping. This is particularly important for the present study, where in a number of cases, marker associations were detected in just a single year/environment which may suggest environmental sensitivity. The following section describes an experiment to assess the durability of a subset of the detected markers in selecting for mechanical strength. This will provide important information regarding the value of these genetic markers in crop improvement.

3.3 *B. NAPUS* MARKER VALIDATION STUDY

The discovery of marker-trait associations through methods such as Associative Transcriptomics has the potential to contribute greatly to crop improvement through marker-assisted breeding. The durability of such markers can however depend on many factors including environmental interactions, where markers durability may vary depending on the environmental conditions. It is also possible that the associations may be the result of false-positive errors due for example to any unaccounted for population structure or relatedness between individuals used within the study. It is therefore important that the efficacy of detected markers in selecting for the trait of interest is tested. This is of particular importance in the present study, due to the lack of across year replication of the more comprehensive, JIC trial. To achieve this, a panel of *B. napus* accessions of previously unknown genotype were screened for variation at a subset of the marker loci detected through Associative Transcriptomics. Following mechanical testing, it was then possible to assess whether the allelic variation at these marker loci, segregate with the target trait as would be expected based on the Associative Transcriptomics results.

3.3.1 Methods

3.3.1.1 Plant material and harvest

96 accessions of *B. napus* (genotype unknown) were harvested from the ASSYST panel grown at Nottingham University. All plants were grown in individual pots within a Polytunnel. For each accession, where possible, three plants were harvested. However, in some cases, only one or two plant replicates were available. All plants were harvested as described in Section 3.1.1. However, for these plants, only the

lower part of the stem was taken and the upper, seed-carrying parts of the stem left to allow for further research being carried out at Nottingham University.

3.3.1.2 Collection of ASSYST panel leaf material and DNA extraction

Prior to harvesting (whilst plants were still growing), leaf material was collected from the 96 accessions taken from the ASSYST panel. Leaf material was collected from the youngest leaf available using forceps (washed in 70% ethanol between samples). Samples were placed directly into capped Eppendorf tubes within a 96-well rack (Qiagen Ltd, Hilden, Germany.), and stored on dry ice. All samples collected were recorded on a 96 well grid sheet to ensure that each leaf sample could be traced back to the correct genotype. This sheet was placed within an envelope and held by a work colleague to allow for the accessions to remain anonymous. In all further analyses, these accessions were identified by their grid position rather than the unique identifier issued during harvest. This was important in removing bias for the analysis.

The samples were submitted to the DNA extraction facility at the JIC. DNA extraction was carried out according to the protocol described in Section 2.4.1.

3.3.1.3 Primer design for screening marker variation detected through Associative Transcriptomics in *B. napus*

Primers were designed for a subset of strongly associated markers identified through Associative Transcriptomics for the stem absolute strength trait Fmax. The Arabidopsis AGI codes were obtained for each marker of interest (taken from the V4 *Brassica napus* pseudomolecule spreadsheet (see Supplementary Data file 3.2f)) and both the full genomic and the coding sequence for each of the Arabidopsis gene models downloaded from TAIR (The Arabidopsis Information Resource: https://www.arabidopsis.org/index.jsp). To obtain the relevant sequence in *B. napus*, the genomic sequence obtained from TAIR was analysed using V4 *B. napus* pseudomolecule BLAST tool, using the n61225 database available at: http://n61225.nbi.ac.uk/blast.html (a JIC internal database). This allowed identification of the pseudomolecule position of the markers of interest and also any paralogous regions to be identified. Using the BLAST output, the homoeologues of the relevant sequences within the pseudomolecule were obtained. This information was then used to extract the corresponding *B. napus* sequence using the “extract

segment" tool which can be found at: http://n61225.nbi.bbsrc.ac.uk/napus.html. The retrieved homoeologues and paralogous sequences corresponding to the target gene of interest were aligned using AlignX (Invitrogen Life Technologies, Paisley, UK). In addition, both genomic and cDNA sequences for Arabidopsis were added to the alignment. This allowed identification of intron:exon boundaries which should be avoided when designing primers based on transcriptome sequence. Finally, transcriptome sequence was obtained for two *B. napus* accessions from the Associative Transcriptomics panel which, based on the transcriptome data used, were expected to show allelic variation for the marker of interest. This sequence was retrieved from the "retrieve consensus" tool which can be found at: http://n61225.nbi.bbsrc.ac.uk/osr_diversity.html. These sequences were added to the alignment, to detect the marker of interest. All NBI web tools cited were provided by Martin Trick, JIC.

NetPrimer (Premier Biosoft International, Palo Alto, USA) was used to design primers. Owing to the high level of conservation between homoeologues (on average differing in sequence by only 3.5 % (Trick et al., 2009)) in *B. napus*, it was not possible to target the expected variation to a specific genome. It was however possible to design primers which would select against the amplification of paralogous regions. In cases where highly similar paralogous sequences were detected, an induced mismatch primer design method was used. This was achieved by designing primers in regions which, although conserved between the target homoeologues, varied between the target homoeologues and the non-target paralogous sequences. In most cases, this variation was seen in the form of a SNP. This site of variation was used as the final base (if in SNP form) of the primer sequence. To further enhance the selective ability of these primers, a single mismatch was introduced at the penultimate base-pair position. The presence of two mismatched bases in the primer sequence is effective in preventing primer binding and PCR amplification of non-target paralogous sequences, allowing for preferential amplification of the target homoeologues (Huang and Brule-Babel, 2010). This approach is illustrated in Figure 3.34. In this example, the target variation resides on chromosome A8/C3 and the induced mismatch (marked with a white "x") is used to select against amplification of paralogous sequences on chromosome A1/C1.This method was employed for all the marker assays included in this study. All primers

were ordered online from Sigma Genosys (http://www.sigmaaldrich.com/life-science/custom-oligos.html).

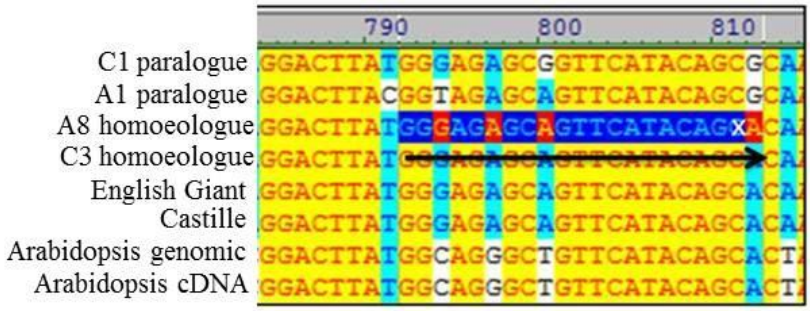

Figure 3. 34. Approach used in selecting against the amplification of paralogous sequences in PCR. The highlighted region marks the chosen primer sequence. The site of induced mismatch within this primer sequence can be seen marked with an "x". This allowed for preferential amplification of (in this example) A8/C3 homoeologues and selection against A1/C1 paralogues.

3.3.1.4 Testing marker assays and genotyping ASSYST accessions

All primer combinations (see Table 3.5) were initially tested using gradient PCR. Standard PCR was then carried out on six *B. napus* accessions of known genotype (included in the Associative Transcriptomics analysis) to assess the efficacy of the marker assays in screening the targeted marker variation. The genotypes of these accessions were confirmed through sequencing. The same marker assays were then used to screen the 96 ASSYST accessions of unknown genotype. PCR and sequencing was carried out as described in Section 2.4.

3.3.1.5 Mechanical testing and exploring marker-trait segregation patterns in ASSYST *B. napus* accessions

All plants harvested from the ASSYST panel were mechanically tested and processed as described in Section 3.1.1.4. The relationship between the uncovered marker genotypes and the mechanical strength data was then explored using ANOVA. This approach tested the following null hypothesis: There is no difference in trait value between accessions carrying the increasing and decreasing allele for marker *x*. Rejection of this hypothesis suggests that the markers identified through Associative Transcriptomics have trait selection capability.

3.3.2 Results

3.3.2.1 Confirming expected allelic variation at marker loci and screening ASSYST accessions for marker variation

To test the power of Associative Transcriptomics to detect markers for trait selection, a subset of markers, (Table 3.5) detected for Fmax, were included in this marker validation study. For this trait three markers, JCVI_39914:1128, JCVI_639:522 and JCVI_31359:1723, were chosen. The first two of these markers, JCVI_39914:1128 and JCVI_639:522 were the most significant markers within the association peaks on chromosomes A5/C5 and A8/C3 respectively. However, for the association detected on chromosomes A2/C2, due to low sequence read depth of the most significant markers (JCVI_31359:1657 and JCVI_31359:1651), an alternative marker JCVI_31359:1723, was selected. This marker was assigned a P value of 9.32E-05 and a trait effect of 26.5% in the Associative Transcriptomics analysis for Fmax. The SNP associations to which these markers correspond are described in Section 3.3.2.1.

As described in Section 3.3.1.3 primers were designed to select against the amplification of all detected paralogous sequences for each target sequence. An example of this can be seen in Figure 3.34 for JCVI_639:522. Through BLAST analysis, in addition to the expected homoeologues, found on A8 and C3, two paralogous sequences were also identified (on chromosomes A1 and C1). By designing the forward primer for this marker in a region found to vary between the target sequence and the paralogous sequences, and through introducing a mismatch at the penultimate base of the primer sequence (marked with an "x" in Figure 3.34), it was possible to achieve preferential amplification of the target homoeologues.

As described in Section 3.3.1.4, all marker assays were initially tested on *B. napus* accessions of known genotype (included previously in the Associative Transcriptomics analysis). This not only allowed for the efficacy of the marker assays in screening the target variation to be tested, but also for the expected variation, initially detected at the transcriptome level, to be confirmed in genomic sequence. For each of the markers included in this predictive study, the target variation was identified. An example of this is given in Figure 3.35, where a comparison can is made between the expected variation based on the transcriptome sequence and the uncovered genomic sequence variation detected through PCR and

sequencing for the Fmax marker, JCVI_31359:1723. Based on the transcriptome sequence a hemi SNP (variation between accessions residing in just a single homoeologue) was expected to be seen between the two *B. napus* accessions "Bolko" and "Hanna" (Figure 3.35A). This was confirmed at the genomic level. Figure 3.35B shows example trace files of the target region. As expected, a double (or hemi) peak can be seen for Bolko. This is indicative of variation between homoeologues. For Hanna, the trace file shows just an "A".

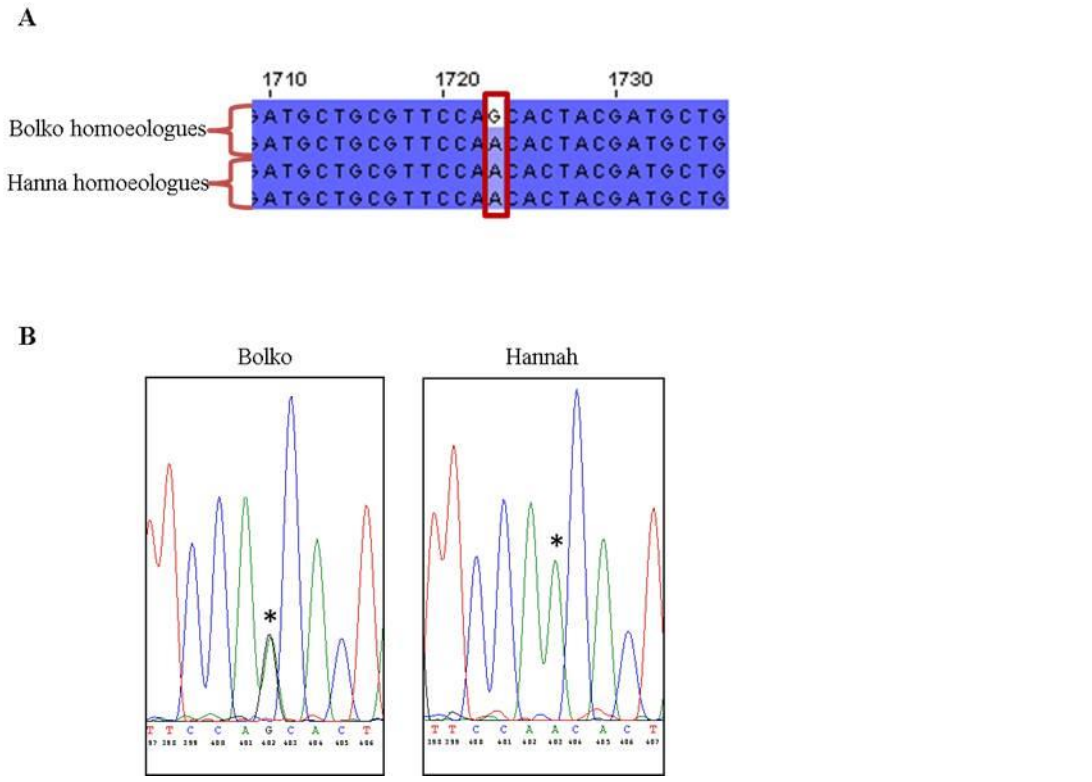

Figure 3. 35. Marker assays were tested through PCR and sequencing on *B. napus* accessions of known genotype. Here, based on the transcriptome sequence (A) the two accessions Bolko and Hannah were expected to have a G/A and an A/A respectively for JCVI_31359:1723. Through PCR and sequencing, the marker assay designed for this locus was found to work effectively and confirm the expected variation at the genomic level (B).

Table 3.5 summarises the primer assays effective in amplifying the targeted marker variation. All effective marker assays were then used to uncover variation in the unknown genotypes taken from the ASSYST panel. Table 3.6 gives an overview of the marker variation detected in these accessions.

Table 3. 5. Primer assays used in screening ASSYST accessions for marker variation detected through Associative Transcriptomics for Fmax.

Marker ID	Linkage group	Increasing allele	Decreasing allele	Forward primer	Reverse primer	Paralogs selected against	Fragment size	Tm°
JCVI_31359:1723	A2/C2	A/G	A	GAAGCTGAAGG ATGCAATGAC	GTTCATCTTCT AAACAG	A7/C7	460	53.1
JCVI_639:522	A8/C3	A	A/C	GGGAGAGCAGT TCATACAGAA	TGAAACGGAAT GTATTCCAC	A1/C1	310	55
JCVI_39914:1128	A5/C5	C	C/G	CAGAAGAGGCT GAAAGGCG	GCCTTTCATCAG TTTCTGCA	A1/C1	360	53.5

Table 3. 6. Marker variation identified across ASSYST accessions for three marker loci detected through Associative Transcriptomics for Fmax.

Marker ID	Number of increasing alleles identified	Number of decreasing alleles identified	Number of sequencing failures
JCVI_31359:1723	27	64	5
JCVI_639:522	62	9	25
JCVI_39914:1128	48	48	0

In all cases allelic variation at the marker loci was successfully identified. A high level of variation was detected for JCVI_39914:1128 and JCVI_31359:1723. The third marker for this trait however, JCVI_639:522 showed a much lower level of variation with just 9 accessions carrying the decreasing allele.

For JCVI_31359:1723 and JCVI_639:522 a number of sequencing reactions failed (particularly for JCVI_639:522). Due to time constraints these reactions were not repeated. Genotype information for each of the 96 ASSYST accessions can be found in Supplementary Data file 3.3a.

3.3.2.2 Exploring variation in stem mechanical strength across marker loci in *B. napus* ASSYST accessions

Through Associative Transcriptomics, JCVI_639:522 and JCVI_31359:1723 were found to have the greatest trait effect sizes of 30.94 % and 33.6 % respectively. JCVI_39914:1128 showed a slightly lower trait effect of 26.89 %. It would therefore be expected that this would be recapitulated in the results obtained from the ASSYST trial. Furthermore, in each case, it would be expected that accessions carrying the increasing allele for the selected markers, would exhibit higher mean Fmax values in comparison to those carrying the decreasing alleles for each marker.

When phenotyping these accessions for mechanical strength, it was noticed that a number of accessions were missing from the panel. It is likely that these were lost during transit. For this reason, together with the lack of information resulting from sequencing failure, the following analyses did not include all 96 accessions for which genotype data were available.

Through ANOVA, significant differences in Fmax were observed between the ASSYST accessions ($P<0.001$) and normally distributed residuals were observed. A summary of the uncovered trait variation, including the obtained ANOVA output can be found in Supplementary Data file 3.3b.

ANOVA was further used to assess variation in Fmax between accessions carrying the increasing and decreasing alleles for each marker individually (see Supplementary Data file 3.3c for trait data and ANOVA output for each marker locus). For JCVI_31359:1723, it was possible to assess marker durability across 83 accessions. Figure 3.36 summarises the trait values for accessions carrying the

alternate alleles for this marker. ASSYST accessions carrying the increasing allele show a mean Fmax of 8.29 kg/sec. Those carrying the decreasing allele exhibited a mean Fmax of 6.436 kg/sec. Through ANOVA, this difference was found to be significant (P=0.017), suggesting that this marker has good trait selection capabilities for this absolute strength trait.

For the second marker for Fmax, JCVI_639:522, 61 accessions were included in the final analysis. Although a higher mean trait value was seen in accessions carrying the increasing allele (7.142 kg/sec) in comparison to those carrying the decreasing allele (4.75 kg/sec), this difference was non-significant (P=0.104). This suggests that this marker does not have good trait selection capability for Fmax. However there were very few accessions carrying the decreasing allele for this marker which is likely to have limited the power to detect a significant difference here.

The final marker for Fmax, JCVI_39914:1128, 87 accessions were included in the analysis. No significant difference in Fmax was observed between accessions carrying the alternate alleles (P=0.062). It is however interesting to note that those carrying the decreasing allele for this marker, on average exhibit a higher trait value than those carrying the increasing allele.

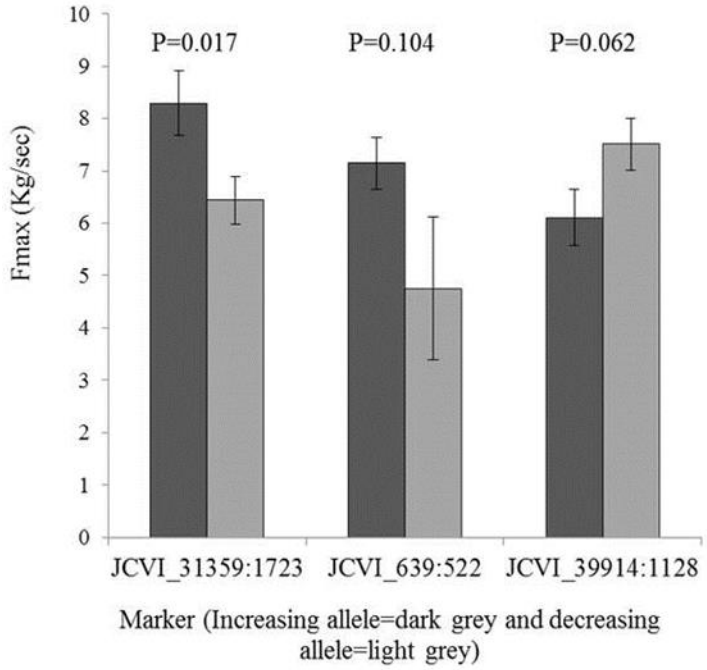

Figure 3. 36. Variation in Fmax seen at three marker loci detected between accessions grown as part of the ASSYST panel. Error bars represent the standard error of the means calculated through ANOVA. P values presented were also calculated using ANOVA.

3.3.3 Discussion

This section has described an experiment carried out to test the trait selection capability of markers detected through Associative Transcriptomics for the stem absolute strength trait, Fmax, in *B. napus.* The results reveal that one of the markers, JCVI_31359:1723, has good selection capability for this mechanical strength trait. This marker could therefore be utilised in breeding to promote improved (or maintain existing) stem mechanical strength in *B. napus*.

JCVI_639:522 and JCVI_39914:1128, were not found to segregate with Fmax as would be expected based on the Associative Transcriptomics results. There are several potential explanations for these negative results. Firstly, it may suggest that these markers are more sensitive to genotype x environment effects and therefore the efficacy of these markers varies depending on the environmental conditions. Secondly, it may suggest that the genomic PCR products analysed, do not represent the same loci as the mRNA-seq data. This may be the result of PCR amplification

bias for example. It is also possible, that the SNP alleles defined by mRNAseq are artefacts caused by differential expressions i.e. a resolved base being assigned because one homoeologue has been silenced. Although, the primer assays were initially tested on *B. napus* accessions of known genotype, given that this was carried out on only a small selection of accessions, it is possible, by chance, that the expected variation was uncovered. This, although unlikely, is a possible risk when seeking SNP variation based on transcriptome sequence. The lack of predictive power may also suggest that these markers are false positive associations, which may arise due to unaccounted for population structure between accessions, or may be the result of a statistical artefact (this will be discussed further in Chapter 5).

The lack of marker durability may also relate to high trait variance for Fmax in the JIC trial data. To improve the chances of marker associations describing important variation at the population level, it is important that population sample means are of high accuracy. Although increased power was gained from the JIC panel owing to a greater number of accessions, plant replication for this trial was relatively low which would undoubtedly limit the power of accurately summarising variation for this complex trait.

For JCVI_639:522, an increase in Fmax was seen for individuals carrying the increasing allele. However, in the ASSYST panel, very few accessions with the decreasing allele at this locus were identified which may, in part, explain the lack of statistical significance observed here. For the absolute strength marker, JCVI_39914:1128, a higher trait value was observed in accessions carrying the decreasing allele. Based on the Associative Transcriptomics results this would not be expected. The difference in trait values observed between the two alleles for this locus was however found to be non-significant. It is therefore possible that the trend observed is merely a product of random variation. Regardless of the cause for this trend, this marker has been found lack trait selection power for absolute strength and is therefore not a good marker for the selection of this trait.

Overall, this study has illustrated the importance of validating genetic markers in order to assess their true value for crop breeding. However, it also shows that even for very complex traits, such as stem mechanical strength, Associative

Transcriptomics can be used for the identification of genetic markers with true trait selection capability in *B. napus.*

3.4 SCREENING ARABIDOPSIS T-DNA INSERTION LINES FOR ALTERED STEM MECHANICAL PROPERTIES

Through Associative Transcriptomics, it was possible to propose potential candidate genes for stem mechanical strength in *B. napus*. However, the detection of plausible genes in close proximity to associated markers does not prove causality. A common method used to explore the potential role of a given gene in contributing to a trait of interest, is mutagenesis. This allows for the assessment of phenotypes following the alteration or loss-of function of the target gene (May et al., 2002). Although there are methods through which this can be achieved in *B. napus* directly, such as TILLING (Sikora et al., 2011), given the genome complexity and long generation time of this species, such approaches may not in all cases be feasible. A very attractive alternative is to study the role of any genes of interest in the closely related species Arabidopsis, which has a much simpler, diploid genome and short generation time in comparison to *B. napus.* In addition, with the availability of a complete genome sequence, resources are now available to facilitate gene function discovery in this species. An example of this is the Arabidopsis T-DNA insertion line library. The development of such mutants, involves the transformation of Arabidopsis with *Agrobacterium tumefacians* carrying a T-DNA. The incorporation of this sequence into the Arabidopsis genome can result in the disruption of gene function due to the integrated foreign sequence. Although this process is random, it is possible, given the known sequence of the inserted T-DNA, to determine which gene has been disrupted (Krysan et al., 1999). This approach has allowed the development of an Arabidopsis T-DNA library, the seed for which is available to the research community (European Arabidopsis Stock Centre). This section describes the screening of Arabidopsis T-DNA insertion mutants for a subset of the candidate genes proposed for stem mechanical strength in *B. napus* in this present study. This allowed any alteration in mechanical properties within these mutants to be assessed and therefore the plausibility of these candidate genes as the source of the causal variation detected through Associative Transcriptomics to be determined.

3.4.1 Methods

3.4.1.1 Plant material used

T-DNA insertion lines (Table 3.7) were ordered from the National Arabidopsis Stock Centre (NASC) using the catalogue database system which can be found at: http://arabidopsis.info/. Where possible, multiple mutant lines for each candidate gene were obtained. To maximise the chance of identifying an adequate number of homozygous mutants and to allow for increased confidence of plant genotypes, where possible, mutants were assessed across two generations. For some T-DNA lines however, due to the late arrival of seed, this was not possible. These mutants were analysed in just a single generation. All seeds were grown in Scotts Levington F1 compost in single pots (9x9 cm). For each line, all seeds obtained (approximately 30) were sown. Prior to sowing, all pots were thoroughly watered. Seeds were then sprinkled sparingly across the surface of the moistened soil. These seeds were stratified for three days at 5 °C to encourage coordinated germination. Following two weeks of growth, 24 well-established seedlings for each line were pricked out into P40 trays containing Scotts Levington F2 mix (plus 25 % grit) compost treated with Intercept insecticide (for aphid control) (gloves should be worn when handling soil treated with Intercept). All seedlings were grown in a Controlled Environment Room (CER3508, building 35, JIC) (22 °C, 75 % RH and 10hr day). All plants were watered by hand, twice a day. To prevent contact between neighbouring plants all plants were held in an upright position using wooden stakes. Once the plants had reached a stage where browning of the seed pods was visible, they were bagged using Glassine ungummed bags (330X180 mm – Global Polythene). At this point, watering was reduced and then, after a further week, stopped completely. All mature plants were harvested by removing the plant and stake from the soil within the bags. Each bag was labelled with a unique identifier to avoid mixing of samples. All harvested plants were left at room temperature for one week prior to threshing, to ensure complete drying. All plants were threshed by hand using a sieve to remove unwanted plant debris, and stored in small seed bags (Kristal ungummed, 73X41 mm-Global Polythene), again labelled with a unique identifier.

3.4.1.2 Arabidopsis leaf sample collection and DNA extraction

After three weeks of growing in P40 trays, leaf samples were collected from each seedling. One new leaf was removed from each plant using forceps (sterilised between samples with 70 % ethanol) and transferred directly into 96well plate tube racks (Qiagen. Ltd, Hilden, Germany). Each sample collected was recorded on a 96 well grid sheet, to ensure that each DNA sample could be traced back to a specific plant. During sample collection, samples were kept on dry ice at all times. All leaf samples were submitted to DNA extraction facility at The John Innes Centre. The DNA was extracted according to the method described in Section 3.3.1.2.

3.4.1.3 Genotyping Arabidopsis T-DNA lines

Primer design was carried out using the Salk Institute Genomics Laboratory (SIGnal) T-DNA primer design tool: http://signal.salk.edu/tdnaprimers.2.html. Genotyping was carried out using a double reaction PCR approach. Reaction one contained the right border primer (RP) and the left border primer (LP) to amplify product only in the presence of a wild type (WT) copy of the target gene i.e. no insertion. Reaction two included RP and the insertion primer (Lb1.3 for Salk lines and Lb3 for SAIL lines). This would amplify only in the presence of the T-DNA insertion. For homozygous WT individuals a PCR band would be expected in reaction one only. For plants homozygous for the insertion i.e. no WT gene copy, PCR product would be expected in reaction two only. Finally, in the case of heterozygous individuals, a PCR product would be expected in both reaction one and reaction two. Information regarding this method can be found at: http://signal.salk.edu/.

Genotyping was carried out through PCR according to the following protocol:

For a single 10 µl reaction:

1- 0.05 µl Takara Ex taq (250u-Clontech-Takara Bio Europe, Saint-Germain-en-Laye, France)
2- 1 µl 10x Takara buffer
3- 0.8 µl 2 mM DNTPs
4- 0.2 µl Primer 1(10 mM)
5- 0.2 µl Primer 2 (10 mM)
6- 6.75 µl Distilled water
7- 1 µl DNA

Using the above mix for each reaction, PCR was carried out using G-Storm -GS1 thermal cycler PCR machine. The cycles used were as follows:

1- Hot start 94 °C for 5 minutes
2- 40 cycles
3- Touchdown 63 °C-48 °C 15 cycles
4- Start cycle 40 times
5- Denaturation 94 °C for 30 s
6- Annealing 56 °C for 30 s
7- Extension 72 °C for 60 s
8- End cycles
9- 72 °C 60 s
10- Store at 10 °C infinite

3.4.1.4 Plant material and growth of the second generation of Arabidopsis T-DNA lines

For each T-DNA line grown in the first generation, where possible one homozygous and one heterozygous parent plant was selected for analysis in a second generation. This allowed for the expected genotype of these plants to be confirmed and for any differences in mechanical strength between homozygous and heterozygous mutants to be explored. 24 seeds per genotype were sown, stratified and grown in FP9 pots as described in Section 3.4.1.1. However, following two weeks of initial growth, seedlings were transferred into the 51-cell ARRASYSTEM (Beta Tech bvba, Gent, Belgium) trays (http://www.arasystem.com/). For each genotype, 12 plants were transplanted. This not only allowed for a good level of replication in all further analyses, but also provided extra plant material in case of failure to establish following transplantation. In cases where all plants in the first generation were found to be heterozygous, 24 seedlings were transplanted to the ARRASYTEM trays. Following four weeks of growth, all plants were placed within the ARRASYSTEM growth tubes (See Figure 3.37). This system allows plants to grow freely without any mechanical perturbation caused by staking of the plants following stem bolting. All plants were grown in the West 1, JIC glasshouse facilities under (20°C day and 18°C night-16hr day length).

After 35 days of growth, photographs were taken, where possible, of a representative homozygote and, where possible, a heterozygote plant for each line. After 35 days no further stem growth is expected to occur (The Arabidopsis Information Resource (TAIR)). A typical Col-0 plant was also selected to be photographed. All plants were photographed next to a 30 cm ruler to allow for any variation in height to be visualised clearly. All plants where assessed for any changes in growth at day 45 and if necessary photographed a second time.

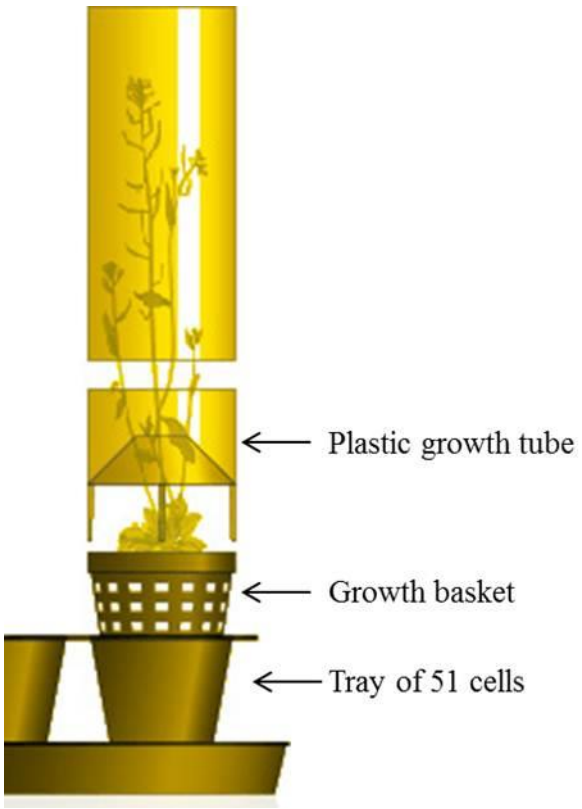

Figure 3. 37. ARRASYSTEM tubes were used to allow the Arabidopsis plants to grow without the need to stake plants for support during growth. This image was adapted from http://www.arasystem.com/.

3.4.1.5 Leaf collection and DNA extraction of 2nd generation T-DNA lines

Leaf material was collected from all plants and stored within 1.5 ml Eppendorf tubes in liquid nitrogen. DNA extraction was then carried out according to the following protocol:

Each leaf sample was macerated using a clean plastic pestle (cleaned in 70 % ethanol) for 15 seconds. 200µl of Edwards extraction buffer (200 mM TrisHCL pH 7.5; 250 mM NaCl; 25 mM EDTA and 0.5 % SDS), was then added to each tube and

the sample further ground, again using a pestle. All samples were then vortexed for 5 seconds before being centrifuged for 3 minutes at 13,000 rpm in the centrifuge. The resulting supernatant was then added to a fresh Eppendorf tube containing 150 µl 100 % isopropanol (previously stored at 4 °C). The tubes were inverted twice to ensure mixing of the two liquids. Samples were then centrifuged for 7 minutes at 13,000 rpm. Supernatant was removed using a pipette and the pellet washed using 70 % ethanol. The samples were then left for 20 minutes in a flow hood to ensure complete drying. Once dry, 30 µl of distilled water was added to each tube and the samples stored overnight at 4 °C to allow for the DNA to fully suspended.

3.4.1.6 Genotyping 2nd generation Arabidopsis T-DNA insertion lines

Genotyping of this second generation of plants was carried out as described in Section 3.4.1.3.

3.4.1.7 Harvesting of Arabidopsis and preparation for mechanical testing

Once the first silique could be seen turning brown, watering of the plants was stopped. All plants were then harvested from the base of the stem and placed within breathable plastic bags. The harvested plants were left at room temperature for 7 days to allow for further drying. For mechanical testing, a 2.5 cm section was taken from the base of the stem using a scalpel. This was carried out, where possible, from six plants per genotype. This section was then stored in a small breathable bag and stored within a silica humidity tank at 55 % RH and 23 °C for two days. Two days had previously been found to be adequate for the equilibration of the moisture content of wheat stems with that of the silica tank. Given the much smaller stems of Arabidopsis in comparison to wheat, it was decided that two days would be more than adequate in this species. In addition to the 2.5 cm section, a further 1 cm section was taken (just above the first section) for determination stem cross-sectional properties (required for later calculation of stem second moment of area.). Mechanical testing was carried out using the same Texture Analyser used in mechanically testing *B. napus* stems. To accommodate the much smaller stem samples, a support stand was made using two razor blades (blade covered and pointing downwards) held within a clamp. These supports were set to be 1.2 cm apart. The Texture Analyser was fitted with a 1 kg load cell. The probe was set at a start height of 4 cm from the test bed and the probe was set to descend a further 0.5

cm following contact with the stem sample. All other parameters were set as described in Section 2.2.1.

3.4.1.7 Obtaining stem cross-sectional measurements in Arabidopsis

The 1cm stem samples taken from the base of the stem were placed vertically into petri dishes (making sure all stems were touching the base of the petri dish to ensure correct scaling of each image) containing agarose gel. This allowed easy visualisation of stem cross-sections under a light microscope. All samples were photographed at the same magnification, with the inclusion of a ruler (to allow for the scaling of each image individually). All stem cross sections were then analysed using Sigma Scan, as described in Section 2.2.2. The results obtained were then used to calculate stem second moment of area for each sample according to the equations described in Section 2.2.2.

3.4.1.8. Data processing and analysis

The stress-strain data (obtained from the three-point bend test) together with the results obtained from the Sigma Scan analysis were processed using an adapted version of the Genstat script described in Section 2.2.2. To allow for the detection of significant differences between the T-DNA insertion lines and WT plants, the student T-test was used. This was carried out using standard errors calculated through ANOVA based on all plants included in the study.

3.4.1.9 FT-IR analysis

With assistance from Ian Wood (IFR) for a subset of mutants, Fourier Transform Infra-red (FT-IR) spectra were collected across the spectral range of 800–4000 cm^{-1} using a dynamic alignment FT-IR spectrophotometer (Bio-Rad FTS 175C, Bio-Rad Laboratories, Cambridge, USA). Prior to analysis of stem tissue, a spectra of air was generated which acted as the reference medium. Each stem sample was macerated using a pestle and mortar and the ground tissue clamped against a diamond element. To reduce error, the crystal was completely covered with stem tissue. Between samples, the crystal and clamp were cleaned with 100 % ethanol. This was left to dry thoroughly before analysing the next sample. For each sample, two spectra were taken, with three plant replicates per mutant line analysed. All spectra obtained were truncated to the fingerprint region (800-1800 cm^{-1}). The baseline was anchored to 1800 cm^{-1} and the curve areas normalised. Average spectra were then obtained for

each mutant line and the values subtracted from those of WT Arabidopsis. This allowed for any difference in cell wall composition of the T-DNA insertion lines relative to WT to be detected.

3.4.2. Results

Trait means and standard errors calculated through ANOVA, and student T-test results can be found in Supplementary file 3.4a.

3.4.2.1. Genotyping Arabidopsis T-DNA insertion lines and assessing mutants for altered stem mechanical performance.

Table 3.7 summarises the T-DNA insertion lines included in this study. This Table also lists the primers used in successfully genotyping these mutants. For each Salk line, the primer used (along with the appropriate right genomic primer (shown in Table 3.7)) to amplify the insertion was: ATTTTGCCGATTTCGGAAC. For all SAIL lines the primer used for the insertion was: TAGCATCTGAATTTCATAACCAATCTCGATACAC.

Table 3. 7. Arabidopsis T-DNA insertion lines screened for altered mechanical strength and the primers used in genotyping them.

Candidate gene	Salk/SAIL ID	T-DNA site	Arabidopsis gene model	Left genomic primer (LP)	Right genomic primer (RP)	Size of WT product (bp)	Size of insertion product (bp)
GAUT5	Salk_050186C	Exon	AT2G30575	GAGGCCAAA ACATTGTCAGAG	GTCATTATCGGGG AAGCTTTC	1068	548-848
GAUT5	SAIL_401_E11	Promoter	AT2G30575	TGTTAGGCAAAA ATGAATCACAAC	GACGGAAGTGATGC TCTTGAG	1116	531-831
PMEI/PME	Salk_085493	Promoter	AT3G12880	CATGCTTAGCCG TTGAGAAAC	TGAAGAAGCGTGGA AGAGAAG	1111	596-896
PMEI/PME	SAIL_602_G12	Promoter	AT3G12880	CATGCTTAGCCG TTGAGAAAC	TTCAGATTTGTACCG TACCGC	1222	536-836
PMEI/PME	SAIL_649_D02	Exon	AT3G12880	TGTTGGTGTATC TGCGATCTG	GAATGTTGAGTGCCT CGAGAC	1066	518-818
VND2	Salk_022124	Intron	AT4G36160	TCTACAGAATCG AACCATGGG	AGTATGCCAAACCTT TAGGCC	1124	548-848
SAUR72	SAIL_564_B07	Promoter	AT3G12830	ATGGACGTGTTT ATGGTTTGC	CAGATTTCTCTGCTG ATTCGG	1050	498-798

3.4.2.1.1. GALACTURONOSYLTRANSFERASE 5 (GAUT5, AT2G30575)

Two T-DNA insertion lines were obtained for the candidate gene, *GAUT5*. This candidate was selected based on the SNP marker association detected on chromosome A4/C4. This association was for stem diameter in the JIC trial with the most significant marker being JCVI_38024:112 (see Section 3.2.2.5). This region was also detected in the analysis of the KWS 2010 material for second moment of area. The first insertion line for this gene was Salk_050186. Analysis of the first generation of this mutant line, revealed only heterozygous individuals (see Figure 3.38 for representative genotypes detected for this line). By analysing the genotype of plants taken from a single heterozygous parent in a second generation, again only heterozygous individuals were observed (Figure 3.39). This may suggest that the homozygous genotype of this insertion is lethal. Given this, it was only possible to analyse heterozygous plants. It is however interesting to note that in no cases were homozygous WT genotypes detected. This may suggest the presence of an additional sequence within the genome which is being amplified (based on sequence similarity this would most likely be *GAUT6* (Sterling et al., 2006)), and therefore masking the presence of any homozygous mutants or WT individuals in the analysis. To assess this possibility it would be necessary to sequence the PCR products obtained. However, due to time constraints and given that an additional insertion line for this candidate gene was available, this anomaly was not explored further.

The second insertion line for *GAUT5* was SAIL_401_E11. Screening of SAIL_401_E11 in the first generation detected both homozygotes and heterozygotes (see Figure 3.38). Analysis of a representative of each genotype in a second generation, confirmed the expected genotypes (Figure 3.40). The fact that it was possible to successfully identify non-heterozygous individuals for this insertion line (which carries a mutation in the same target gene as Salk_050186) may suggest that the lack of homozygous plants detected for Salk_050186 is not the result of amplification of a non-target gene. It is also possible however, that the RP+LP primers used in amplifying this region in SAIL_401_E11, are more specific to the target sequence than that of Salk_050186.

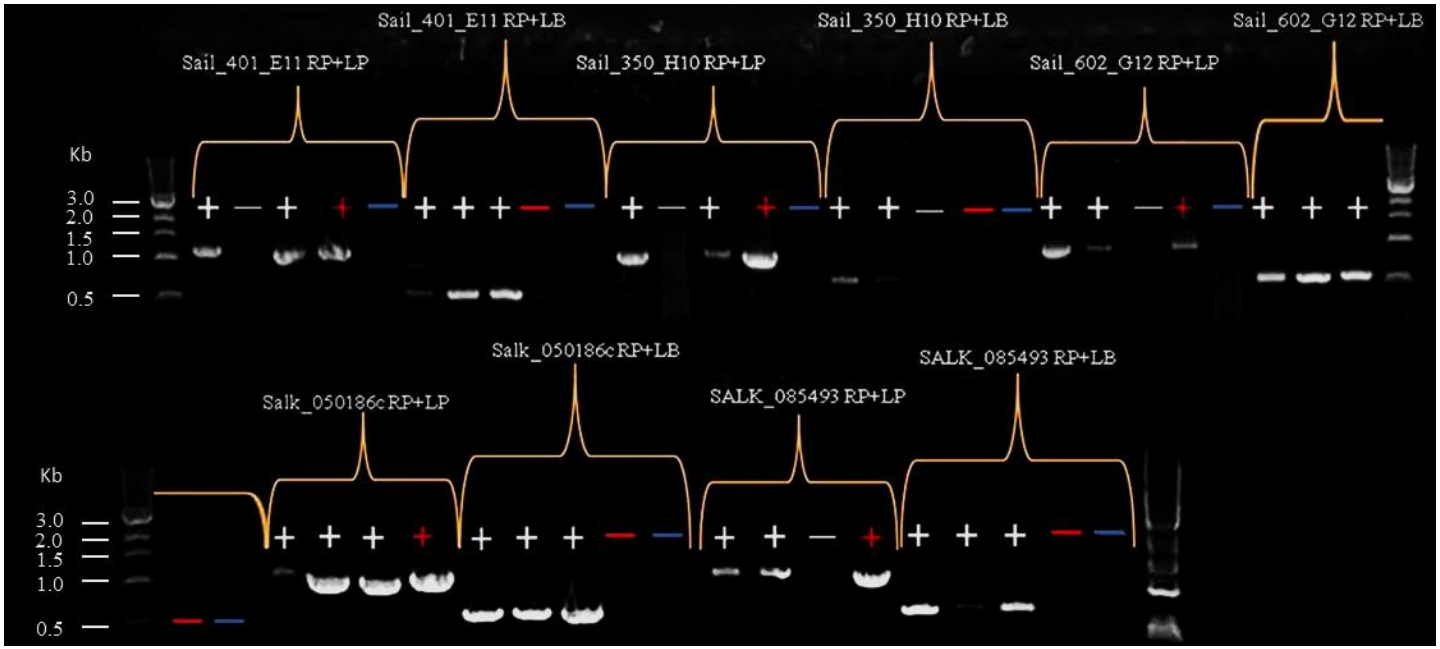

Figure 3. 38. Genotypes uncovered for the T-DNA insertion lines screened in the first of two generations. Each set of a given T-DNA line is represented across two reactions. The first, RP+LP targeted the amplification of any WT gene copies present. The second reaction RP+LB screened for the presence of the T-DNA insertion. "+" and "-"signs indicated the presence and absence of PCR product respectively. The signs are shown in red and blue for PCR reactions carried out with WT DNA and water controls respectively (1Kb ladder used).

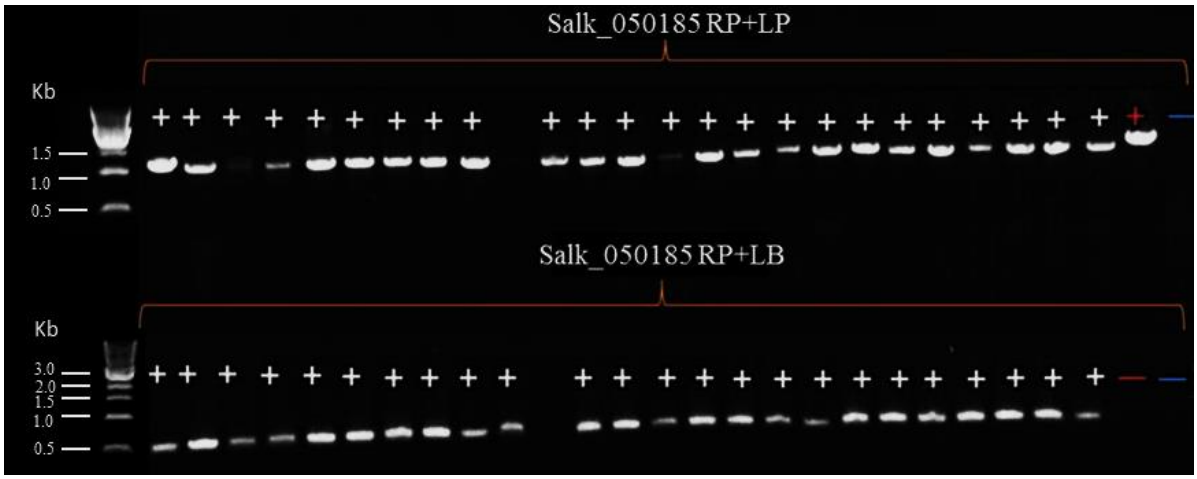

Figure 3. 39. Genotyping of a Salk_050186 plants in a second generation. The seed were taken from a heterozygous parent from the first generation. "+" and "-" signs indicated the presence and absence of PCR product respectively. The signs are shown in red and blue for PCR reactions carried out with WT DNA and water controls respectively (1Kb ladder used).

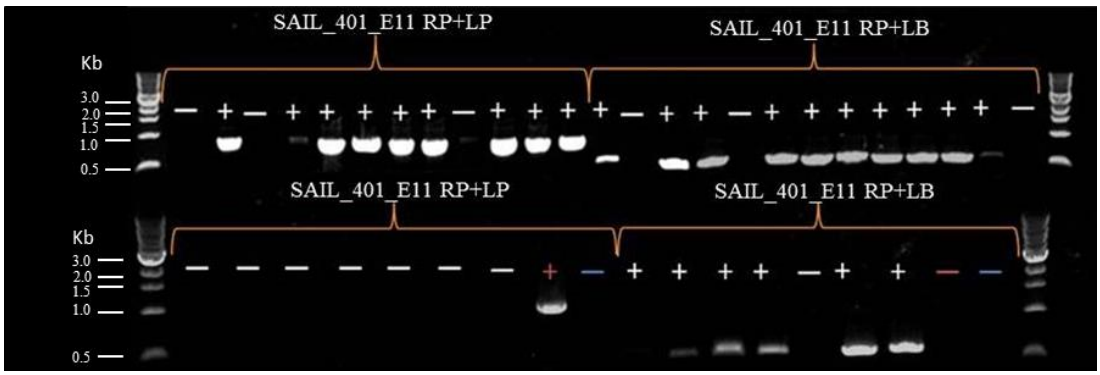

Figure 3. 40. Confirmation of heterozygous (A) and homozygous (B) genotypes selected from the first generation of genotyping the mutant line, SAIL_401_E11. "+" and "–" signs indicated the presence and absence of PCR product respectively. The signs are shown in red and blue for PCR reactions carried out with WT DNA and water controls respectively (1Kb ladder used).

For Salk_050186, although a reduction in stem structural and absolute strength in comparison to Col-0 was observed, these were found to be non-significant (Figure 3.41). For the second T-DNA insertion line, SAIL_401_E11, the homozygous plants

were found to differ significantly from Col-0 for both stem diameter and second moment of area (both at $P<0.05$) (Figure 3.41A and B). The heterozygous plants for this line were also found to vary significantly from WT for these structural traits ($P<0.01$). Both homozygous and heterozygous plants also exhibited significantly lower Fmax and F/V in comparison to WT ($P<0.001$) (Figure 3.41C and D). No significant changes in material strength were observed ($P>0.05$). These results add evidence to the hypothesis that *GAUT5* has a role to play in determining stem structural strength.

In the Associative Transcriptomics analyses, no clear signals were detected for the absolute strength traits Fmax and F/V in the A4/C4 region containing *GAUT5*. However, the results obtained from this T-DNA line screen show a clear effect on absolute strength. The lack of association seen in *B. napus* may be the result of the much more complex genetic makeup of this crop species in comparison to Arabidopsis. The Association signal for this A4/C4 locus was much weaker than those observed for the absolute strength traits. It is therefore possible, that while this locus may have an effect on absolute strength, the effect is concealed by the additional loci of larger effect.

In addition to the results obtained from the three-point bend test, a number of interesting phenotypes were observed in the SAIL_401_E11 mutants during development. After 35 days of growth, heterozygous plants were slightly shorter than Col-0. This difference was found to be statistically significant ($P<0.05$) (Figure 3.42B). Homozygous plants were however much shorter than Col-0 (Figure 3.42C). However, after 45 days of growth, some increase in stem elongation was seen. These mutants still however remained significantly shorter than Col-0 ($P<0.01$). In addition, after 45 days, these homozygous plants were unable to maintain an upright growth habit without support (Figure 3.42D). This is particularly interesting given that these mutants are considerably shorter than WT (and therefore have reduced leverage imposed by stature). Similar phenotypes were also observed for the Salk_050186 mutants.

Very little work has been carried out on *GAUT5*. It is however known to be highly expressed in the plant stem of Arabidopsis (Caffall et al., 2009). In addition, very few galacturonosyltransferase genes have been fully characterised. Work carried out

on mutants lacking a functional copy of *GAUT8* (*qua-1* mutants) however, have been found to exhibit similar phenotypes to those observed in the present study for the mutants defective for *GAUT5*. *Qua-1* mutants are known to exhibit reduced plant height and stem thickness in comparison to WT (Orfila et al., 2005). *QUA-1* has been proposed to be a gene with galacturonosyltransferase activity, involved in the synthesis of the cell wall pectin, homogalacturonan. Orfila et al (2005), identified this gene as a catalyst which promotes the incorporation of galacturonic acid residues onto growing homogalacturonan chains (Orfila et al., 2005). Mutants for this gene have been found to exhibit reduced levels of cell wall galacturonic acid and a 25% reduction in cell wall homogalacturonan content. It has also been proposed that this gene may be involved in determining the pectin methylesterification state of the cell wall (Bouton et al., 2002) which may, at least in part explain the reduced cell expansion also displayed in this mutant. Finally, *qua-1* mutants exhibit a reduction in xylem cell differentiation in comparison to WT Arabidopsis, which may suggest a reduction in cambial activity (Orfila et al., 2005). Similar phenotypes have also been observed in mutants lacking a functional copy of *GAUT6* (Caffell, 2008), which, based on sequence similarity, is the galacturonosyltransferase most closely related to *GAUT5* in Arabidopsis (Sterling et al., 2006). Based on this information, it seems likely that the reduced stem thickness phenotype observed in the *GAUT5* mutants, is the result of reduced cell size in the stem tissue. It may also be in part the result of reduced cell differentiation from the cambial tissue. To confirm this, an in depth microscopy analysis would be required.

The results presented here are very promising and provide further evidence to suggest that *GAUT5* is the candidate gene to which the SNP association detected on chromosome A4/C4 corresponds.

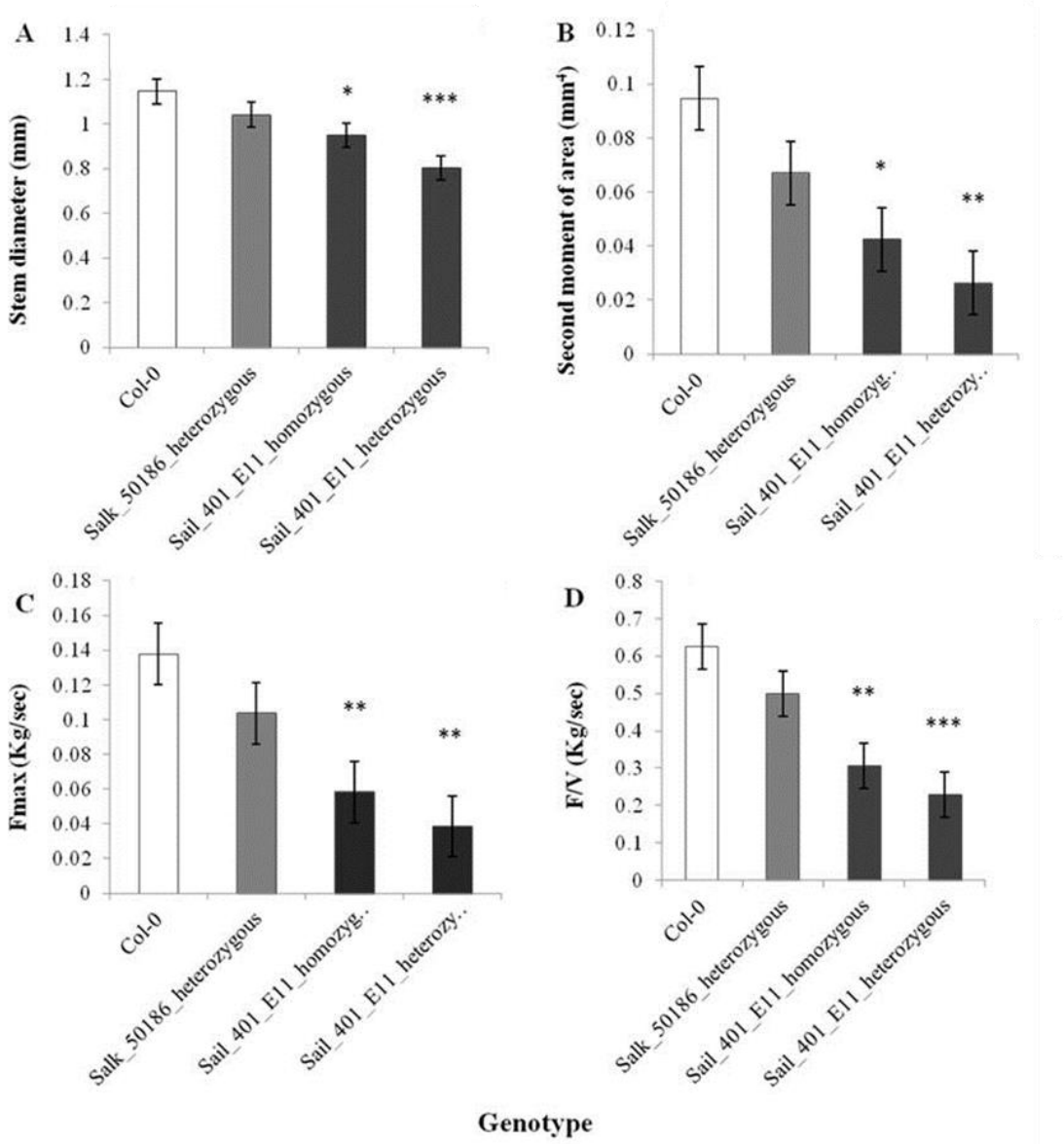

Figure 3. 41. Variation in stem diameter (A), second moment of area (B), Fmax (C) and F/V (D) observed between the GAUT5 T-DNA insertion lines and Col-0. The asterisks are indicative of the significance level of the differences observed in these traits between the different mutant genotypes and Col-0 where *<0.05, **<0.01 and *<00.1 as determined through the student T-test. The error bars represent the standard error of the means calculated through ANOVA.**

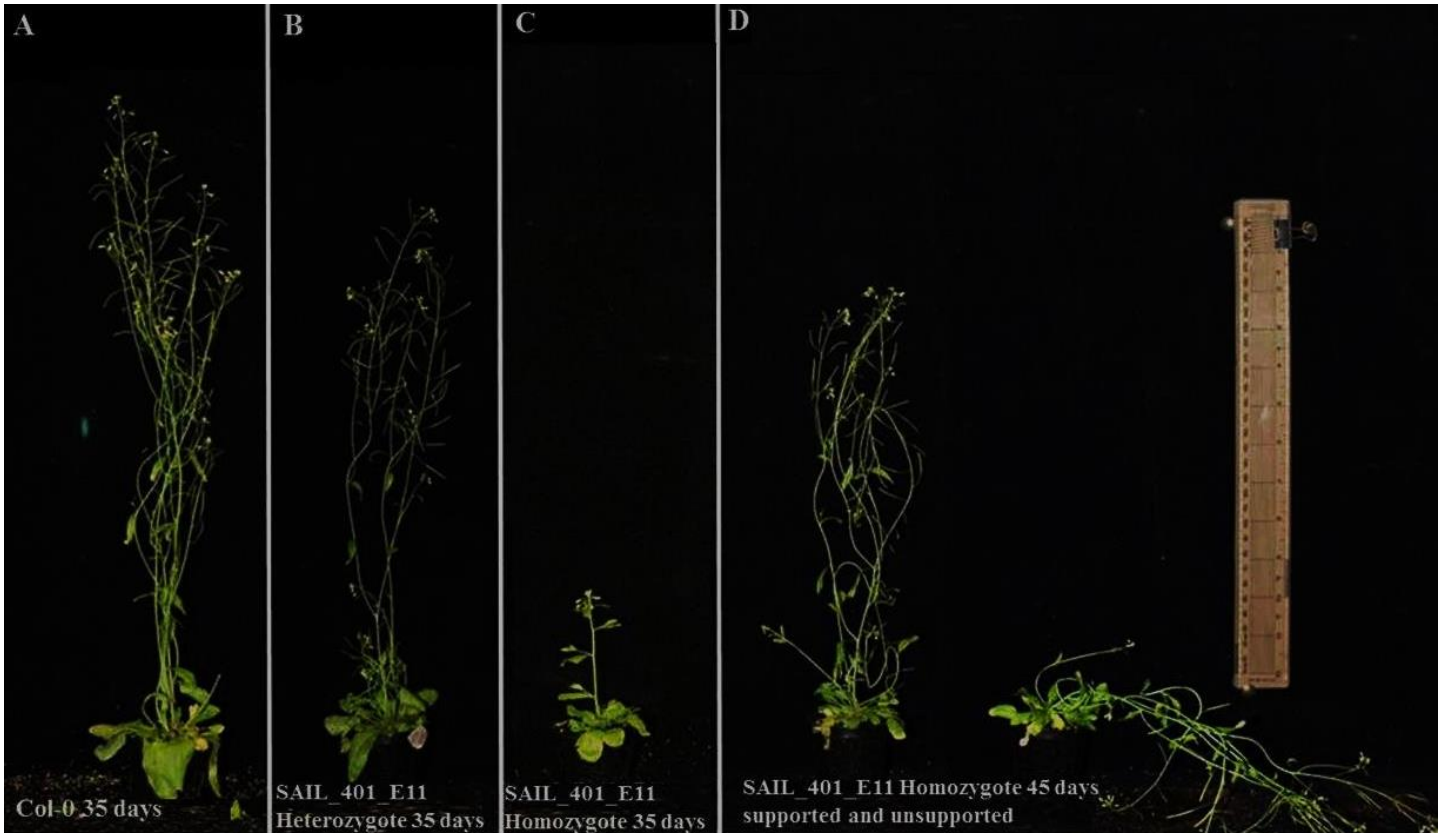

Figure 3. 42. A typical Col-0 plant (A) alongside a representative heterozygous plant for the Galuronosyltransferase-5 T-DNA insertion line at 35 days (B) and a representative homozygote for this SAIL_401_E11 at 35 (C) and 45 day (D).

3.4.2.1.2 Pectin methylesterase (PME)/pectin methylesterase inhibitor (PMEI) (AT3G12880)

The second candidate gene included in this T-DNA insertion line screen was a gene which has been described as having pectin methylesterase (PME)/pectin methylesterase inhibitor (PMEI) function, AT3G12880. This candidate was selected based on the marker association detected on chromosome A5/C5 where the most highly associating marker was JCVI_39914:1128 (see Section 3.2.2.1). This association peak was detected in the Associative Transcriptomics analysis for absolute stem strength (Fmax and F/V in the JIC analysis) as well as for stem diameter (JIC analysis) and stem weight (KWS analysis). For this candidate, it was possible to obtain seed for three different insertion lines, Salk_085493, SAIL_602_G12 and SAIL_649_D02. The first two of these were obtained in time to allow for genotyping across two generations. For Salk_085493 both homozygotes and heterozygotes were identified in the first round of genotyping (Figure 3.38). One representative for each of these genotypes was taken forward for a second year of genotyping. However, in doing this it was revealed that both parental plants were homozygous for the insertion. This anomalous finding could be the result of a number of possible factors. It is possible that the samples were mixed up or mislabelled. It is also possible that the heterozygous phenotype observed in the first generation was in fact the result of contamination with WT DNA. Given the presence of the WT control band seen in genotyping the second generation, it is unlikely that the unexpected homozygous genotype revealed was the result of PCR reaction failure. For this reason, only homozygotes confirmed in both generations were kept on for further analysis. The genotyping of the second generation of this genotype also revealed some unexpected banding on the gel (although the bands cannot be seen clearly, non-specific banding was seen in wells marked with an orange “X” in Figure 3.43). Although most likely the result of contamination during leaf collection/DNA extraction, these samples were omitted from all further analysis, leaving just those plants confidently confirmed as homozygous.

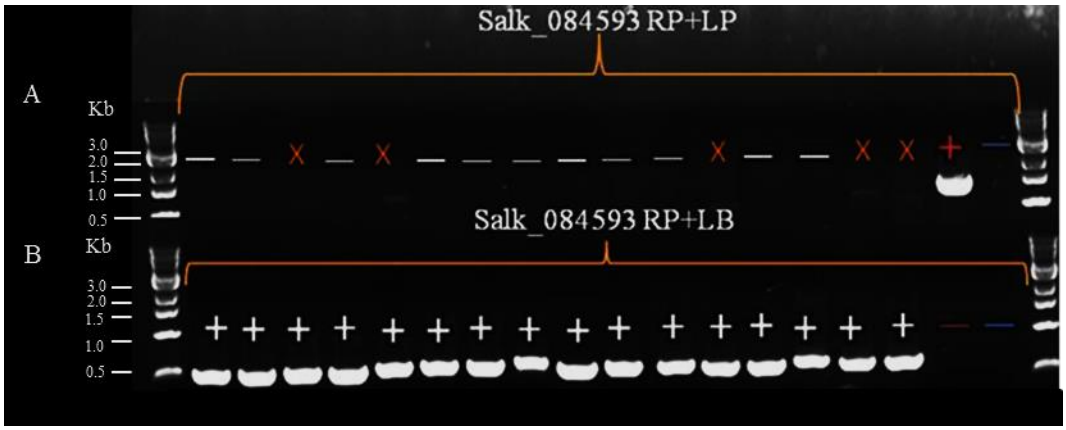

Figure 3. 43. Genotyping results of Salk_085493 in the second generation. Seed originated from a parent of homozygous genotype in first generation. "+" and "–" signs indicated the presence and absence of PCR product respectively. The signs are shown in red and blue for PCR reactions carried out with WT DNA and water controls respectively. Potential contamination can be seen marked with an orange "X"(1Kb ladder used).

For SAIL_602_G12, homozygous and heterozygous plants were identified in the first round of genotyping (Figure 3.38). These genotypes were confirmed in a second generation (Figure 3.44). However, again, there was some non-specific banding for the homozygotes. In most cases, these bands are much smaller than those seen for the WT control. However a couple of bands similar in size to that of the WT control were also observed (marked with an "x" in Figure 3.44A). These, although likely to be due to contamination of the DNA samples, were again omitted from all further analyses.

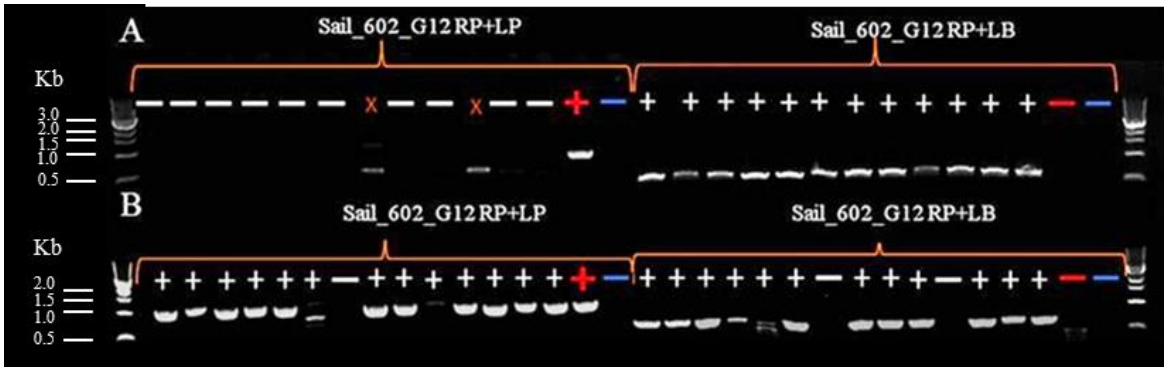

Figure 3. 44. Genotyping results of SAIL_602_G12 in a second generation. This screen allowed for the confirmation of both homozygous (A) and heterozygous (B) genotypes initially detected in the first generation of genotyping. "+" and "– " signs indicated the presence and absence of PCR product respectively. The signs are shown in red and blue for PCR reactions carried out with WT DNA and water controls respectively. Non-specific banding can be seen marked with an orange "x" (1Kb ladder used).

The final T-DNA insertion line obtained for this candidate gene was SAIL_649_D02. Due to the late arrival of seed, it was not possible to genotype these plants across two generations. To allow detection of a suitable number of both homozygous and heterozygous individuals in a single generation, 24 plants for this mutant line were genotyped. Both homozygotes and heterozygous plants were identified (Figure 3.45).

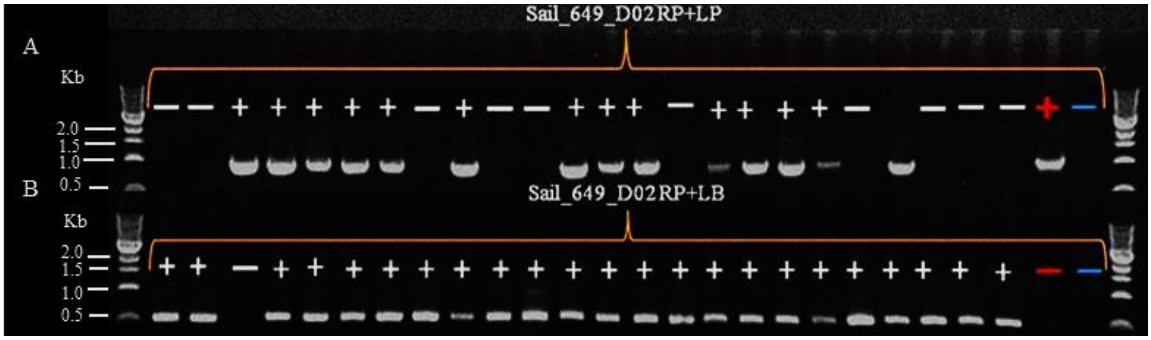

Figure 3. 45. Genotyping of SAIL_649_D02 allowing for the detection of both homozygotes and heterozygotes in a single generation. "+" and "–" signs indicated the presence and absence of PCR product respectively. The signs are shown in red and blue for PCR reactions carried out with WT DNA and water controls respectively (1Kb ladder used).

For Salk_085493, a reduction in second moment of area and both Fmax and F/V in comparison to WT was observed. However, these differences were not statistically significant (P>0.05) (Figure 3.46). Although, slightly reduced plant height was observed in this mutants, this was not found to be significantly different to WT (P>0.05). These plants exhibited no obvious mechanical weakness during growth.

For SAIL_602_G12, no significant changes in stem mechanical strength were observed in the heterozygous plants (Figure 3.46). These plants were also morphologically very similar to WT (Figure 3.47B) and no significant reduction in plant height was observed (P>0.05). In contrast, the homozygous individuals displayed significantly reduced stem diameter (P<0.05) and second moment of area (P<0.01) in comparison to WT. These homozygotes were also found to exhibit significantly reduced Fmax (P<0.05) and F/V (P<0.01) relative to Col-0 (Figure 3.47C). At 35 days these plants were considerably shorter than WT and were found to display a much reduced level of branching (Figure 3.47C). At 45 days, although some homozygous individuals remained considerably shorter than WT (Figure 3.47D), there was some variation in final height between these plants. This variation resulted in no significant difference in plant height relative to WT (P>0.05).

For SAIL_649_D02 both the homozygotes and heterozygotes exhibit altered structural and mechanical properties in comparison to Col-0 (Figure 3.46). Homozygous plants have significantly reduced stem diameter and second moment of

area ($P<0.05$). Reduced Fmax and F/V was also seen in homozygous plants ($P<0.05$). Heterozygous plants for this insertion line also exhibited significantly reduced stem diameter; second moment of area; Fmax and F/V ($P<0.001$). Although heterozygotes were more highly significantly different to Col-0, no significant differences were observed between homozygous and heterozygous plants for these traits. Both homozygote and heterozygotes displayed very similar phenotypes throughout growth (Figure 3.48). Both genotypes exhibit significantly reduced plant height ($P<0.05$ for homozygotes and $P<0.001$ for heterozygotes).

In Arabidopsis (TAIR) this candidate gene has been annotated as having pectin methylesterase (PME)/pectin methylesterase inhibitor (PMEI) activity. Some pectin methylesterases have been found to contain domains of high similarity to those seen in pectin methylesterase inhibitors. This may be the reason for the ambiguity in gene annotation observed. The gene annotation given in TAIR, describes the presence of an inhibitor Pro domain (pectinesterase inhibitor: IPR006501). This domain is known to be a feature in pectin methylesterases, but is very similar to the PMEI domain of pectin methylesterase inhibitors (Pelloux et al., 2007; Wang et al., 2013). Given this, based on sequence information alone, it is difficult to propose the mode of gene function here i.e. whether it has a role in cleaving methyl ester pectin side groups, or whether it is an inhibitor of this process. To further assess the role of this gene in contributing to stem mechanical strength, FTIR analysis was carried out on the stem material used in mechanical testing for homozygous SAIL_649_D02. This was carried out as described in Section 3.4.1.9 (see Supplementary Data file 3.4b for raw FTIR spectra data). Figure 3.49 summarises the results obtained from this analysis. Reports comparing the FTIR spectra of high and low methylesterified pectin, suggest that it is a shift in the ratio of two spectral peaks which is the important discriminating factor. The first of these peaks is found at1740 cm^{-1}. The second is a peak at between 1600 and 1630 cm^{-1}, which would show a higher or lower absorbance relative to the 1740 cm^{-1} in low methylesterified and high methylesterified pectins respectively (Szymanska-Chargot and Zdunek, 2013). As can be seen from Figure 3.49, the SAIL_649_D02 mutants show enrichment relative to WT at 1740 cm^{-1} and depletion at 1624 cm^{-1}. This finding may suggest that this pectin-related gene is functioning as a pectin methylesterase as opposed to an inhibitor of demethylesterification.

As previously described, there are two opposing hypotheses for the effect of pectin demethylesterification on the cell wall. One proposed effect is that of cell wall loosening resulting from the removal of methyl ester groups from the cell wall pectins. In this scenario, it is thought that the reduced methylesterification promotes the degradation of the cell wall pectin by enzymes, and that this degradation promotes cell wall loosening. The second scenario proposed, suggests that the removal of methyl esters by PMEs, increases the likelihood of calcium bridges forming, which ultimately rigidifies the plant cell wall (Hongo et al., 2012; Müller et al., 2013). It seems most likely that the first of these hypotheses provides the best explanation for the results presented here. Organs reach their final size through a combination of cell proliferation and expansion (Breuninger and Lenhard, 2010). For a cell to expand, the cell wall must first loosen, allowing for the incorporation of additional cell wall material (McCann et al., 2001). The reduced stem thickness seen in mutants lacking a functional copy of a gene which seems to have PME function may be the results of reduced cell expansion. Based on the above hypothesis, this could be the result of a high level of pectin methylesterification (indicated in FTIR analysis results) acting as a barrier to cell wall hydrolysing enzymes which act to promote cell wall loosening for cell growth (Xiao et al., 2014).

It must be noted, that in addition to the changes in parts of the FTIR spectra which describe the ratio of methylesterified and demethylesterified cell wall pectin, the spectra for this mutant also shows a great deal of variation in other cell wall components relative to WT. Between 800-1090 cm^{-1} there is a general enrichment in this mutant relative to WT. This region is known to contain information describing a wide range of cell wall components including pectin, but also cellulose, xyloglucan (a hemicellulosic component) and galactose (Kacuráková et al., 2000). It is therefore very difficult, based on the results presented here, to propose which of these components are causing this signal. Furthermore, it is difficult using such reverse genetics approaches, given the large changes observed in the FTIR spectra, to confidently attribute the mechanical defects observed in the described mutant to altered pectin methylesterification alone. It is likely that the cell wall of this mutant is compensating for the changes induced in cell wall structure by altering the amounts/composition of other cell wall components. This, although biologically very interesting, complicates the interpretation of the results. What can be said however is

that the disruption of this gene, which based on the results presented here is acting as a pectin methylesterase, brings about cell wall changes which ultimately contribute to stem mechanical strength.

This candidate gene was, as previously mentioned, selected based on a SNP association detected for absolute strength and stem structural strength. This association was also detected (although to a weaker extent) for stem weight. Given the above discussion, it seems very plausible that a gene involved in the pectin demethylesterification could have an effect on each of these traits and the evidence suggests that this may be the correct candidate gene to which the A5/C5 SNP association corresponds.

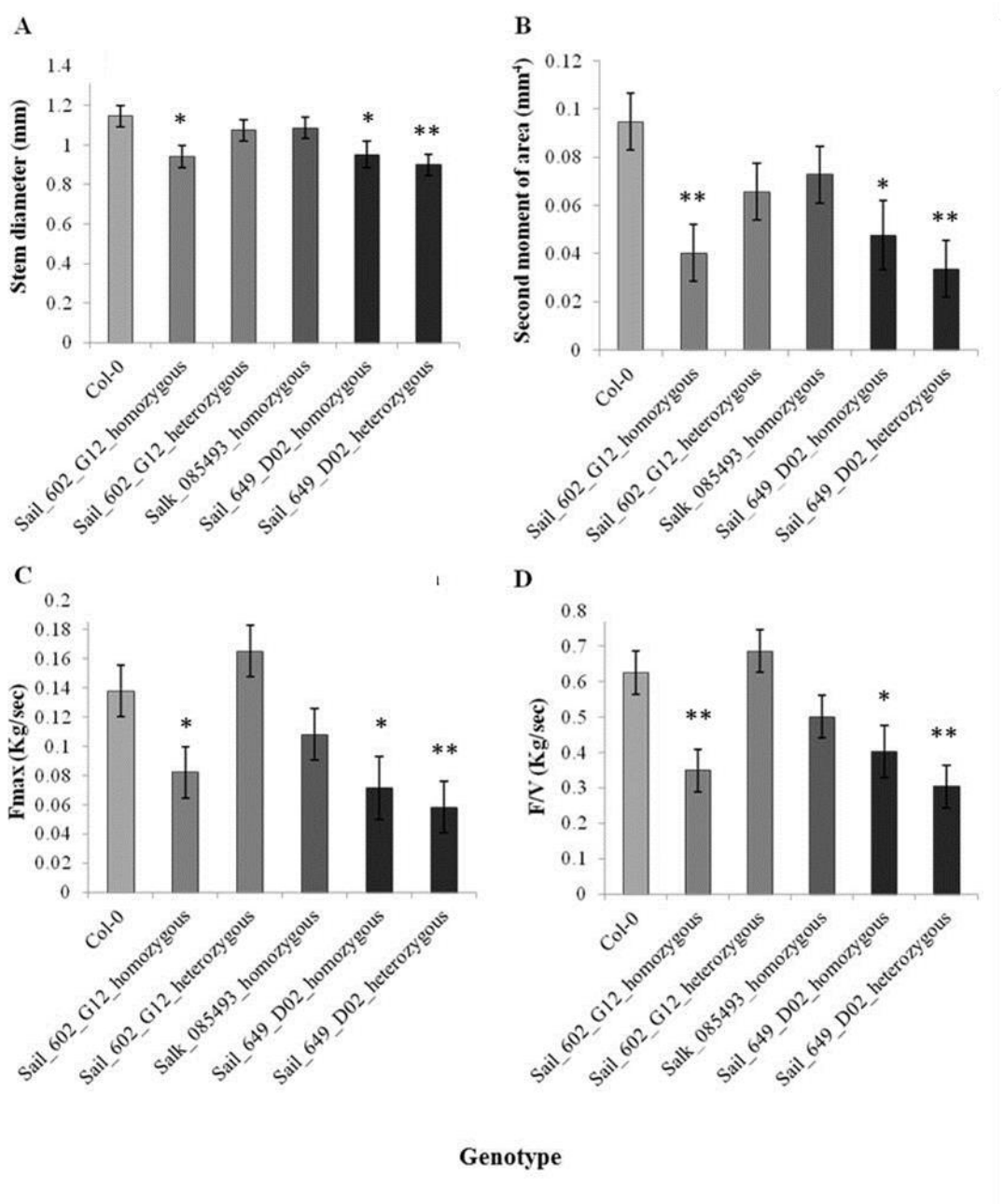

Figure 3. 46. Variation in stem diameter (A), second moment of area (B), Fmax (C) and F/V (D) observed between the PMEI/PME T-DNA insertion lines and Col-0. The asterisks are indicative of the significance level of the differences observed in these traits between the different mutant genotypes and Col-0 where *=<0.05, **=<0.01 and *=<00.1 as determined through the student T-test. The error bars represent the standard error of the means calculated through ANOVA.**

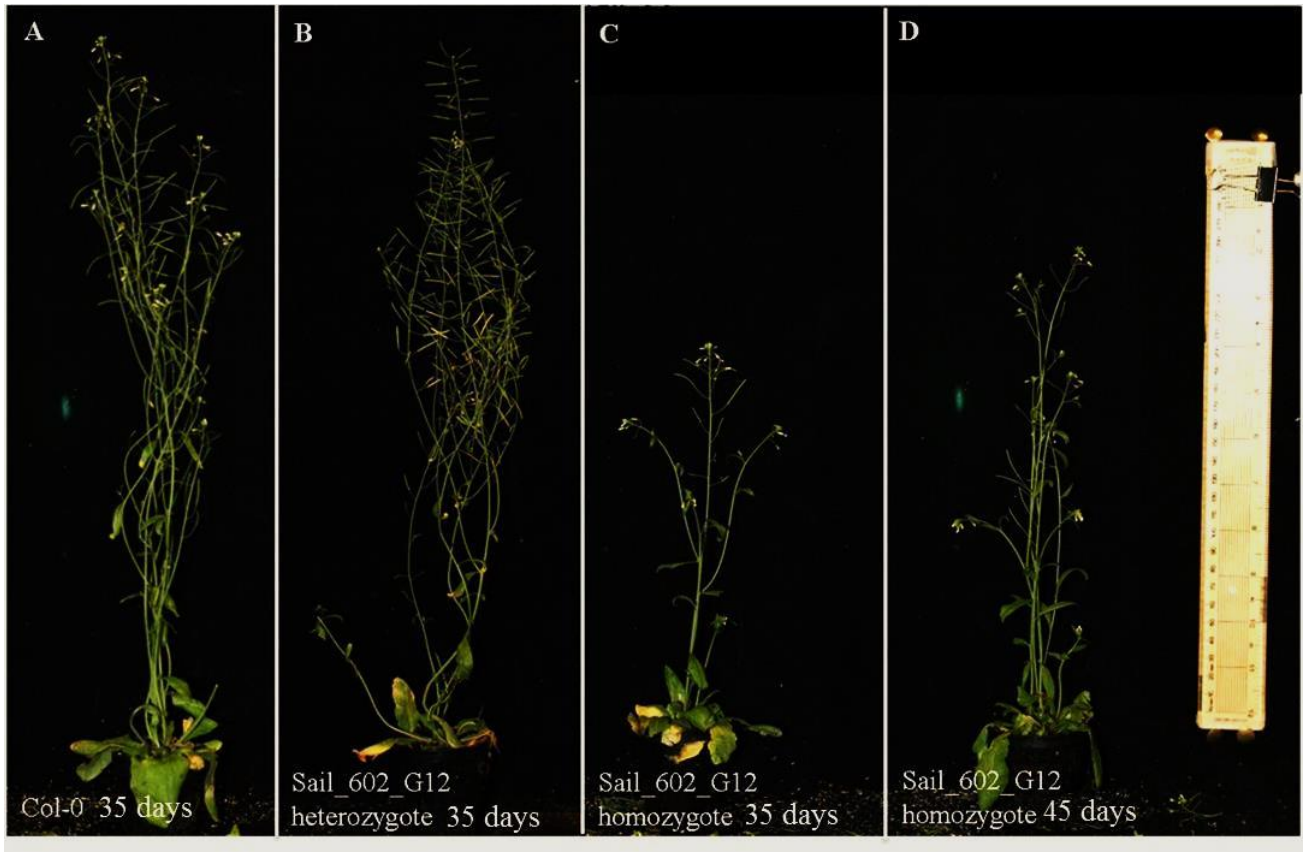

Figure 3. 47. The growth of a typical Col-0 plant (A) alongside a heterozygous SAIL_602_G12 plant at 35 days (B) and a homozygous SAIL_602_G12 plant at 35 (C) and 45 days (D).

Figure 3. 48. The growth of a typical Col-0 plant (unsupported) (A) alongside a heterozygous SAIL_649_D02 supported and unsupported plant at 35 days (B) and a homozygous SAIL_602_G12 supported and unsupported plant at 35 (C).

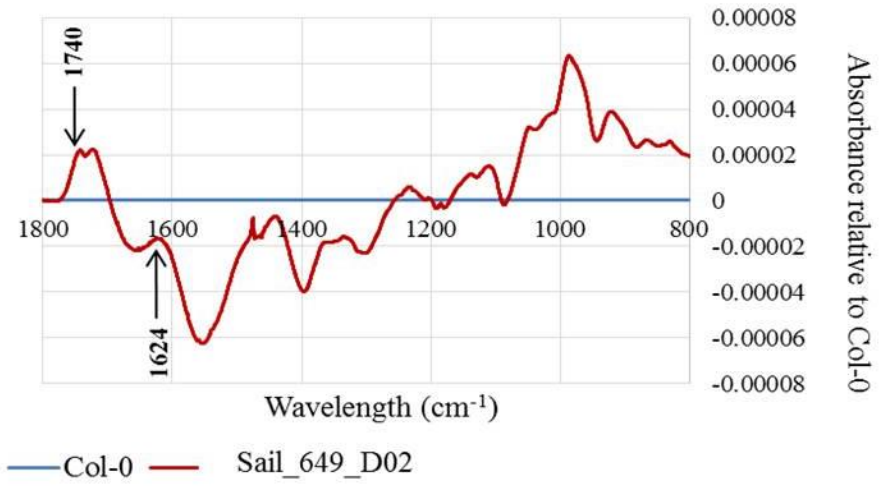

Figure 3. 49. FTIR results describing the chemical composition of the stems of homozygous plants for the T-DNA insertion line SAIL_649_D02 (dark blue line), relative to that of Col-0 (pale blue base line) The peak marked at 1740 and 1624 indicate spectral regions known to describe high methylesterification and demethylesterification respectively (Szymanska-Chargot and Zdunek, 2013).

3.4.2.1.3. *SAUR-LIKE AUXIN RESPONSE 72* (AT3G12830)

Two potential candidate genes were detected in close proximity to JCVI_39914:1128 (see Section 3.2.2.1). The first of these was described previously in Section 3.4.2.1.2. However, in addition to this, an orthologue of the Arabidopsis *SAUR72* gene was also detected. Only a single T-DNA insertion line, SAIL_564_B07, was available for this candidate gene. Again, due to the late arrival of seed, it was not possible to genotype this line across two generations. However, as describe for SAIL_649_D02, it was possible to identify both homozygous and heterozygous individuals in a single generation (Figure 3.50).

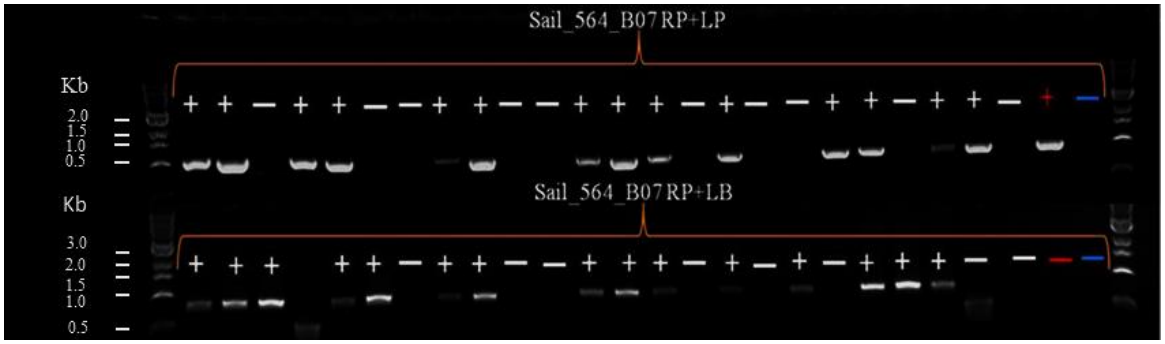

Figure 3. 50. Genotyping results for SAIL_564_B07 allowing for the detection of both homozygote and heterozygote plants in a single generation. "+" and "–" signs indicated the presence and absence of PCR product respectively. The signs are shown in red and blue for PCR reactions carried out with WT DNA and water controls respectively (Ladder key shows 1Kb ladder, New England BioLabs Inc., Massachusetts, USA).

Through mechanical testing, although the homozygous plants were found to exhibit reduced stem diameter and Fmax in comparison to Col-0, these differences were found to be non-significant ($P>0.05$). Second moment of area and F/V were however found to be significantly reduced in comparison to Col-0 in individuals homozygous for this insertion ($P<0.05$). No significant differences were observed for any mechanical traits in the heterozygous plants (Figure 3.51). Throughout development, these mutants were found to display a number of interesting phenotypes. At 35 days of growth, both homozygous and heterozygous plants were seen to be slightly shorter than WT ($P<0.05$). They also exhibited reduced branching (Figure 3.52). The detection of a branching phenotype in this mutant is not surprising, given the known role of auxin in controlling outgrowth (Bennett et al., 2006). Unlike the heterozygous plants, homozygous individuals exhibited a pendant stem phenotype at 35 days (Figure 3.52C).

Although many studies have focused on the importance of auxin in various aspects of plant development, relatively little is known about SAURs specifically. Recent work carried out in Arabidopsis found that *SAUR72* is highly expressed in the hypocotyl (Qiu et al., 2013). It has also been suggested that SAUR genes may be important in coordinating cell proliferation and expansion (Qui et al., 2013). Given this, it is logical that a mutant defective in *SAUR72* function, is observed to have a reduction in the structural strength trait, second moment of area (and as a result,

reduced absolute strength), which is determined by the size and geometry of the stem.

SAU72 and what seems to be a PME gene were both selected as potential candidate genes for a single SNP association detected on chromosome A5/C5. Given that significant alterations in stem mechanical strength are observed in the mutants for both candidates, it is difficult to conclude whether the expected PME gene or the orthologue of *SAUR72* is the gene to which the association peak corresponds. It may seem logical to assume that given the greater effect seen in the PME mutants, that this is the most likely candidate gene. However, given that one of these mutants showed no effect on stem structural and absolute strength and that only a single insertion line was available for *SAUR72*, this is not a fair assumption. It is of course possible that the association peak detected corresponds to both of these genes which may act independently or together to contribute to stem mechanical strength. Recent work carried out by Braybrook and Peaucelle (2013), provides evidence to suggest that auxin may be involved in the induction of pectin demethylesterification (Braybrook and Peaucelle, 2013). To explore any potential role of *SAUR72* in contributing to cell wall pectin demethylesterification, FTIR analysis was carried out on the stem material taken from homozygous SAIL_564_B07 plants (see Supplementary Data file 3.4b for raw FTIR spectra data). Figure 3.53 shows a plot of the differences in FTIR spectra in this mutant relative to WT. Enrichment is seen at 1743 cm^{-1} relative to WT. In addition, at 1605 cm^{-1} a small depletion relative to WT can be seen. This pattern is very similar to that observed previously for the gene proposed to act as a pectin methylesterase. This may suggest that *SAUR72* is also involved in determining the methylesterification state of cell wall pectin. It is possible that this gene interacts directly with the gene proposed here to be a PME or may influence the function of other genes involved in pectin demethylesterification. To explore this, further experimentation would be required. It may for example be of interest to assess any expression patterns between these genes. Co-expression patterns would provide further evidence of an interaction between these genes. It would also be advantageous to obtain a more sensitive measure of pectin methylesterification in these mutants.

When comparing the FTIR spectra for the gene proposed to act as a PME and that of the *SAU72* mutant, although commonalties were observed in the regions assigned to

varying levels of pectin methylesterification, other parts of the spectra show a very different pattern between mutants. This suggests that these genes, although may have a similar role in the controlling the methylesterification state of cell wall pectins, are having very different on other cell wall components.

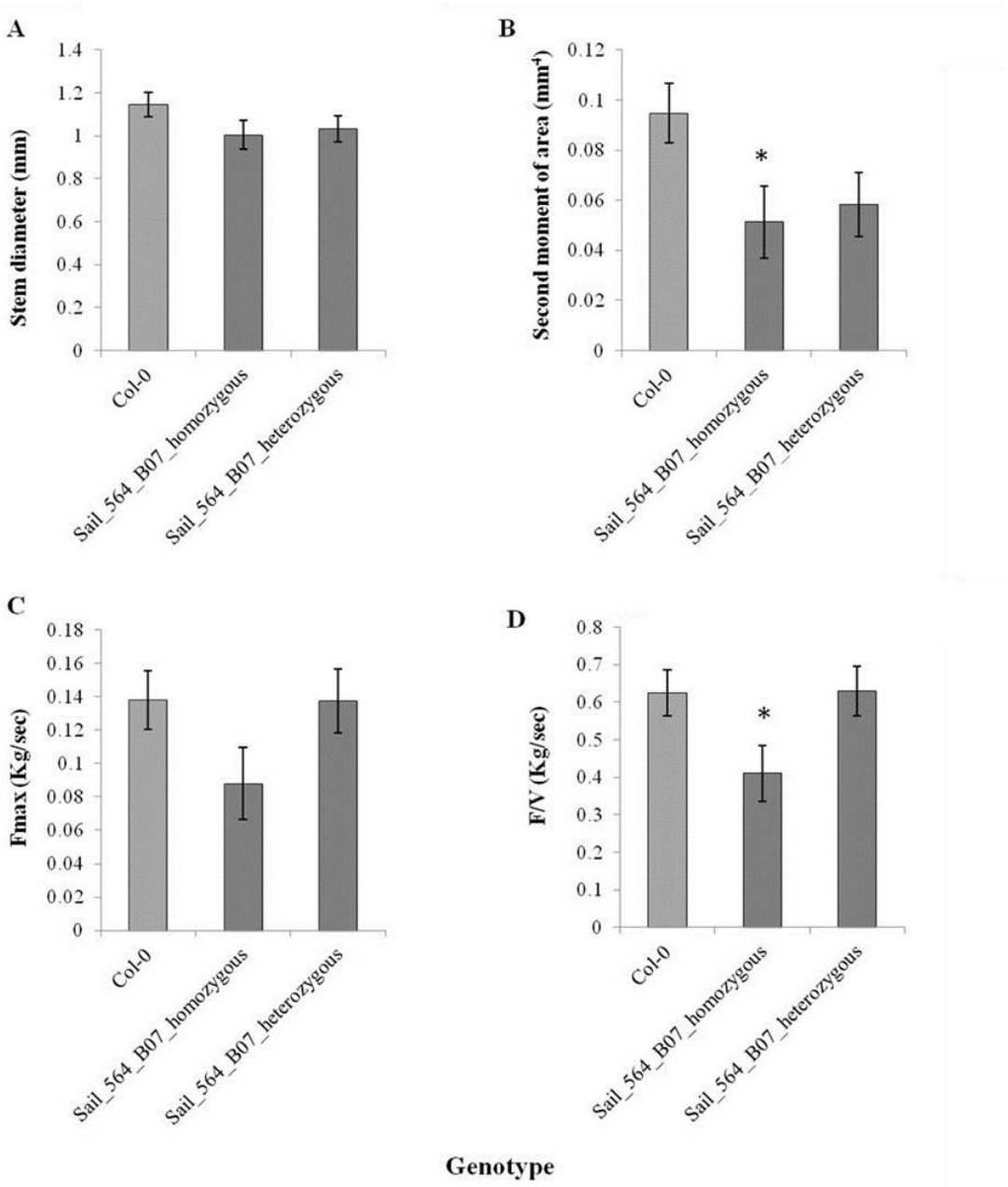

Figure 3. 51. Variation in stem diameter (A), second moment of area (B), Fmax (C) and F/V (D) (A and B) and stem absolute (C and D) strength observed between the homozygous and heterozygous SAUR mutants (SAIL_564_B07) and Col-0. The asterisks are indicative of the significance level of the differences observed in these traits between the different mutant genotypes and Col-0 where *=<0.05, **=<0.01 and *=<00.1 as determined through the student T-test. The error bars represent the standard error of the means calculated through ANOVA.**

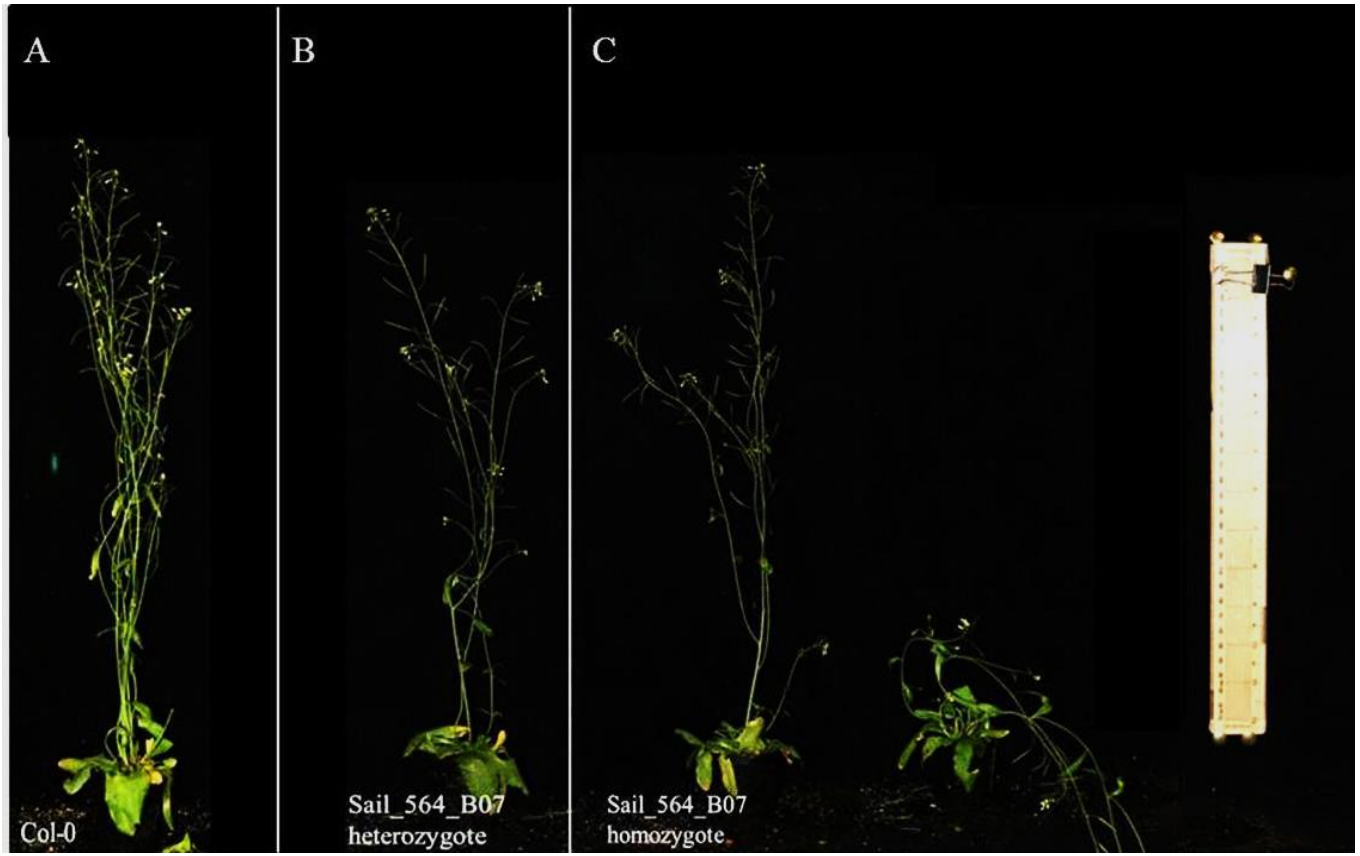

Figure 3. 52. The growth of a typical Col-0 (unsupported) (A) alongside heterozygous (unsupported) (B) and a homozygous Sail_564_B07 (supported and unsupported) (C) as seen after 35 days of growth

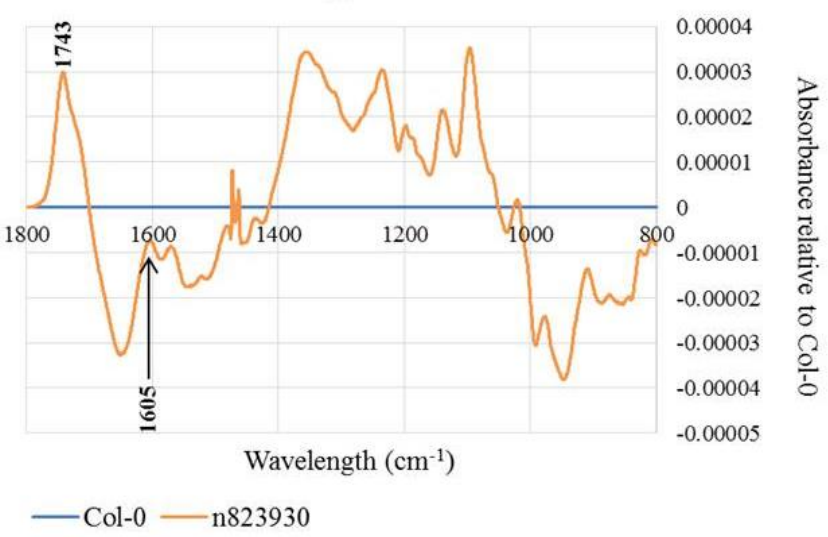

Figure 3. 53. FTIR results describing the chemical composition of the stems of homozygous plants for the T-DNA insertion line SAIL_564_B07 (orange line), relative to that of Col-0 (pale blue base line). The peak marked at 1743 and 1605 indicate spectral regions know to describe high methylesterification and demethylesterification respectively (Szymanska-Chargot and Zdunek, 2013)

3.4.2.1.5. *VASCULAR-RELATED NAC-DOMAIN TRANSCRIPTION FACTOR 2* (AT4G36160)

The final T-DNA insertion line to be included in this study was for a NAC transcription factor, *VND2* (AT4G36160). This gene belongs to a family of transcription factors known to be involved in cell wall biosynthesis in plants. For example, *VND7* is known to be an important regulator of xylem differentiation and is thought to act upstream of pathways known to be important in cell wall lignin and cellulose biosynthesis. *VND7* has also be found to interact with *VND2* (Hussey et al., 2013). Although the exact role of *VND2* is not clear (Hussey et al., 2013), the fact that it has been found to interact with a gene known to alter the biosynthesis of cell wall components thought to be important for stem mechanical strength (Jones et al., 2001), suggests that it may be a good potential candidate gene in the present study.

This candidate was selected based on the close proximity of this gene to the most significant marker, JCVI_19955:172 on chromosome A1/C1 for the material strength traits MOE and MOR (see Section 3.2.2.4). For this candidate gene, only a

single T-DNA insertion line was available, Salk_022124. This mutant line was genotyped across just a single generation due to the late arrival of seed. The genotyping of these plants revealed completely homozygous seed (Figure 3.54). Although many of the PCR reactions failed in this screen, given the great number of homozygotes observed, it is unlikely that this batch of seeds were segregating for the mutation. However, to be confident of the genotype of those plants being mechanically tested, only those for which the PCR reaction was successful were selected.

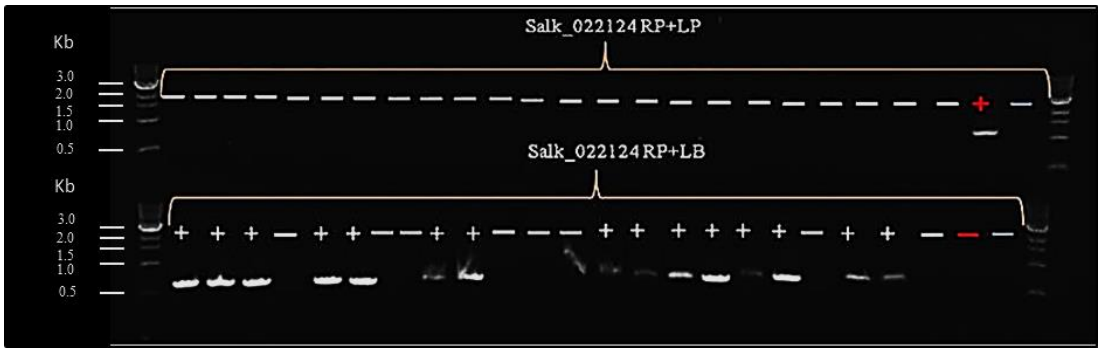

Figure 3. 54. Genotyping Salk_022124 revealed plants completely homozygous for the insertion. "+" and "–" signs indicated the presence and absence of PCR product respectively. The signs are shown in red and blue for PCR reactions carried out with WT DNA and water controls respectively (Ladder key shows 1Kb ladder, New England BioLabs Inc., Massachusetts, USA).

No significant changes in MOR or MOE (Figure 3.55A and B) were observed between these homozygous mutants and Col-0. Furthermore, no significant changes were seen for the absolute strength and stem structural traits for these plants. The screening of this mutant therefore provides no evidence to suggest that this gene has a role to play in contributing to stem mechanical strength. During growth, these mutants were morphologically very similar to WT. However Salk_022124 plants were found to show a slight reduction in branching (Figure 3.56). No significant difference in plant height relative to WT was observed ($P>0.05$).

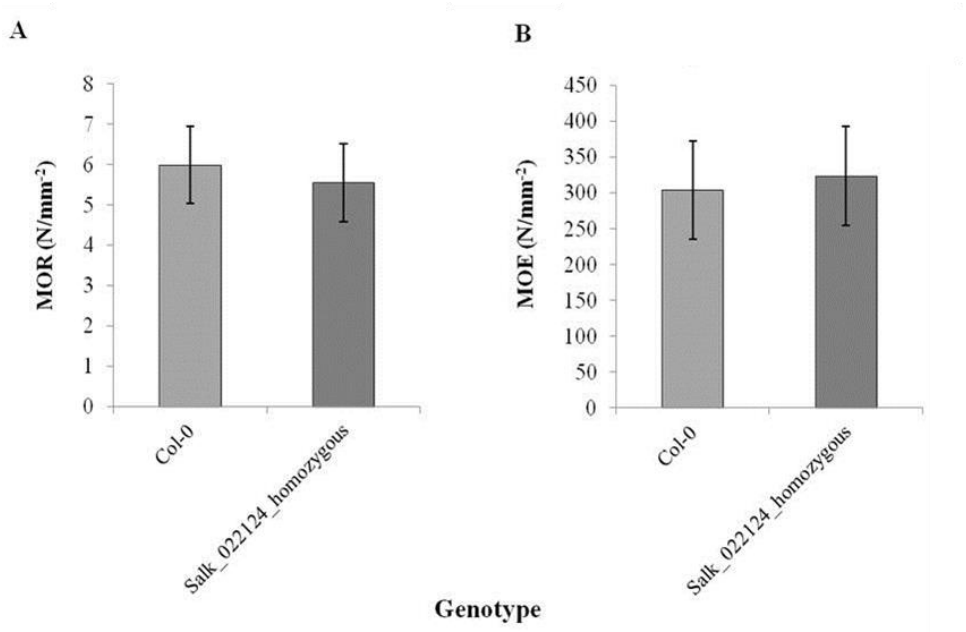

Figure 3. 55. Variation in MOR (A) and MOE (B) between homozygous mutants, Salk_022124 and Col-0. The asterisks are indicative of the significance level of the differences observed in these traits between the different mutant genotypes and Col-0 where *=<0.05, **=<0.01 and *=<00.1 as determined through the student T-test. The error bars represent the standard error of the means calculated through ANOVA.**

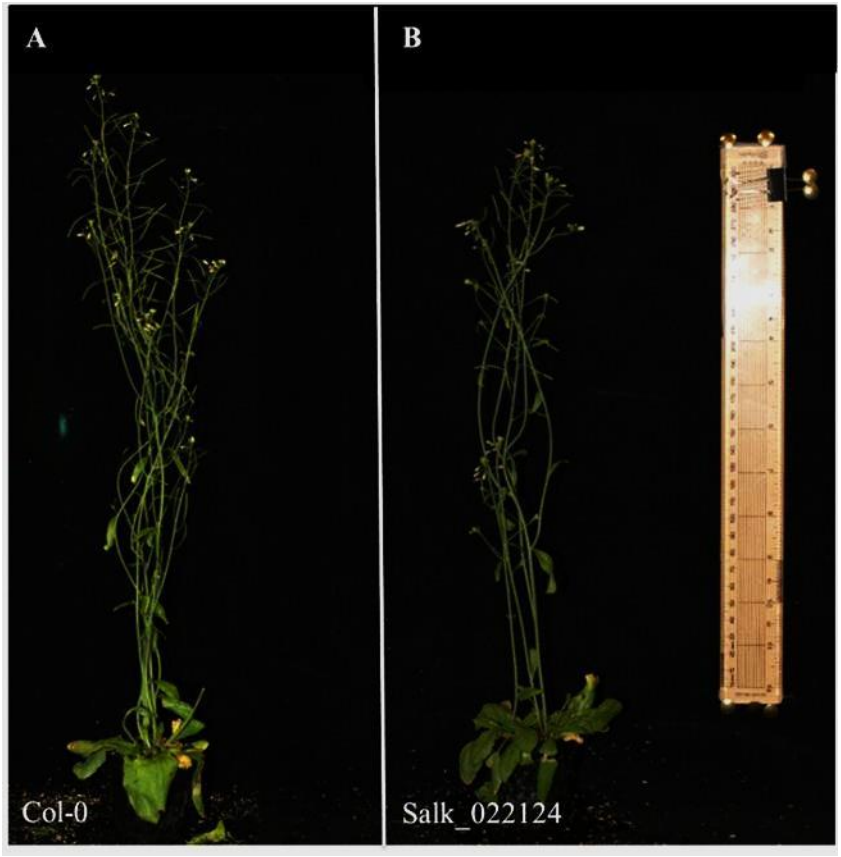

Figure 3. 56. Growth of a typical Col-0 plant (A) alongside a typical homozygous plant for the T-DNA insertion line Salk_022124 (B) as seen after 35 days of growth.

3.4.2.2. Discussion

Through Associative Transcriptomics it was possible to propose several candidate genes for the mechanical strength traits explored in *B. napus*. This section has described the work carried out to further explore the function of some of these genes, through the screening of Arabidopsis T-DNA insertion lines for altered stem mechanical strength.

For a number of the mutants included, significant differences in stem mechanical strength were observed. The results presented here, suggest that *GAUT5* (AT2G30575) is important in contributing to stem structural strength in Arabidopsis, as SAIL_401_E11 homozygotes and heterozygotes exhibit reduced stem diameter and second moment of area relative to WT. These plants were also found to have altered absolute strength. No significant effect on these traits was however seen in the *GAUT5* insertion line, Salk_050186. This, as previously mentioned, may be due

to inaccuracies in genotyping, possibly the result of non-specific amplification of additional DNA fragments of high sequence similarity. This would explain the apparently completely heterozygous genotype of these plants. If this was the case, it is possible that the phenotypic data collected for these plants were based on a combination of different genotypes for the insertion. This would be likely to introduce a high level of error in the phenotypic measurements between samples, therefore explaining the lack of significant difference observed between these plants and WT. It is of course possible that these plants were true heterozygotes. However, based on the segregation ratio expected from a (presumably) heterozygous parent, this seems unlikely.

In addition to the role of *GAUT5* in contributing to stem mechanical strength, the screening of these Arabidopsis T-DNA insertion lines also supports the proposed role of a gene described as having pectin methylesterase inhibitor/ pectin methylesterase activity, AT3G12880, in contributing to these traits. For this candidate gene, three different T-DNA insertion lines were analysed, two of which (SAIL_649_D02 and SAIL_602_G12) were seen to exhibit reduced stem structural and absolute strength relative to WT. The third T-DNA line for this candidate gene, Salk_085493, showed no significant difference in mechanical strength in comparison to WT. There are a number of potential reasons for the lack of phenotype observed here. The T-DNA insertion line, Salk_085493, carries an insertion in the promoter region. Previous reports have shown that promoter-based insertions may cause a reduction in the level of gene expression (sometimes referred to as a "knock-down") but may fail to altogether knock-out gene function (Wang, 2008). This may explain why no significant changes in stem mechanical strength were seen in this mutant, where even a low level of expression may be enough to rescue any phenotypes related to complete loss of function. To explore this further, it would be useful to assess transcript levels in this mutant through RT-PCR.

Based on the results obtained from FTIR analysis, it seems likely that this gene may in fact be acting as a pectin methylesterase as opposed to an inhibitor. Pectin methylesterases are involved in the removal of methyl ester groups from cell wall pectins. This process (demethylesterification) has been proposed to have two opposing effects on the cell wall. One effect of demethylesterification is cell wall stiffening, resulting from the formation of intermolecular calcium cross-bridges. The

second effect is cell wall loosening, resulting from the degradation of cell wall pectins by enzymes, such as polygalacturonases, which is important in cell growth (Liu et al., 2013; Xiao et al., 2014). Given that the mutants were found to exhibit reduced stem thickness (the likely cause of the reduced absolute strength also seen in these mutants), which may result from reduced cell expansion, it seems likely that it is a decrease in the latter effect which is being seen in SAIL_649_D02.

Based on phenotypes reported in the literature, it is expected that *GAUT5* may also contribute to stem mechanical strength through its effect on cell expansion. Mutants defective in other genes encoding galacturonosyltransferases have been found to exhibit reduced cell size. They also display fewer xylem cells in the stem, which may be indicative of reduced activity in the vascular cambium (Nieminen et al., 2004; Guo et al., 2009). In plants, the growth of an organ occurs as a result of two consecutive (but overlapping) phases: a cell proliferation phase and a cell expansion phase (Feng et al., 2011). Given the important role of stem diameter in contributing to stem absolute strength in *B. napus* (see Section 3.1.2.2), it is very plausible to propose that such genes expected to alter cell size (and potentially controlling cell proliferation) are likely candidates. To allow for this to be explored further in the mutants included in the present study, an in depth microscopy screen would be required, to assess any changes in cell size and differentiation.

The results presented here also suggest that a gene involved in auxin signalling, *SAUR72*, may be involved in contributing to stem mechanical strength in Arabidopsis. For this candidate gene, only a single T-DNA insertion line was analysed. However, significant differences in both absolute and structural strength were observed in the homozygotes for this mutation. Like the pectin-related genes described so far, SAURs have also been implicated in determining cell size. Furthermore, auxin has been implicated in the regulation of pectin demethylesterification, which based on the results obtained from the FTIR analysis may be the case for *SAUR72*. It must however be noted, that the mutants defective in *SAUR72* and PME function were seen to be phenotypically very different during development. However, this may be indicative of pleiotropic effects of the induced mutations. This may be particularly likely for the mutants defective in *SAUR72* function, given the many proposed roles of auxin in plant development (Vanneste and Friml, 2009). Furthermore, based on the FTIR analyses, it seems that although

commonalities in pectin methylesterification state were seen, many additional differences in cell wall composition were also observed, which may contribute to the differences in plant morphology observed. It is also possible that the morphological differences are the result of additional mutations in other parts of the genome. Although background mutations are rare in T-DNA insertion lines (in comparison to EMS-mutagenised plants for example), it is expected that on average approximately 50% of the T-DNA insertion lines developed will carry more than one insertion (The Salk Institute Genomic Analysis Laboratory). It is also possible that *SAUR72* and the gene proposed to have PME activity are acting independently on these strength traits. It could be for example, that the pattern of methylesterification seen in the *SAUR72* mutants is the result of a knock-on effect caused by the direct effect of the mutation on some other cell wall component. There are a number of ways through which this potential interaction could be further explored. For example, it would be interesting to explore the expression patterns of both the PME gene and *SAUR72*. Finding that they are co-expressed and that they are expressed in common tissue types may add evidence to the presence of an interaction.

No effect on material strength was observed for the insertion line for *VND2* (proposed as a candidate gene for MOE and MOR in *B. napus*). There are a number of potential explanations for the lack of phenotype observed in this mutant. Firstly, it may suggest that *VND2* is not the correct candidate gene for this trait. As previously discussed, the detection of a potential candidate gene in close proximity to an associating marker by no means proves causality. It must also be considered, that despite being closely related species, there may be differences in the genetic control of these traits between Arabidopsis and *B. napus. VND2* may indeed have a role to play in determining material strength in *B. napus*, while having no effect on the same traits in Arabidopsis. If this is the case, it would be necessary to analyse the effect of these genes in *B. napus* directly using methods such as TILLING. As a final potential explanation, the lack of phenotypes observed in these mutants could be explained by genetic redundancy. For a given biological process, there may be many genes with overlapping function which contribute. Where this is the case, the loss of function of just a single copy, may be compensated for by additional genes with shared function and therefore any expected changes in phenotype may be masked. This has been found to be a particular problem in cell wall related genes which often

belong to large gene families with similar biological function (Cao et al., 2008). Work carried out in Arabidopsis, found that there is an overlap in expression pattern between genes encoding VASCULAR-RELATED NAC domain proteins (the gene family to which *VND2* belongs), which may suggest functional redundancy within this gene family. This would also explain why in the work carried out on mutants defective in *VND6* and *VND7* gene function, no clear phenotypes were observed (Kubo et al., 2005). In such situations, it may be necessary to create multiple gene knockouts before a clear phenotype can be observed.

Although the mechanical testing of these mutants has been very insightful, it would be advantageous to further explore the biological nature of the mechanical defects observed. It would for example be of interest to assess variation in cell size and/or number in the PME, *GAUT5* and *SAUR72* mutants. It would also be of interest to obtain more information regarding the methylesterification state of the cell walls in the *SAUR72* and the mutant which has been proposed to be a pectin methylesterase. Although the spectral peaks which describe this cell wall characteristic are more easily distinguishable in comparison to other cell wall features, it would be advantageous to acquire further evidence to support the observed spectra. One method which could be used to further analyse the methylesterification of cell wall pectins in these mutants is antibody probing. There are specific probes now available which target pectins of different methylesterification states. The level of antibody binding can then be viewed clearly using miscroscopy methods (Lehner et al., 2010).

In cases where just a single T-DNA insertion line was assessed for altered mechanical strength, it would be advantageous to screen additional alleles for the insertion. Detecting common phenotoypes in mutants carrying different insertions for the same gene, provides much more convincing evidence that the phenotype observed is actually the direct result of the mutated gene. In addition, complementation experiments would also be of interest. The rescuing of mutant phenotypes following transformation of the mutant plant with a functional copy of the Arabidopsis target gene, provides evidence to suggest that the phenotypes observed are the result of the loss-of-function. Following this confirmation, it would then be of interest to see if the mutant phenotypes can also be rescued by introducing a functional copy of the *B. napus* orthologue of that gene. This would provide evidence to suggest that the Arabidopsis and *B. napus* orthologues share biological

function. (Krysan et al., 1999). These experiments were unfortunately beyond the scope of the present study, but may be valuable areas for future research.

Many additional candidate genes were proposed for stem mechanical strength and related traits in the Associative Transcriptomics analysis carried out in *B. napus*. Due to time constraints, not all of these genes were explored through the screening of Arabidopsis T-DNA insertion lines. This may also be an area for continued research in the future.

3.5 CHAPTER SUMMARY

This Chapter has described the work carried out to explore the genetic control of stem mechanical strength and related traits in *B. napus*. Through statistical analyses a high level of genetic variation was detected across the 79 *B. napus* accessions included. This suggests that there is good breeding potential for the mechanical strength and related traits analysed.

The results presented here suggest that stem strength in this species is largely determined by stem structural attributes. In Section 3.1.2.2, it was shown that approximately 70% of the variation in absolute stem strength can be explained by stem diameter alone. Structural traits such as stem outer cortex were also found to be important contributors to absolute strength. These traits may therefore be the most important traits on which to focus breeding efforts for improved stem mechanical strength in *B. napus*.

Due to the lack of lodging observed in the field, it was not possible to obtain an accurate measure for this trait. However, it was possible to obtain a measure of plant standing ability under field conditions experimentally, using a pulley system connected to a digital force gauge. This provided information relating to the force required to pull the entire plant body from its upright position, and therefore more closely simulates the effect of wind and heavy rainfall (key contributors to lodging) on the mechanical stability of the plant stem. The finding that these measures relate positively to those obtained from the three-point bend test, suggests that this study has been successful in uncovering variation which is of importance for stem mechanical strength in field conditions. This provides evidence to suggest that the

traits included in this study may be useful in breeding for *B. napus* accessions with improved standing ability.

Through correlation analysis many traits were seen to show close relationships with stem mechanical strength. To obtain a thorough overview of the potentially important genetic components of stem strength, each of these traits were analysed through Associative Transcriptomics. In doing so, a number of common associations were identified, suggesting that the correlations observed have a genetic component. In other cases, association peaks were observed which were unique to just a single trait. An example of this was seen with the association peak detected on chromosome A8, where JCVI_639:522 was the most significant marker. This association peak was detected for Fmax only. As previously proposed, this may be the result of a number of different traits contributing to absolute strength in combination, where the individual contributions alone are not sufficient in explaining the associating variation. Association peaks unique to MOE and MOR were also identified. As previously explained, the relationship between absolute and material strength in *B. napus* is complex. Based on the data presented here, the genetic variation relating to material strength is relatively rare. This may explain the lack of commonality between the association peaks detected for material and absolute strength, where it is simply masked by the much greater contributions made by the structural stem components for example.

The statistical analyses described in Section 3.1.2.1, revealed that there may be a high level of environmental sensitivity for the stem mechanical strength and related traits explored in the present study. Given this, it was important to assess the robustness of the markers detected for these traits for trait selection capability. This validation may be of particular importance given the risk of false positive associations in association studies. To assess marker durability, a marker validation study was carried out. This study screened *B. napus* accessions of previously unknown genotype for variation at marker loci detected through Associative Transcriptomics for the absolute strength trait, Fmax. This experiment revealed that one marker, JCVI_31359:1723, has good trait selection power for this mechanical strength trait. For the remaining markers, in some cases, the lack of trait selection power may, at least in part, be explained by the low level of genetic variation detected at the marker loci in the accessions grown as part of the ASSYST panel.

However, one of the markers, JCVI_39914:1128, detected for absolute stem strength in the Associative Transcriptomics analysis, showed no trait selection power despite there being a high level of variation at this locus in the ASSYST breeding panel. Potential reasons for these negative results were proposed in Section 3.3.3.

Although a number of markers were found to be ineffective in selecting for mechanical strength, it is very encouraging to have identified a durable marker for such a complex trait in this study. The next stage will be to assess whether this marker is successful in reducing true stem lodging susceptibility under field conditions.

In light of the marker validation experiment, it may be advantageous to re-evaluate which of the marker associations detected through Associative Transcriptomics may hold most promise for crop improvement. The SNP marker, JCVI_31359:1723, which proved durable for the selection of Fmax in the marker validation study, was assigned a P value of 9.32E-05. JCVI_39914:1128 and JCVI_639:522 (both of which were found to lack durability for trait selection) obtained respective P values of 4.07E-04 and 5.86E-04. Given this, it may be appropriate to introduce a retrospective significance threshold level of 9.23E-05. All SNP associations found to reach this level of significance can be seen marked with an asterisk in Table 3.3 Based on the results obtained in the marker validation study, these markers may hold most value for the improvement of these traits through marker assisted breeding.

Based on conserved synteny with Arabidopsis, it was possible to propose candidate genes for the mechanical strength and related traits included in the present study. To further explore the function of a subset of these genes in contributing to mechanical strength, an Arabidopsis T-DNA insertion line screen was carried out. Given that *B. napus* and Arabidopsis are closely related species, it is expected that there will be a high level of conserved gene function between them.

Three main SNP associations were detected for Fmax through Associative Transcriptomics. One of these, detected on chromosome A8 was found to be in close proximity to an orthologue of the Arabidopsis *IFL1* gene. As previously described, this is a very good candidate gene for stem strength. However, given that this gene has already been explored in terms of its role in stem mechanical strength (Zhong and Ye, 1999), little would have been gained from its inclusion here. The strongest

SNP association for this trait was detected on chromosome A2/C2 (with the most significant marker being JCVI_31359:1657), an association peak found to be in close proximity to two candidates orthologous to Arabidopsis genes expected to have PMEI/PME activity. Unfortunately, there were no T-DNA insertion lines available for these candidates. However, it was possible to assess the role of an additional PMEI/PME gene, AT3G12880, proposed as a candidate gene for the third marker association detected for this trait. This SNP association was the aforementioned A5/C5 locus, where JCVI_39914:1128 was found to be the most significant marker. In Arabidopsis, two mutant lines lacking a functional copy of this gene were found to exhibit altered stem structural and stem absolute strength. This is very encouraging, given that this SNP association was detected for both stem diameter and the absolute strength traits in the Associative Transcriptomics analysis for *B. napus*. In addition to this candidate gene, a gene involved in auxin signalling, *SAUR72*, was also selected for this A5/C5 association peak. The mutant defective in this gene also exhibited reduced structural and absolute strength. This may suggest that the association peak detected corresponds to variation between accessions for both the PEMI/PME gene and *SAUR72*, which may act together or independently to alter stem mechanical strength. This potential interaction was further investigated through the screening of these mutants in an FTIR analysis. This allowed for a broad comparison of stem cell wall composition of these mutants relative to Col-0 to be made. This experiment revealed that homozygous mutants defective in PME and *SAU72* function, exhibit a higher level of methylesterified pectin in comparison to Col-0. This provides evidence to suggest that a) the gene proposed to have PMEI/PME activity is acting as a PME, and b) that *SAU72* may also alter the methylesterficiation state of the plant cell wall.

Through Associative Transcriptomics, a SNP marker association was also identified for stem structural strength on chromosome A4/C4. This association was detected in the results obtained for both the JIC and KWS 2010 trial. Within this A4/C4 region is an orthologue of the Arabidopsis *GAUT5* gene which, based on high sequence similarity to other Arabidopsis galacturonosyltransferases, may be involved in pectin biosynthesis. For this candidate gene, one of the mutant lines analysed exhibited reduced absolute and structural strength. This again suggests that pectin may have an important role to play in contributing to stem mechanical strength. Given that each of

these genes are implicated in controlling cell expansion in plants (a process which is very likely to contribute to stem structural/absolute strength in *B. napus*), it is possible that they are working as part of a common pathway which contributes to stem mechanical strength in *B. napus.*

As previously mentioned, the marker association on chromosome A5/C5, for which a PME and *SAUR72* were the proposed candidate genes, was found in the marker validation study to lack power to select for Fmax. This may suggest that these genes contribute to stem mechanical strength to varying degrees depending on environmental conditions. As previously mentioned, it is also possible that this A5/C5 association is a false positive. If this is the case, the identification of candidate genes, for which loss of function mechanical defects in Arabidopsis are observed, may have been the result of random chance i.e. the phenotypes observed in Arabidopsis are in no way related to the variation seen at this A5/C5 locus in *B. napus*. The Arabidopsis genome contains many genes involved in cell wall biosynthesis, many of which may contribute to stem mechanical strength, but which may not be the site of causal variation to which nearby SNP associations correspond. Regardless of this, the results presented here suggest that these genes have a role to play in determining stem structural strength in Arabidopsis, and which may therefore still be good targets for the improvement of stem strength in *B. napus.*

One candidate gene, *VND2,* was explored through the screening of T-DNA insertion lines for the stem material strength traits, MOE and MOR. This mutant did not exhibit significant changes in stem material strength. As previously discussed, it may be that this is simply not the correct candidate gene to which the marker association on A1/C1 corresponds. Although, genome-wide association studies are a powerful way through which QTL can be identified in a non-biased way, there is still a large element of bias in the selection of candidate genes. It may for example be that the true causal variant lies within a gene which has not yet been implicated in a process which may plausibly contribute to stem mechanical strength. The lack of stem mechanical phenotypes observed in these mutants may also be explained by redundancy within the Arabidopsis genome, where any phenotypes resulting from the loss of function of one gene, may be masked by the compensatory effect of others. Other potential explanations for this negative result were explored in Section 3.4.2.2.

The analysis of several traits of potential importance for stem mechanical strength and lodging resistance in *B. napus*, provides a powerful way through which breeders can develop a crop ideotype for lodging resistance in the future development of elite accessions. However, many additional markers and candidate genes were identified for the stem mechanical and related traits analysed in *B. napus* which were not explored further in the present study. These loci should be the focus of future work, to allow their biological importance to be further assessed and for their potential role in crop improvement to be explored.

This Chapter has shown that through Associative Transcriptomics, it is possible to identify durable markers for complex traits of agronomic importance in *B. napus*. It has also however shown that given the potentially high rate of false positive association, transcriptome sequencing artefacts, high trait variances and potential environmental sensitivity of the detected loci, that validation of such results is crucial if the true relevance of such studies to crop improvement is to be realised.

Chapter 4. Stem mechanical properties in hexaploid wheat

CHAPTER OVERVIEW

This Chapter describes the work carried out to explore the genetic control of stem mechanical strength in wheat. The first section will describe the variation available for stem strength and related traits across a panel of 100 wheat accessions. This section will also explore the relationship between these traits and describe the relevance of studying stem mechanical strength for reduced stem lodging susceptibility. The second section will describe an Associative Transcriptomics analysis, conducted for the identification of both SNP and gene expressions markers (GEMs) which may be used in the marker-assisted breeding of wheat accessions with improved stem mechanical strength. This section will discuss potential candidate genes which may be responsible for the causal variation to which the marker associations correspond. Following this will be the description of a marker validation experiment, conducted to allow the efficacy of a subset of the SNP markers detected through Associative Transcriptomics for their use in marker-assisted breeding of elite wheat accessions to be tested. This section will then be followed by a summary of the most important findings from the work carried out in wheat.

4.1 VARIATION IN STEM MECHANICAL PROPERTIES AND RELATED TRAITS ACROSS A PANEL OF 100 HEXAPLOID WHEAT ACCESSIONS

The key objective of this study was to uncover genetic markers which may be used in the marker-assisted breeding of elite wheat accessions with improved stem mechanical properties. These markers are chosen based on the observation of patterns between marker allele segregation and trait variation. Given this, it is important that a good level of phenotypic variation can be detected within the panel of accessions included in the study. The studies to date aimed at exploring variation in stem mechanical strength in wheat, have been based on very small collections of wheat accessions known to vary for lodging susceptibility(Wang et al., 2012), or have utilised mapping populations resulting from the cross of parental plants known to vary for stem strength (Hai et al., 2005). These approaches, although informative

to some extent, are only able to explore a very limited pool of allelic variation, which is very unlikely to represent the true level of variation available for exploitation through breeding. In contrast, the present study has made use of a panel of 100 wheat accessions. This panel, although mainly consisting of winter wheat accessions (and therefore unlikely to represent the variation available at the whole species level), is more likely to represent the level of variation readily available to European wheat breeders. This section describes the analysis of stem mechanical strength across this panel to allow for a more accurate estimation to be made regarding the level of variation available. In addition, this Chapter will explore the relationship between stem mechanical strength and potentially related traits (including stem structural, plant morphological and chemical composition traits). This will not only promote a more in-depth understanding of stem strength in wheat, but will reveal additional traits which may be important for breeding elite wheat accessions and which may therefore qualify for further analyses through genetic mapping.

4.1.1 Methods

4.1.1.1 Plant material and harvest

A panel of 100 accessions of hexaploid bread wheat, *T. aestivum*, were grown at a single site (KWS, Thriplow, UK) across two years (2011 and 2012). Table 4.1 gives an overview of the wheat accessions included within this panel, and provides information regarding the level of across year replication for each accession. For both field trials, each accession was grown in single (non-replicated) 1 m^2 plots. The harvesting of these plants was carried out by hand using secateurs. All plants were cut at the base of the stem, taking as much above-ground material as possible. In 2011, ten plants were harvested at random from the centre of each 1m^2 plot per accession. In 2012, five plants per accession were harvested. Once harvested, each plant sample was labelled with a unique identifier, including cultivar name and year of harvest. To avoid the separation of tillers during transportation, each plant was held together using gardening twine. Prior to further processing, all plants were dried thoroughly at room temperature on mesh drying racks (Building 50, JIC).

Table 4. 1. Wheat accessions grown across two years at KWS, Thriplow. The presence of an accession within a given field trial, is indicated with an "x". 4.1a

Accession name	Sequence name	KWS 2011	KWS 2012	Date	Type
Access	W1_Acc_80.1	X	X	2002	Winter
Alba	W002_Alb_80.1	X	X	1928	Winter
Albatross	W003_ALB	X	X	1977	Winter
Avalon	W009_AVA	X	X	1980	Winter
Alchemy	W004_ALC	-	X	2006	Winter
Ambrosia	W006_AMB_rep	-	X	2005	Winter
Apache	W008_APA	-	X	1949	Winter
Bacanora	W010_BAC	X	X	1988	Spring
Battalion	W011_BAT	-	X	2007	Winter
Beaver	W012_BEA	X	X	1990	Winter
Boreonos	W13_Bor_80.1	X	X	-	-
Buster	W15_Bus_80.1	X	X		Winter
Calif	W017_CAL	X	X	1985	Winter
Cap desp	W020_CapD_80.1	X	X	1946	Winter
Capo	W019_CAP	X	X	1989	Winter
Cezanne	W021_CEZ	X	X	1997	Winter
Charger	W121_Cha_80.1	X	X	1997	Winter
Claire	W22_Cla_80.1	X	X	1999	Winter
consort	W23_Con_80.1	X	X	1995	Winter
Cordiale	W024_COR	-	X	2004	Winter
Courtot	W025_COU	X	X	1974	Winter
Deben	W026_DEB	X	X	2001	Winter
Einstein	W028_EIN	-	X	2003	Winter
Equinox	W029_EQU	X	X	1997	Winter
Erla kolben	W030_ERL	X	X	1961	Winter
Escorial	W31_Esc_80.1	X	X	1987	Winter
Etoile de choisy	W32_EdC_80.1	X	X	1950	Winter
Exsept	W034_EXS	X	X	2001	Winter
Extrem	W35_Ext_80.1	X	X	1969	Winter
Fanal	W36_Fan_80.1	X	X	1953	Winter
Flair	W037_FLA	X	X	1996	Winter
Flame	W38_Flm_80.1	X	X	-	-
Florida	W40_Flo_80.1	X	X	-	-
Galahad	W041_GAL	X	X	1983	Winter

4.1b

Accession name	Sequence name	KWS 2011	KWS 2012	Date	Type
Gatsby	W042_GAT	-	X	2006	Winter
Gladiator	W043_GLAD	-	X	2004	Winter
Glasgow	W044_GLAS	-	X	2005	Winter
Haven	W46_Hav_80.1	X	X	1990	Winter
Hereward	W047_HER	X	X	1991	Winter
Hobbit	W48_Hob_80.1	X	X	1989	Winter
Holdfast	W49_Hol_80.1	X	X	1936	Winter
Humber	W050_HUM	-	X	2007	Winter
Huntsman	W51_Hun_80.1	X	X	1971	Winter
Hustler	W52_Hus_80.1	X	X	1978	Winter
Hybrid 46	W053_HYB	X	X	1946	Winter
Hyperion	W054_HYP	-	X	2006	Winter
Istabraq	W55_Ist_80.1	X	X	2004	Winter
Kawkas	W57_Kaw_80.1	-	X	1972	Winter
Kontrast	W058_KON	X	X	1990	Winter
Leda	W59_Led_80.1	X	X		
Longbow	W60_Lon_80.1	X	X	1983	Winter
Malacca	W62_Mal_80.1	X	X	1999	Winter
Marco	W063_MAR	X	X	1969	Winter
Maris widgeon	W64_MaW_80.1	X	X	1965	Winter
Mega	W067_Meg_80.1	X	X	1974	Winter
Mendal	W68_Men_80.1	X	X	1926	Winter
Mercia	W070_MER	X	X	1986	Winter
Miras	W72_Mir_80.1	X	X	-	Winter
Mironowskaja	W73_Mka_80.1	X	X	-	Winter
Muck	W74_Muc_80.1	X	X	1962	Winter
Multiweiss	W75_Mul_80.1	X	X	1966	Winter
Nautica	W077_NAU	X	X	1977	Winter
Norman	W78_Nor_80.1	X	X	1981	Winter
Oakley	W079_OAK	X	X	2007	Winter
Obelisk	W80_Obe_80.1	X	X	1985	Winter
Orlando	W81_Orl_80.1	X	X	1972	Winter
Palur	W82_Pal_80.1	X	X	1986	Winter
Paragon	W083_PAR	X	X	1999	Winter

4.1c

Accession name	Sequence name	KWS 2011	KWS 2012	Date	Type
Perlo	W084_PER	X	X	1979	Winter
Piko	W119_Pik_80.1	X	X	1994	Winter
Rabe	W085_RAB	X	X	1972	Winter
Recital	W086_REC	X	X	1986	Winter
Renanbansa	W117_Ren_80.1	-	X	-	-
Rialto	W88_Ria_80.1	X	X	1993	Winter
Riband	W89_Rib_80.1	X	X	1988	Winter
Rimpaus braun	W090_RimB_80.1	X	X	1939	Winter
Savannah	W094_SAV	X	X	1990	Winter
Shamrock	W123_Sha_80.1	-	X	-	Winter
Shango	W76_Sha_80.1	X		1994	Winter
Slejpner	W096_SLE	X	X	1986	Winter
Soissons	W97_Soi_80.1	X	X	1992	Winter
Solstice	W098_SOL	X	X	-	Winter
Spark	W099_Spa_80.1	X	X	1991	Winter
Sperber	W100_SPE	X	X	1982	Winter
Stamm	W101_Sta_80.1	X	X	-	-
Starke	W102_Str_80.1	X	X	-	-
Steadfast	W103_STE	X	X	1969	Winter, Spring
Svale	W104_SVA	X	X	1956	Winter
Tadorna	W105_Tad_80.1	X	X	-	Spring
Taras	W106_Tar_80.1	X	X	-	-
Tremie	W108_TRE	-	X	1992	Winter
Trintella	W118_Tri_80.1	X	X	1994	Winter
Tschermaks	W109_Tsc_80.1	X	X		
Vilmorin	W110_Vil_80.1	X	X	1928	Winter
Virgo	W111_VIR	X	X	1968	Winter
Virtue	W112_Vrt_80.1	X	X	1979	Winter
Weebil	W120_Wee_80.1	-	X	-	-
Werla	W113_Wer_80.1	-	X	-	Winter
Xi19	W114_XI19	X	X	2002	Winter
Zebedee	W115_ZEB	-	X	2007	Winter

4.1.1.2 Phenotyping

4.1.1.2.1 Screening a subset of wheat accessions for stem lodging risk

As previously discussed, lodging is a complex trait thought to be the result of either stem breakage or anchorage failure. It may be also be a combination of the two. Lodging is thought to be induced by wind and heavy rainfall. The resistance of a plant to lodging, is likely to be the result of many contributing factors, including plant height, root anchorage strength and stem mechanical strength (all of which are complex traits). Given this complexity, it is difficult to assess the contribution of a single factor, in this case stem strength, by simply scoring the lodging observed in the field. To assess which accessions were more prone to stem lodging, a field-based "stem lodging risk" screen was conducted. This experiment was carried out prior harvesting plants grown as part of the KWS 2012 field trial.

The setup used was similar to that described in Section 3.1.1.2 for *B. napus*. However, in wheat, given greater plant replication and lower stem mechanical strength, it was possible to obtain a score of resistance to stem breakage using the pulley system illustrated in Figure 4.1. This screen was carried out on a subset of 65 accessions across the field (6 plants were tested for each accession). The pulley cord was attached to the highest point of the main stem (just below the ear) using a clip. Each stem was then pulled to ground level through a reproducible arc. Following this, any stem mechanical failure induced by the pulling of the stem was recorded. Stems found to suffer stem breakage were scored with a "1" and those for which no mechanical failure could be detected were scored as "0". To avoid the problems caused by "edge effects" (where plants growing at the edge of a plot may not be "typical" due to having more space etc. or may have been more exposed to the effects of wind and rainfall), this experiment was conducted on within-plot plants (requiring the removal of plants growing at the nearest edge).

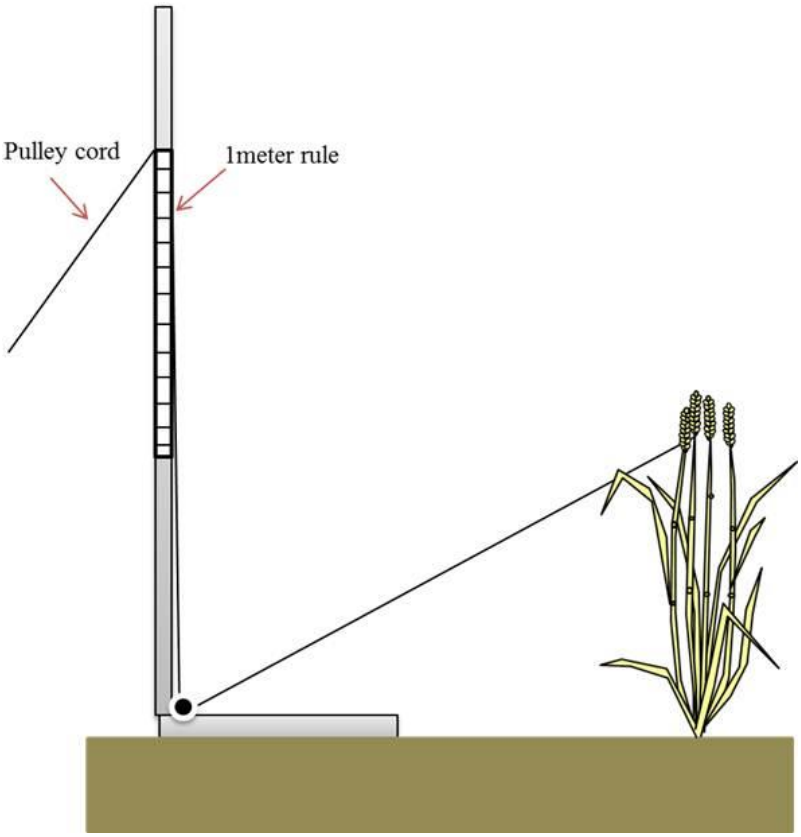

Figure 4. 1. Apparatus used in screening a subset of wheat accessions for stem lodging risk under field conditions.

4.1.1.2.2 Morphological trait measurements

Following harvest and the complete drying of the plant material, several morphological measurements were taken manually from the main stem (determined as the tallest) of each wheat plant. The following measurements were recorded:

1) Plant height (measured from the stem base to the top of ear)
2) Pre-threshed weight
3) Post-threshed weight
4) Length of second internode (counted from the base of the stem)

The subtraction of measurement 3 from measurement 2 above, allowed for a measure of main stem grain weight to be obtained. While this does not represent an estimate of overall grain yield, it is adequate to allow for the direct relationship between grain weight and the mechanical strength of the main stem to be explored. The above measurements can also be seen illustrated in Figure 4.2.

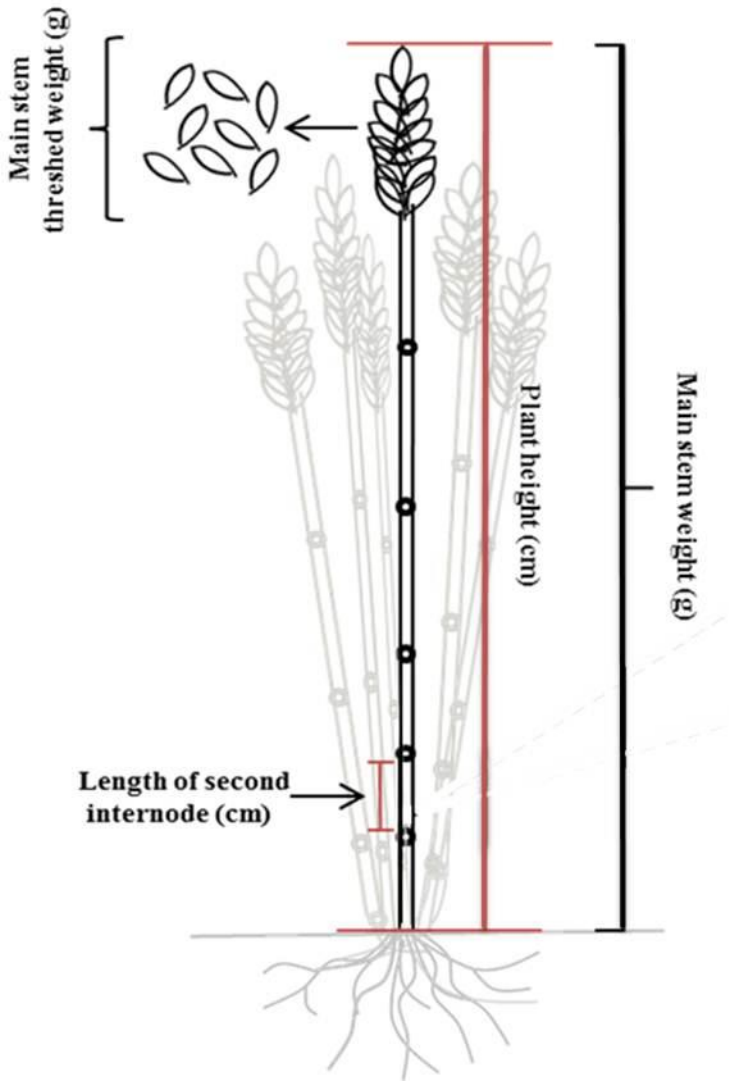

Figure 4. 2. Morphological measurements taken across 100 wheat accessions.

4.1.1.2.3 Screening wheat accessions for stem cross-sectional properties and preparing samples for mechanical testing

Following these initial morphological trait measurements, a 5 cm section was taken from the second internode from the base of the stem. Using a permanent marker pen, the lower end of the internode section was marked. The transverse section of the marked end of each stem was then photographed to allow for the collection of stem-cross sectional measurements (see Section 2.2.2 for a description of the cross-sectional measure taken). To allow for any variation in moisture content between samples to be accounted for, all samples were stored within a silica gel tank at 55 %RH. To allow for the amount of time required for the moisture content of the stem samples to equilibrate with that of the silica tank, a small pilot study was carried out

as described in Section 2.1. This pilot study revealed that two days within the silica tank was an adequate length of time (Figure 4.3). For this reason, prior to mechanical testing, all wheat stem samples were stored for a minimum of two days at 55 % RH.

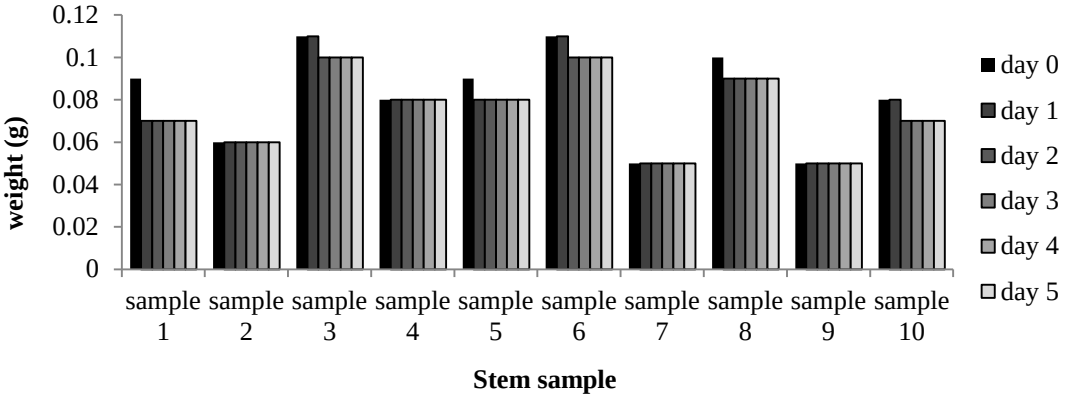

Figure 4. 3. Results obtained from a pilot study assessing the required length of time for wheat stem samples, stored within a silica tank, to reach 55%RH.

4.1.1.2.4 Screening 100 wheat accessions for variation in stem mechanical strength using a three-point bend test.

Mechanical testing was carried out on the 5 cm stem sections using a Texture Analyser with three-point bend test setup. A description of this method can be found in Section 2.2.1. The Texture Analyser was fitted with a 5 kg load cell and the support stands were set to be 2.5 cm apart. The deformation probe was set at a start height of 40 cm and was set to descend 1.5 cm following contact with the stem sample. Following mechanical testing, the mechanical data obtained, along with the stem cross-sectional measures collected, were used in calculating stem absolute, material and structural strength measures. A description of these measures and information regarding all equations used can be found in Section 2.2.2. All calculations were performed using an adapted version of the script which can be found in Supplementary Data file 2.1.

4.1.1.2.5 FT-IR analysis and PLS modelling of wheat straw

In collaboration with the laboratory of Professor Keith Waldron, IFR, chemical composition data were obtained for a subset of wheat accessions.

To assess the level of variation in wheat straw chemical composition Collins et al (accepted), conducted FTmidIR followed by PLS modelling to evaluate carbohydrate and lignin composition across 90 wheat accessions. This method also allowed for the screening of cross-linking phenolics including ferulic and diferulic acids.

FT-IR models were developed based on six glasshouse-grown wheat accessions (Avalon, Cadenza, Paragon, Charger Robigus and Savannah) with four replicate plants per accession. Following the removal of grain tissue, the plants were dissected into four main tissue types: internode, node, leaf and ear. In addition to these six accessions, in 2010, 90 wheat accessions were grown at KWS Thriplow, Rothamsted research and 17 Velcourt-managed field sites. In 2011, these plants were grown for a second year at KWS, Thriplow. All consequent analyses were carried out on whole plant material (minus the grain).

All wheat samples were finely ground using a J&K MF10 analytical sieve mill and a J&K A10 grinder to achieve a fine powder of less than 250 µm. Following mixing, the resulting powder was analysed using FT-IR analysis.

Straw sugar composition was assessed following the hydrolysis of the material using 72 % H_2SO_4. Samples were left in solution for 3 hours with occassional agitation. The solution was then diluted to 1mol L^{-1}, and the hydrolysis repeated for a further 2.5 hours. All hydrolysed sugars were then analysed using GC (Perkin-Elmer Autosystem XL, Massachusetts, USA). Uronic acid levels were determined using colorimetric methods (Blumenkrantz and Asboe-Hansen, 1973). All measurements were carried out in triplicate. Following the hydrolysis of sugar components using sulphuric acid at, a residue remains comprised of lignin, ash and any remaining cell wall proteins.

To assess variation in phenolic acids between accessions, phenolics were extracted using 4 M NaOH and then separated and identified using HPLC.

FTIR-ATR spectra were obtained for the finely milled plant material, with five replicates per sample. This was carried out using a BioRad Fourier Transform

Infrared spectrometer, fitted with a reflection diamond ATR accessory (GoldenGate (Specac), Slough, UK) and an MCT detector. Each sample was loaded on the crystal and compressed using a clamp fixture. This method allowed for a spectrum for each sample to be obtained between 800 and 4000 cm^{-1}. These spectra were then processed using MATLAB V7.14 (MathWorks Inc., Massachusetts, USA). Firstly, the spectra were truncated to 800-1800 cm^{-1}. The absorbance at1800 cm^{-1} was used as the baseline and all spectra were area normalised. Partial Least Squares (PLS) models for each spectral component were developed using the 'plsregress' function in the MATLAB statistics toolbox V8.0 (MathWorks Inc., Massachusetts, USA). PLS models were also generated for total sugar composition as well as for the following individual sugars components: glucose, rhamnose, fucose, arabinose, xylose, mannose, galactose, as well as glucuronic acid. PLS models were also developed for Lignin (with and without the acid insoluble ash correction) and phenolics. The phenolic compounds included were protocatechuic acid, protocatechuic aldehyde, p-OH-benzoic acid, caffeic acid, vanillic acid, p-OH-benzaldehyde, vanillin, trans-p-coumaric acid, sinapic acid, trans-ferulic acid, cis-p-coumaric acid, 8,8' –DiFA, 8,5'-DiFA, cis-ferulic acid, 5,5'-DiFA, 8-O-4'-DiFA, 8,5'-DiFA (benzofuran), and total phenolics).

PLS models were verified and then calibrated using a subset of wheat accessions (including the six aforementioned glasshouse grown accessions). Chemical compositions of the 90 field-grown accessions were predicted data generated by the application of the models to the spectra replicates, then averaged to give a value (Collins et al., accepted).

All chemical analyses and model development was carried out in the lab of Professor Keith Waldron by, Samuel Collins, Klaus Wellner and Isabel Martinez.

4.1.1.4 Data analysis

Following the calculation of the mechanical trait values, Genstat was used to perform a Restricted Maximum Likelihood (REML) analysis. This method allowed for the combining of trait data collected across the two years to obtain predicted means. This method also allowed for the exploration of trait normality. Any traits found to show trait residuals clearly deviating from a normal distribution, where transformed using a $\log_{10}$ transformation. The REML analysis was then repeated using these transformed values. Following the calculation of REML-predicted means, Genstat was further used to conduct correlation analyses, allowing any relationships between the traits included in this study to be determined.

4.1.2 Results

4.1.2.1 Detected variation in stem mechanical properties and related traits across 100 wheat accessions

Although in the majority of cases, trait residuals were seen to follow a normal distribution (an example of which can be seen in Figure 4.4A for the absolute strength trait F/V), skewed residual plots were observed for the following traits: second moment of area, stem outer cortex thickness, MOR and MOE (An example of this can be seen in Figure 4.4B for MOE). However, following a $\log_{10}$ transformation a greater level of normality was achieved (Figure 4.4C). In addition, for many traits (with the exception of stem outer cortex) significant interactions were detected between year and genotype. To allow for the prediction of mean values where such interactions were observed, using REML, Genotype and year were fitted as fixed effects (genotype + year) and the interaction term (genotype.year) was fitted as a random effect. In cases where no interaction was detected, this model was adjusted to include just year as a random term.

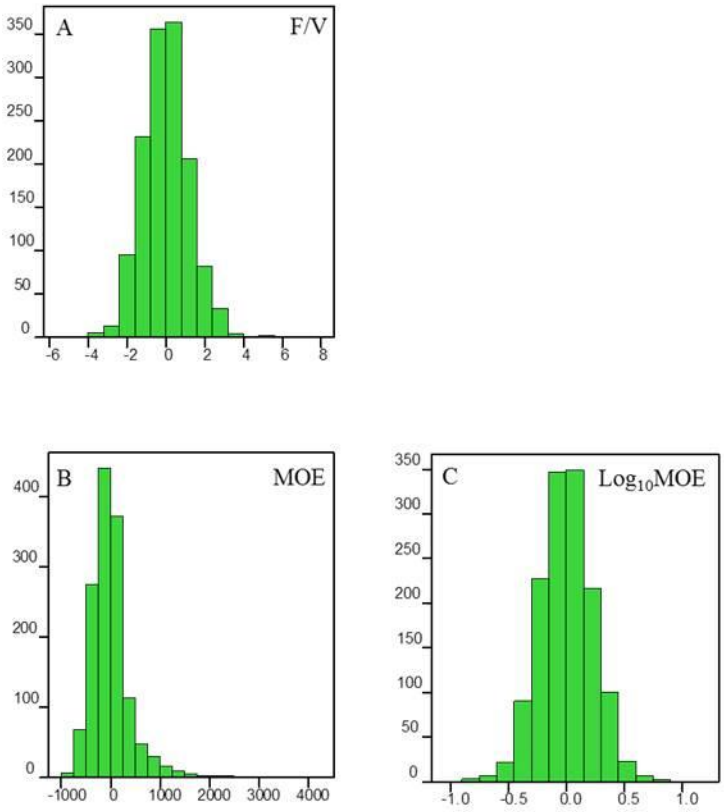

Figure 4. 4. Trait residual plots seen for stem mechanical traits in wheat. F/V is shown here with normal residuals (A). MOE exhibited a skewed distribution (B). Increased normality was achieved following a log_{10} transformation (C).

Using these models, highly significant differences (P<0.001) were detected for the following traits: Fmax; F/V; stem hollow area; main stem threshed weight and plant height. Stem width, main stem grain weight and the length of second internode also varied significantly (P=0.002). Stem parenchyma area and log_{10} outer cortex thickness varied significantly between accessions with respective P values of 0.039 and 0.026. Log_{10} MOR varied significantly between accessions with a P value of <0.001. Finally, log_{10} second moment of area and log_{10} MOE did not vary significantly between accessions with respective P values of 0.51 and 0.32. Although second moment of area was found to not vary significantly between accessions in both transformed and non-transformed form, significant differences between accessions were seen for non-transformed MOE (P<0.001). No genotype by year interaction was detected for this trait. The variation observed for each trait can be seen in Figures 4.5-4.11. The detection of significant variation for the majority of the traits measured, suggests that there is potential for the dissection of these traits

through genetic mapping and consequent improvement through breeding. All trait summary information along with the results obtained from REML analyses are presented in Supplementary Data files 4.1a and b.

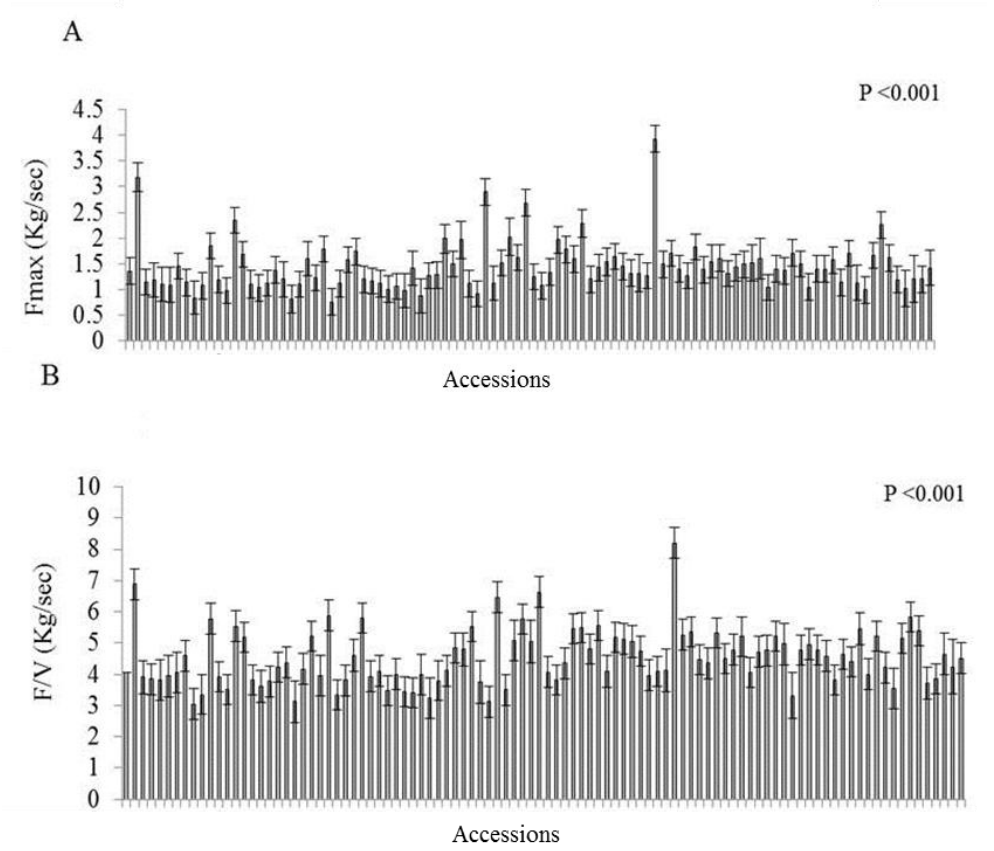

Figure 4. 5. Variation in Fmax (A) and F/V (B) across 100 wheat accessions grown across two years at KWS, Thriplow. Error bars represent the standard error of the mean as calculated through REML.

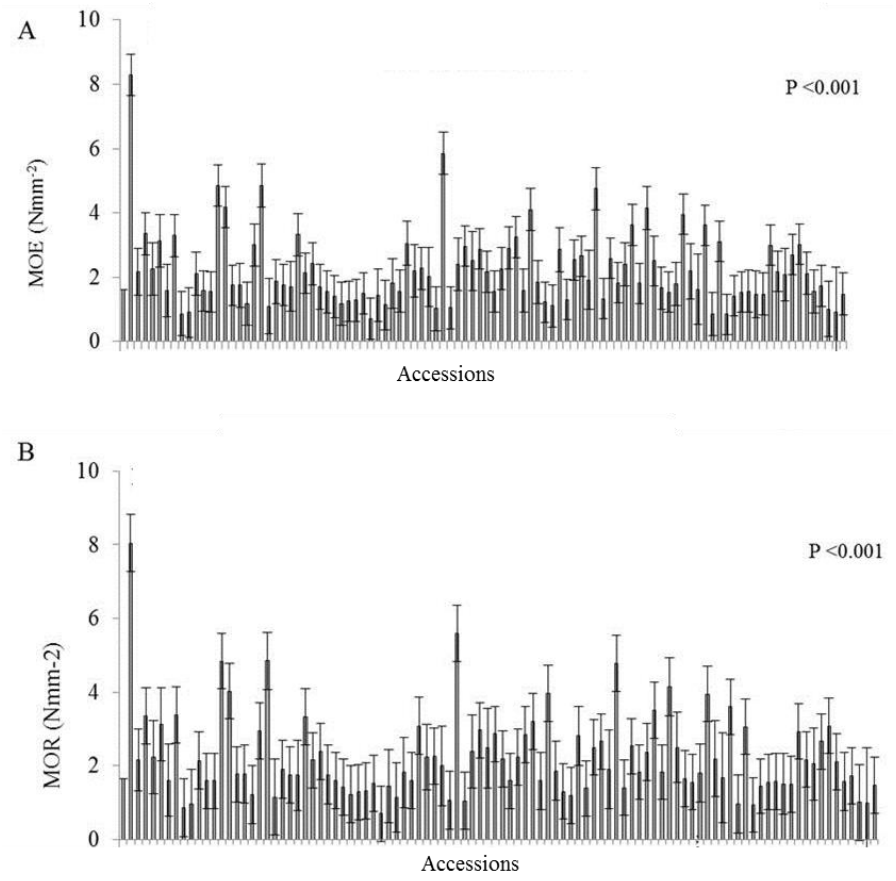

Figure 4. 6.Variation in MOE (A) and MOR (B) across 100 wheat accessions grown across two years at KWS, Thriplow. Error bars represent the standard error of the mean as calculated through REML.

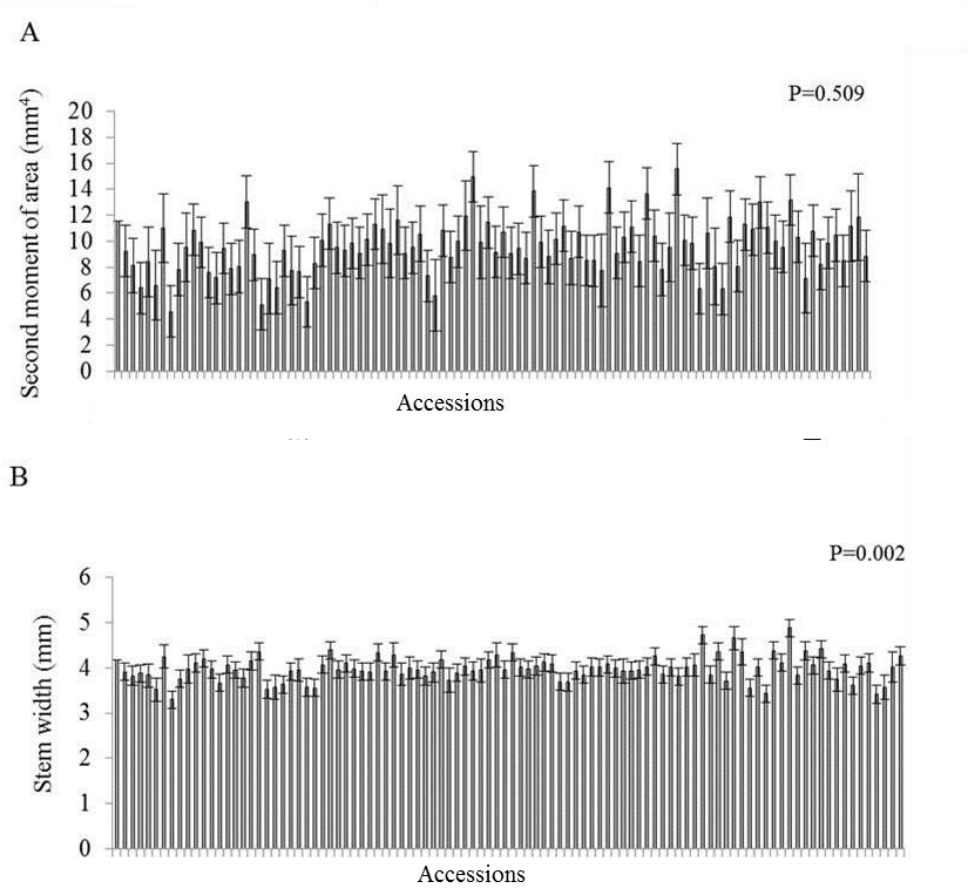

Figure 4. 7.Variation in second moment of area (A) and stem diameter (B) across 100 wheat accessions grown across two years at KWS, Thriplow. Error bars represent the standard error of the mean as calculated through REML.

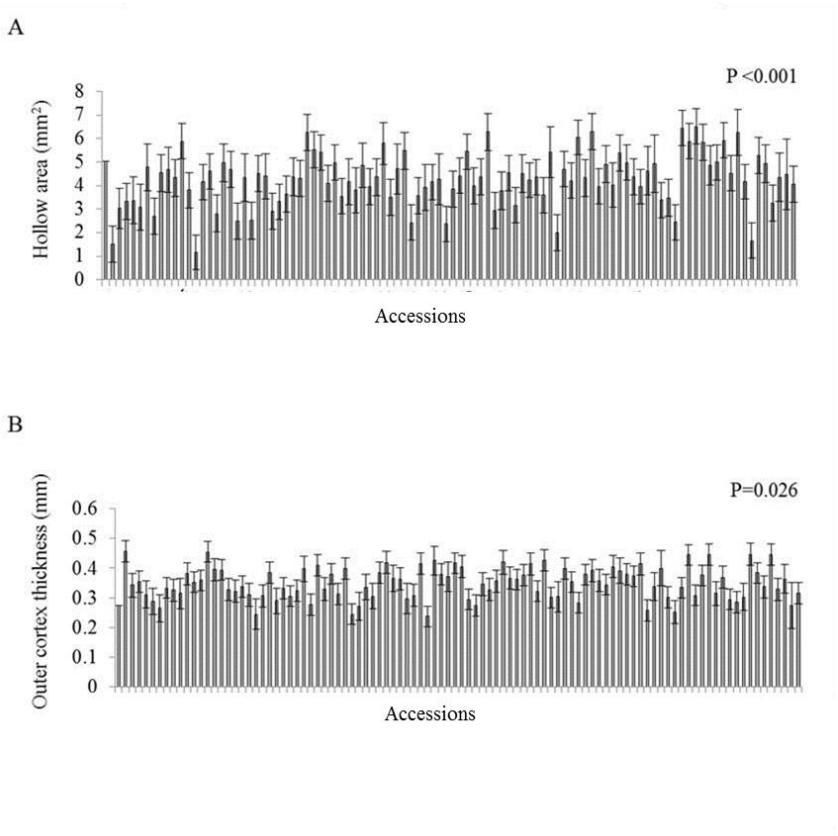

Figure 4. 8. Variation in hollow area (A) and outer cortex thickness (B) across 100 wheat accessions grown across two years at KWS, Thriplow. Error bars represent the standard error of the mean as calculated through REML.

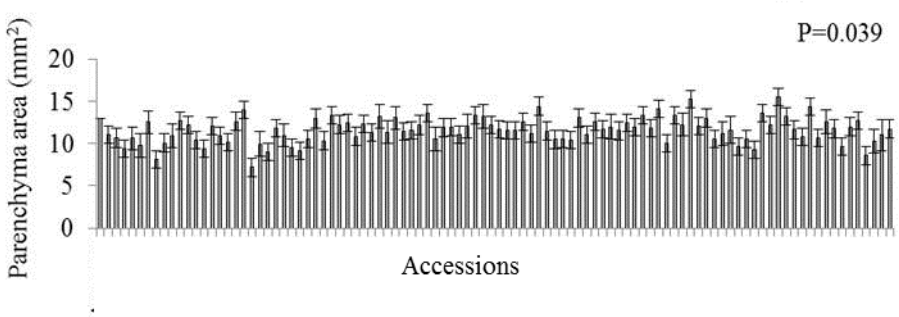

Figure 4. 9. Variation in parenchyma area across 100 wheat accessions grown across two years at KWS, Thriplow. Error bars represent the standard error of the mean as calculated through REML.

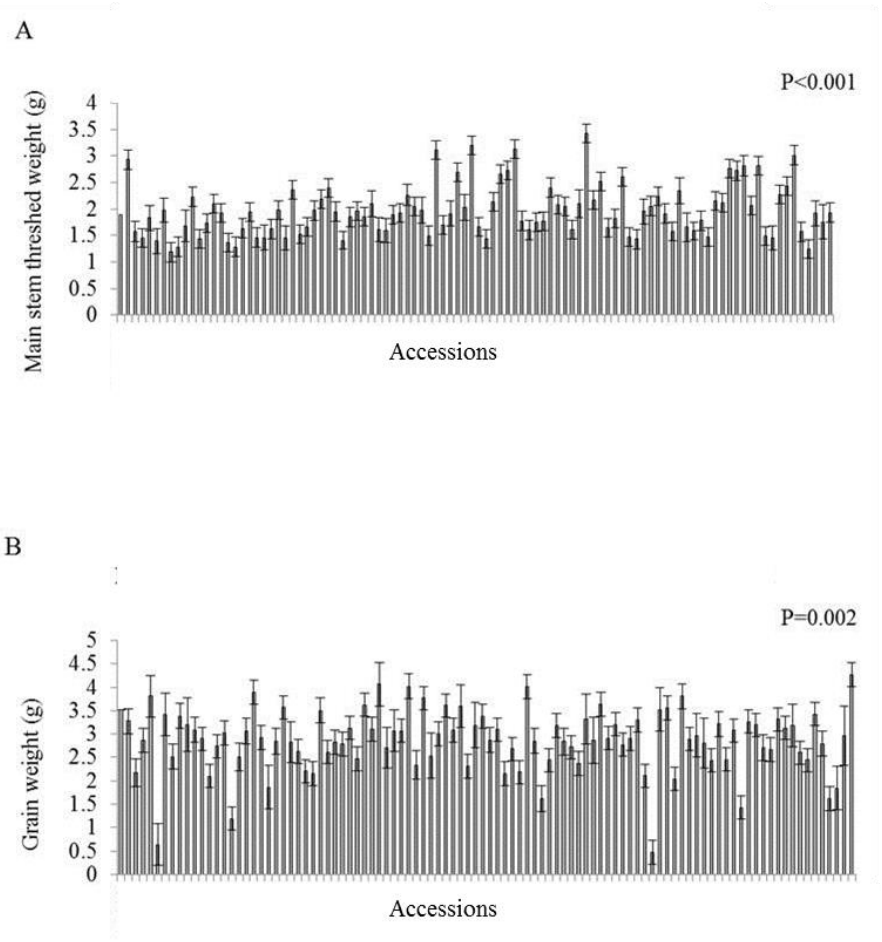

Figure 4. 10. Variation in main stem threshed weight(A) and grain weight (B) across 100 wheat accessions grown across two years at KWS, Thriplow. Error bars represent the standard error of the mean as calculated through REML.

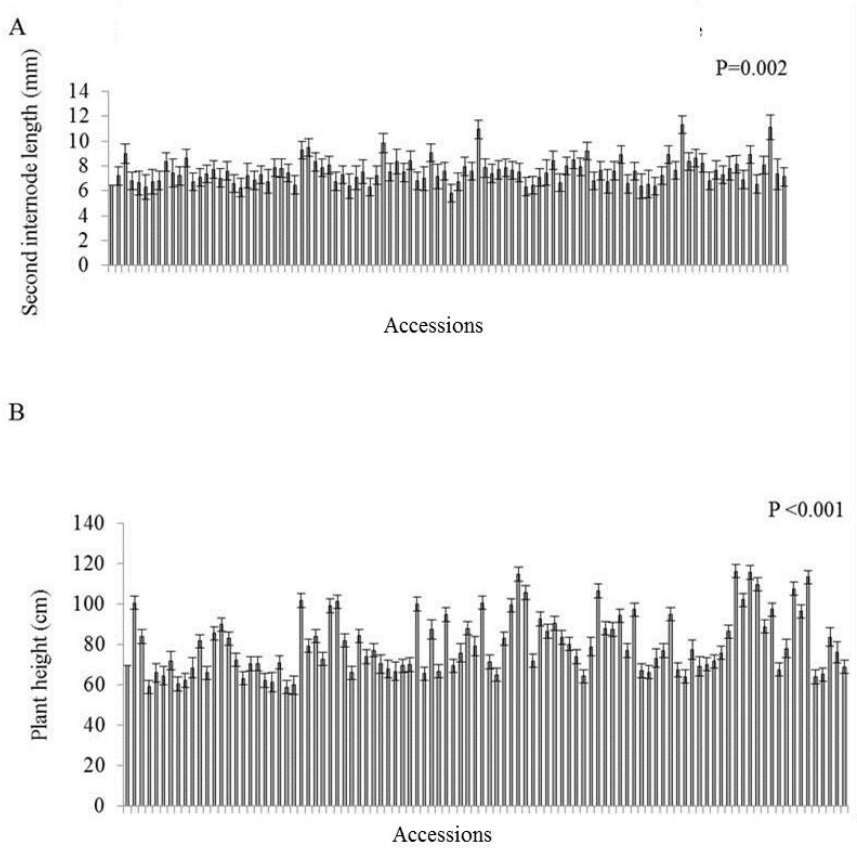

Figure 4. 11. Variation in second internode length (A) and plant height (B) across 100 wheat accessions grown across two years at KWS, Thriplow. Error bars represent the standard error of the mean as calculated through REML.

4.1.2.3 The relationship between stem structural and morphological characteristics and stem mechanical strength in wheat

To explore the relationships between the mechanical, morphological and structural traits measured, a correlation analysis was carried out (Table 4.2). No significant relationship was detected between the absolute strength measures and stem width. Weak but statistically significant positive correlations were however detected between the absolute strength traits Fmax and F/V and the stem structural strength trait, second moment of area giving R^2 values of 0.08 (P=0.004) and 0.06 (P=0.009) for Fmax and F/V respectively. Significant positive correlations were also detected between the absolute strength measures and the thickness of stem outer cortex and plant height giving R^2 values of 0.23 (P<0.001) and 0.15 (P<0.001) for Fmax and 0.16 (P<0.001) and 0.23 (P<0.001) for F/V respectively. Stem hollow area was found to show a significant negative correlation with stem absolute strength giving an R^2 of 0.18 (P<0.001) for Fmax and 0.13 (P<0.001) for F/V. The strongest

correlations however were seen between the main stem threshed weight and stem absolute strength giving R^2 values of 0.42 ($P<0.001$) and 0.47 ($P<0.001$) for Fmax and F/V respectively. These traits correlate positively with absolute strength. A weak but significant positive correlation was also detected between grain weight and the absolute strength measure, giving an R^2 values of 0.08 ($P=0.003$) and 0.04 ($P=0.04$) for Fmax and F/V respectively. Further significant positive correlations were detected between the absolute strength traits and area of stem parenchyma tissue (R^2= 0.29 ($P<0.001$) for Fmax and 0.18 ($P<0.001$) for F/V). No significant correlations were detected between stem strength and the length of second internode. The relationship between F/V and second internode length is however only marginally below the 5 % significance threshold ($P=0.29$ and 0.051 for Fmax and F/V respectively).

Significant correlations were also detected for the material strength traits, MOE and MOR. As expected (due to the auto-correlative nature of their derivation), significant negative correlations were seen between the material strength traits and the stem structural traits such as the stem hollow area and stem width and second moment of area. Positive correlations were detected between the stem parenchyma tissue area (R^2=0.15, $P<0.001$); outer cortex thickness (R^2=0.12, $P<0.001$) and main stem threshed weight (R^2=0.09, $P<0.01$) and MOR. In the case of the stem parenchyma tissue, this may be the result of the inverse relationship observed between stem parenchyma tissue and stem hollow area, the latter of which is negatively auto correlated with MOR. Outer cortex thickness and main stem threshed weight however, do not show a significant negative correlation with any of the other structural traits. It is therefore possible that these are true correlation i.e. the material comprising the outer cortex for example, may truly contribute to overall material strength of the stem.

Finally, grain weight and plant height as previously mentioned, correlate significantly with the absolute strength traits. They do however not show significant correlations with the material strength traits, suggesting that the relationship with absolute strength is structural, for example the second moment of area or stem width.

	Fmax	F/V	Log_{10} MOR	Log_{10} MOE	Log_{10} second moment of area	Stem width	Hollow area	Parenchyma area	Log_{10} outer cortex thickness	Length of second internode	Threshed weight	Grain weight	Plant height
Fmax	-												
F/V	***0.84	-											
Log_{10} MOR	***0.56	***0.46	-										
Log_{10} MOE	***0.12	***0.21	***0.51	-									
Log_{10} second moment of area	**0.08	**0.06	**-0.09	***-0.52	-								
Stem width	0	0	**-0.06	***-0.24	***0.34	-							
Hollow area	***-0.18	***-0.13	***-0.38	***-0.33	***0.12	***0.28	-						
Parenchyma area	***0.29	***0.18	***0.15	-0	***0.10	***0.12	***-0.11	-					
Log_{10} outer cortex thickness	***0.23	***0.16	***0.12	0.02	0.02	0.02	-0.01	0	-				
Length of second internode	0.01	0.04	-0	-0	0.03	-0	***0.11	-0.07	0	-			
Threshed weight	***0.42	***0.47	**0.09	0	***0.2	***0.16	0.01	***0.14	**0.07	***0.12	-		
Grain weight	**0.08	*0.04	0	-0.03	***0.12	***0.13	0	***0.14	0	*-0.05	**0.08	-	
Plant height	***0.15	***0.23	0.02	0	**0.09	0.04	*0.05	-0	*0.05	***0.42	***0.60	-0	-

Table 4. 2. Correlation matrix for all stem mechanical, structural and morphological traits analysed in wheat. The correlations are given here as R^2 values. The asterisks represent the statistical significance of each correlation where * represents a significance level of < 0.05; ** represents a significance level of < 0.01 and * represents a significance level of < 0.001. Negative correlations can be seen marked with a "-".**

4.1.2.2 The importance of stem mechanical strength for stem lodging risk in wheat

To obtain a measure of wheat standing ability under field conditions, a lodging experiment was carried out as described in Section 4.1.1.2.1. The results obtained from this experiment, revealed significant variation in stem lodging risk between the 65 accessions tested (P=0.02).

When exploring the relationship between these stem lodging risk results and the results obtained through the three-point bend test, some very encouraging relationships were observed (Figure 4.1.2).

Negative correlations were observed between the lodging risk score and both absolute and material strength. Fmax and MOR showed the most significant correlations with respective R^2 values of 0.27 (P<0.001) and 0.39 (P<0.001). This suggests that material strength may be of most importance here. Furthermore, no correlation was detected between the lodging risk score and plant height (despite height varying from 60.4 cm to 121.8 cm across accessions) (R^2=0.02 with a P value of 0.21). This suggests that this measure of stem lodging was not confounded by the variation in height between the accessions scored (an important observation given that each plant was tested at the base of the ear and not at a standard height for each sample).

These findings suggest that the three-point bend test method employed is effective in identifying important variation relevant to stem mechanical strength under field conditions. Trait summary data and the results obtained from ANOVA are presented in Supplementary Data file 4.1c.

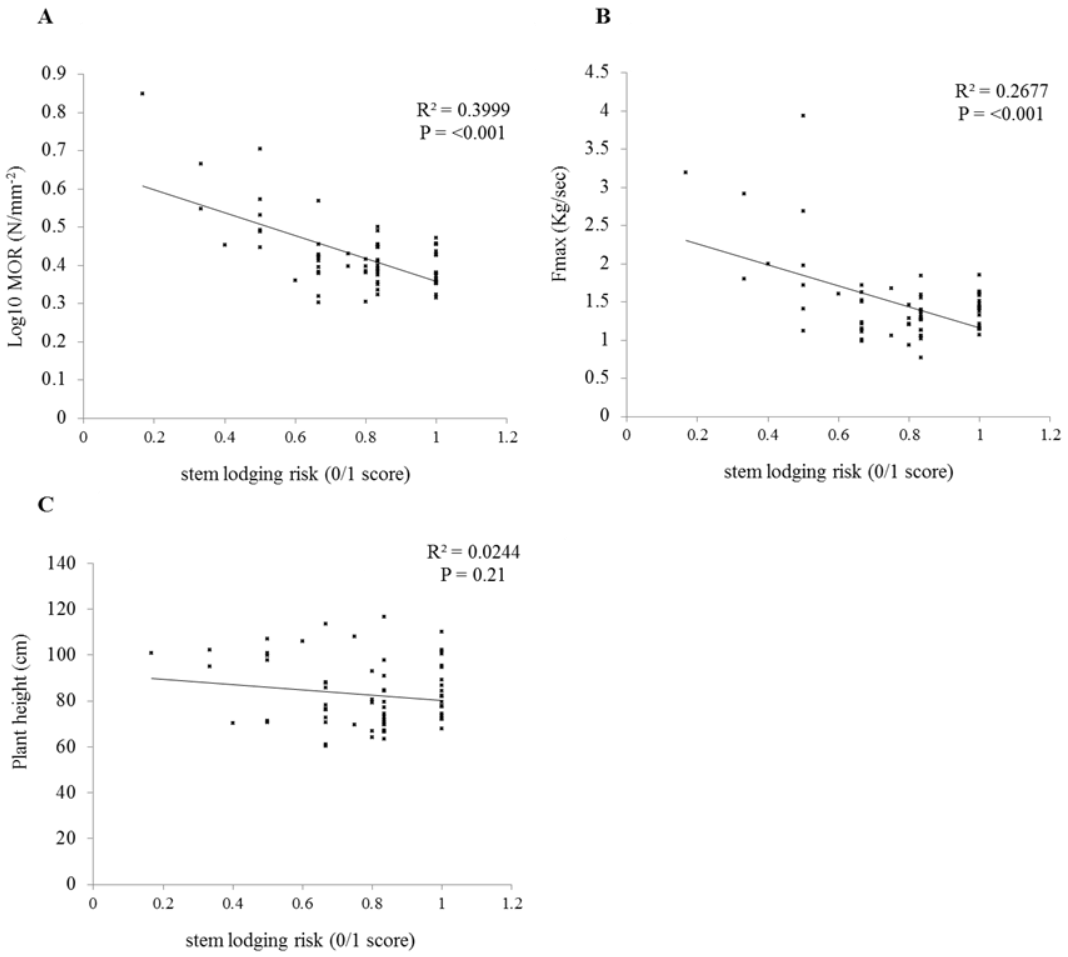

Figure 4. 12. The relationship between stem lodging risk and MOR (A), Fmax (B) and plant height (C) across 65 wheat accessions. R^2 and P values were calculated using the correlation analysis in Genstat.

4.1.2.4 The relationship between chemical composition and stem mechanical strength in wheat

In addition to the traits so far described, in collaboration with the IFR, data were obtained, describing variation in the chemical composition of the wheat straw used in the present study. These data were collected as part of the work carried out in the laboratory of Professor Keith Waldron, IFR. Using Fourier Transform Infrared (FT-IR) spectroscopy on finely milled plant material, measures of cell wall components were obtained. These included measures of several cell wall sugars including rhamnose, fucose, arabinose, xylose, mannose, galactose and glucose. In addition, functionally important phenolic compounds found in the plant cell wall were also assayed including ferulic acid, coumeric acid and lignin. In addition, a small subset of accessions, were included in an analysis looking at specific tissue types. This

allowed for the variation in chemical composition of different anatomical structures to be determined. The tissue types included in this further analysis were internode, node, leaf and ear. Studies have reported roles of both cell wall sugars and phenolics in determining stem mechanical strength (Jones et al., 2001; Wang et al., 2006) making these data very valuable to the present study. To explore any potential relationships between these traits, a further correlation analyses was conducted. This was carried out including only those traits found to correlate significantly with mechanical strength in the previously described correlation analysis. The FT-IR analysis was carried out on 86 of the 100 wheat accessions for which mechanical trait data was collected. This correlation analyses was therefore based on data collected for this subset of 86 accessions

The results from this analysis revealed several interesting findings (see Supplementary Data file 4.1d for correlation matrix). Firstly, Fmax and F/V were found to correlate significantly to a number of chemical components. For example, both Fmax and F/V correlate positively with glucose (R^2=0.11, P=0.002 for Fmax and R^2=0.18, P<0.001 for F/V). Fmax ad F/V show a negative relationship with P-OH-phenyl acetic acid (R^2 = 0.13, P<0.001 for Fmax and R2=0.18, P<0.001 for F/V).These mechanical strength traits also correlate negatively with galactose (R^2=0.20, P<0.001 for Fmax and R^2=0.25, P<0.001 for F/V). Significant negative correlations were also seen for arabinose where both Fmax and F/V were found to have an R^2 of 0.19 (P<0.001).

In contrast, very few significant correlations were detected between the material strength measures and chemical composition. Very weak positive correlation were seen between MOR and P-OH-phenylacetic acid (R^2 =0.06, P=0.0215). This material strength trait also shows a weak negative with truxillic acid FA (R^2=0.05, P=0.0315) and galacturonic acid (R^2= 0.03, P=0.047).

The finding that very few significant correlations were detected for material strength, suggests that the correlations are the result of some structural or morphological factor(s) which contributes to/is correlated with absolute strength. Indeed similar trends were also seen between these chemical components and plant height, second moment of area, outer cortex thickness and main stem weight, all of which positively correlate with absolute strength. For galactose, plant height and main stem weight

were found to have respective R^2 values of 0.498 (P<0.001) and 0.34 (P<0.001). In addition, for this same chemical component, second moment of area was found to have an R^2 of 0.12 (P<0.001). For glucose, plant height and main stem threshed weight were found to have R^2 values of 0.39 (P<0.001) and 0.22 (P<0.001) respectively. Outer cortex thickness was also correlates positively (although the correction is weak) with glucose (R^2=0.08, P=0.009). This finding suggests that the correlations seen between these traits and the chemical components are related to variation in various plant morphological and structural traits. The various tissue types making up the wheat stem are known to have considerably different chemical composition (Collins et al., accepted). The stem samples which were analysed through mechanical testing exhibited a high level of variation in structural traits such as outer cortex thickness and parenchyma area (both of which correlate positively with absolute strength). These tissue types are likely to vary significantly in chemical composition. It is therefore likely that the correlations observed between chemical composition and absolute strength reflect this, as opposed to the chemical components themselves explaining the variation in absolute strength.

4.1.3 Discussion

Through analysing the trait data collected across two years of field trials through REML analysis, a high level variation has been detected between accessions for many of the traits included in the present study. This suggests that there is potential for the improvement of stem mechanical strength in wheat through breeding. Furthermore, through analysing the relationships between stem mechanical strength and plant morphological and stem structural traits, several traits important in contributing to the overall variation in stem strength observed across the panel of 100 wheat accessions have been identified. The results presented suggest that the weight of the stem tissue is highly related to the overall strength of the stem. In part, this seems to be related to stem height. However, given that the stem weight traits show a much stronger correlation with stem strength than plant height, it may also be due to other factors such as stem structure or density. Indeed, several stem structural traits were seen to correlate positively with threshed weight, including outer cortex thickness and stem parenchyma area, suggesting that the size/density of these stem components are contributing significantly to the overall stem weight.

The results also suggest that stem solidity may be a good target for the breeding of wheat accessions with improved stem strength. This trait may be of further interest given that studies have reported that an increase in stem solidity can provide resistance to damage caused by stem sawfly by impeding the growth of sawfly larvae and reducing the ease of migration throughout the plant body (Kong et al., 2013). It may therefore be possible to select for increased Sawfly resistance and increased stem strength with this stem structural trait. Stem parenchyma area was also seen to correlate positively with absolute strength. Stem parenchyma in wheat is a significantly more densely packed within the stem in comparison to other plant species such as *B. napus.* It is therefore not surprising that a correlation with mechanical strength has been detected. However, the relationship observed may also be indicative of autocorrelation. Indeed, the area of stem parenchyma tissue was found to correlate with additional structural traits also found to be related to strength. An example of this can be seen with the negative correlation between parenchyma area and stem hollow area (a trait found to negatively correlate with strength).

Another potentially important relationship detected in this analysis, was that between plant height and absolute strength. Although this correlation may represent an artefact of the relationship observed between stem threshed weight and height, it would not be surprising that a taller plant would require higher levels of mechanical support given the increase in bending moment imposed by greater stature. As previously discussed, reduced plant height has been a key target trait for the development of lodging resistant wheat accessions (Borojevic and Borojevic, 2005; Pavlista et al., 2010). An ideotype for lodging resistance may be a crop which is short in stature and which has high stem mechanical strength. However, the results presented here suggest that height and stem strength are positively related, and therefore the development of this ideotype crop for lodging resistance may be complex. However, through genetic mapping, the power to uncouple these important target traits should be revealed, and given that the highest R^2 value seen between these traits was 0.26 for F/V, it seems encouraging that this will be possible. This finding illustrates the importance of studying multiple, potentially interrelated traits within a single study.

Unlike previous reports in the literature (Kelbert et al., 2004; Hai et al., 2005), no significant relationships were detected between stem mechanical strength and

internode length or stem diameter. There is therefore no evidence to suggest that these traits should be used as morphological markers as a proxy for stem strength in wheat, or should be explored further through genetic mapping.

Interestingly, although clearly an important measure of structural strength in some species (as seen with the work carried out for *B. napus)*, second moment of area showed little correlation with absolute strength in wheat. We have however seen that other structural traits, such as stem outer cortex thickness and parenchyma area do correlate with absolute strength. However, the combining of these seemingly important traits into a single measure of structural strength, the second moment of area, results in these effects being diminished. This may be due to the complex interaction between these component traits. Indeed, the correlation analysis revealed a negative relationship between stem hollow are and stem parenchyma area and that these traits each contribute to absolute strength with opposite trends. Furthermore, while both stem parenchyma area and stem outer cortex both positively contributes to stem strength, they show no correlation with each other. However, regardless of the cause, this analysis has revealed that stem second moment of area is not a suitable measure of stem structural strength in wheat (or at least within the 100 accessions included in his study). This has however not hindered the present study, given that many additional structural traits were explored and their importance in determining stem strength revealed. It must however be realised, that given the use of the second moment of area in calculating the material strength traits, MOE and MOR, it is likely that the material and structural counterparts of the absolute strength traits, Fmax and F/V have not been entirely uncoupled. This, although not necessarily a great problem, must be considered in all further analyses and discussion.

As described, a number of expected autocorrelations were detected between the stem structural traits and the material strength traits, MOE and MOR. However, as previously mentioned, potentially important observations were also made. For example, the data presented suggests that the outer cortex thickness and the main stem threshed weight may be important contributors to material strength. It must of course, given the above comment, be considered that this could be the result of unaccounted for structural strength in these material strength measures.

One particularly promising finding obtained from the in depth analysis of these mechanical trait data, was that the strength measures obtained through the three point bend test analysis correlate negatively to the lodging risk measures obtained in the field. The three-point bend test method allows for a very accurate measure of stem strength to be obtained, in the case of the present study, for a single lower internode. However, in the field environment, the mechanical forces induced by heavy rainfall and wind for example, impose stress on not just a single internode, but the whole plant body. It is therefore important to see how the results obtained in vitro, translate to the field environment and to assess whether the mechanical strength of a single internode, can reveal variation relevant for the mechanical support of the entire plant body. The results obtained from the pulley-based lodging risk experiment, suggest that the mechanical trait information collected through the employment of the three-point bend test has relevance to what is observed in the field.

Finally, the results obtained from the FT-IR analysis provided potentially very important insights for the importance of chemical composition in determining stem mechanical strength in wheat. As previously described, significant relationships were detected between the absolute strength measures and various chemical components. Very few significant correlations however were detected for the material strength measures. As previously mentioned, this suggests that the significant correlations observed are the result of stem structural or morphological traits which correlate with stem absolute strength (suggesting that, at least to a certain extent, the material and structural strength components have been successfully uncoupled). The lack of correlation detected between the stem material strength traits and chemical composition may seem surprising given the plethora of studies which have reported links between these traits. It is possible that this lack of correlation is due to the material on which the FT-IR analysis was performed. The chemical analysis was carried out on threshed plant material (including leaf material). It is therefore possible that any relationship between the chemical composition of the wheat stem and the stem mechanical strength have been "diluted" by the inclusion of other likely irrelevant tissue types in the analysis. Having said this, the lack of correlation may also be indicative of a potentially important biological explanation. It may for example suggest that while the relative amounts of the chemical components do not significantly contribute to stem mechanical strength in wheat, it may be the way in

which these components are incorporated and held together within the plant cell wall that is of most relevance. For example, studies suggest that the level of pectin methylesterification and xylan acetylation may be important in determining structural integrity (Pawar et al., 2013). These more fine-scale cell wall characteristics were not screened in the chemistry analyses conducted. The lack of correlations observed may therefore provide clues regarding the potential candidate genes involved in determining the variation observed for these mechanical traits.

4.2 Uncovering the genetic control of stem mechanical strength in wheat through Associative Transcriptomics

The use of genetic markers in crop improvement provides a way through which traits of agronomic importance may be selected for with increased efficiency (Collard and Mackill, 2008). Previous studies focusing on the identification of markers for stem strength in wheat have utilised Quantitative Trait Loci (QTL) analysis. Although a widely used technique, as previously discussed, QTL analysis has limitations. Associative Transcriptomics is a novel approach for marker detection (Harper et al., 2012) which incorporates high levels of allelic variation and SNP density through the utilisation of a diversity panel in place of a bi-parental mapping population. In addition, through the use of transcriptome sequencing, as opposed to genomic, this method allows for variation to be explored at both the sequence and gene expression level. This method has only very recently been employed in wheat, where it proved effective in identifying markers for biomass traits including plant height (Harper et al., Under review). This section describes the use of Associative Transcriptomics to uncover genetic markers which may be utilised in the marker-assisted breeding of wheat accessions with improved stem mechanical strength. All absolute and material strength measures described in the previous section were explored using this method. In addition, all traits found to show significant variation between accessions and to correlate significantly with these mechanical traits were also included in the analysis. As explained in Section 4.1.2 a number of the traits included in this study were found to have non-normal residuals. It is commonly advised, that when using linear models, trait variances should follow a normal distribution. However, a number of studies have shown that linear models can be used robustly in modelling non-normal data (Koerhuis and Hill, 1996; Ma and Mazumdar, 2011). Furthermore, studies have raised concern regarding the use of data transformation in association studies, where

for example, it may result in the concealment of important biological information (Shen and Ronnegard, 2013). There are several examples of association studies which have been successfully implemented on trait data which deviates from normality. An example of this can be seen with the aforementioned Associative Transcriptomics analysis carried to assess seed quality traits in *B. napus.* The data used in this study were of bimodal nature and yet the methods employed proved robust in the detection of causal loci (Harper et al., 2012).To explore the potential effects of trait normality in Associative Transcriptomics, the following analysis will (in cases where non-normal residuals were identified) include both transformed and non-transformed data. This will allow for any differences between the results obtained to be assessed and for the importance of trait normality in Associative Transcriptomics to be evaluated.

4.2.1 Methods

4.2.1.1 mRNA-seq and marker detection in wheat

Transcriptome sequence data was obtained for each of the 100 wheat accessions included in the present study. This was achieved using Illumina transcriptome sequencing technology. The sequencing of these accessions was split between two sequencing platforms, with 49 accessions being sequenced using the GAII platform and the remaining 51 sequenced using the Hi-seq platform**.** The alignment of these reads for SNP detection was facilitated by the development of a reference sequence. This reference sequence was generated based on transcriptome sequence of *Triticum urartu*, *Aegilops speltoides* and *Aegilops tauschii* (representing the A, B and D genomes, respectively). The B genome was further improved by "curing" (Higgins et al., 2012) using sequence information from *T. trugidum*, which more closely represents the B genome in hexaploid wheat (Harper et al., Under review).The 100-base paired-end reads from mRNA-seq were assembled into sets of unigenes using the Trinity assembly package (Grabherr et al., 2011). This resulted in a reference transcriptome sequence comprising 105,069, 132,363 and 85,296 unigenes for the A, B and D genomes, respectively. Based on linkage map information and conserved synteny between wheat and *Brachypodium distachyon,* this unigene set was arranged into a hypothetical gene order. This ordered unigene set, or pseudomolecule, may be used in place of a wheat genome sequence (See Supplementary Data file 4.2e for wheat pseudomolecule (Harper et al., Under review)). Based on sequence similarity

to Brachypodium, rice, sorghum and Maize, this pseudomolecule has, where possible, been annotated with probable gene functions. The A, B and D reference unigenes were different enough to enable reads to be assigned in a genome-specific manner. The alignment of the transcriptome sequencing data to this reference sequence enabled the detection of 42,066 SNP markers (see Supplementary Data file 4.2b for SNP data). This alignment was carried out using MAQ. In addition, quantification of sequence read depths (as reads per kb per million aligned reads; RPKM) provided a measure of expression for each unigene. In total, 94,060 transcripts were identified with non-zero transcription (defined as average read depth across accessions of >0.4). This provided the information required to explore any relationships between gene expression and the trait of interest in what has been termed a Gene Expression Marker analysis, or GEM analysis (Harper et al., 2012). A detailed description of the methods used in performing Associative Transcriptomics in wheat can be found in Harper et al (Under review). The transcript abundance data used in the GEM analyses can be found in Supplementary Data file 4.2c.

4.2.1.2 Accounting for population structure and relatedness and performing Associative Transcriptomics in hexaploid wheat.

To reduce potential error caused by sequencing inaccuracies, Associative Transcriptomics was carried out with minor alleles removed (alleles present as a frequency <5%). This results in the dataset consisting of 42,066 SNP markers being reduced to 12,456. This filtered dataset was then used in constructing a kinship matrix. In addition, using STRUCTURE analysis, two main populations (K) between which the genome ancestry of the 100 wheat accessions could be apportioned, were identified. The methods used in performing STRUCTURE analysis and the construction of the kinship matrix can be found in Section 2.3.1. Figure 4.13A shows a graphical representation of the output obtained from STRUCTURE. Figure 4.13B shows the results obtained following the calculations described by Evanno et al (2005) (see Supplementary file 4.2a for full calculation). As can be seen, the highest delta K value was obtained when K=2 (Harper et al., Under review).The output of this analysis was used as a Q matrix for the subsequent Associative Transcriptomics analyses. STRUCTURE analysis was carried out by Dr A. Harper.

Together, the kinship matrix and the Q matrix were used in Associative Transcriptomics using the MLM approach described in Section 2.3.1. Using the

methods described in Section 2.3.2, linear regression was performed using the 94,060 GEM markers identified. The most highly significant GEM associations detected, were further explored by mapping transcript abundance (measured as RPKM) as a trait against the SNP data (as described in Section 2.3.3).

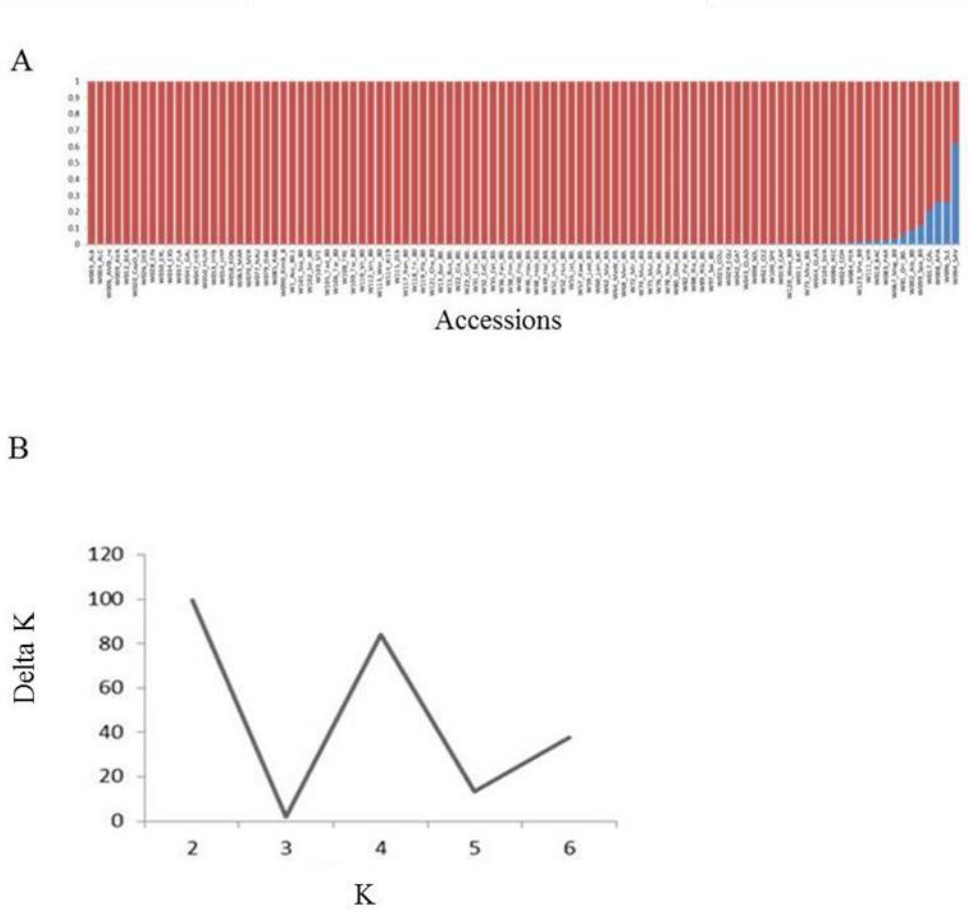

Figure 4. 13. Results obtained from STRUCTURE analysis in wheat. Two main populations (K) were identified (A). Using the methods described by Evanno et al., 2005, K=2 was confirmed to be the best estimate of K (B) (Harper et al., Under review).

4.2.2 Results

The Associative Transcriptomics analysis detected a number of marker associations for each of the traits included in the study. A summary of some of the most highly significant marker associations can be found in Tables 4.3 and 4.4 for the SNP markers and the GEM markers respectively. The script used in performing linear regression for the GEM analysis can be found in Supplementary Data file 4.2d. This file also includes scripts used in plotting the SNP and GEM analyses results. The full SNP and GEM association plots can be found in Supplementary Figures files 4.1-4.12. This section will focus only on the most promising results for each trait. Where relevant, this analysis has been carried out on both transformed and non-transformed data. Any differences observed between these analyses will be described.

Table 4. 3 (a). SNP associations detected through Associative Transcriptomics in wheat.

Trait	Linkage group	Marker ID	P value	Pseudomolecule position (kb)	Increasing allele (no. observations)	Decreasing allele (no. Observations)	Effect %	Potential candidate gene(s)	Candidate position in pseudomolecule (kb)
MOR	1B	B_comp205_c0_seq2:624	1.03E-05	33686.60	C (26)	A (74)	17.4	GDSL-like Lipase (AT1G54790)	33686.6
MOR	2D	D_comp970_c0_seq1:1030	8.98E-04	2040.50	G (61)	A (39)	10.5	acetyl xylan esterase (Os04g01980)	2048.8
MOR	2D	D_comp1058_c0_seq1:1573	1.68E-03	45761.20	C (13)	G (85)	18.9	*COP9 SIGNAOLSOME SUBUNIT 5B* (Os04g56070)	45243.6
Log_{10} MOR	1B	B_comp205_c0_seq2:624	4.79E-05	33686.60	C (26)	A (74)	17.8	GDSL-like Lipase (AT1G54790)	33686.6
Log_{10} MOR	2D	D_comp970_c0_seq1:1030	5.65E-04	2040.50	G (61)	A (39)	11.9	acetyl xylan esterase (Os04g01980)	2048.8
Log_{10} MOR	2D	D_comp1058_c0_seq1:1573	9.23E-03	45761.20	C (13)	G (85)	16.2	*COP9 SIGNAOLSOME SUBUNIT 5B* (Os04g56070)	45243.6
MOE	1B	B_comp205_c0_seq2:624	1.85E-05	33686.60	C (26)	A (74)	17.6	GDSL-like Lipase (AT1G54790)	33686.6
MOE	2D	D_comp970_c0_seq1:1030	8.02E-04	2040.50	G (61)	A (39)	10.7	acetyl xylan esterase (Os04g01980)	2048.8
MOE	2D	D_comp1058_c0_seq1:1573	1.32E-03	45761.20	C (13)	G (85)	19.4	*COP9 SIGNAOLSOME SUBUNIT 5B* (Os04g56070)	45243.6
Fmax	1B	B_comp205_c0_seq2:624	1.84E-05	33686.60	C(26)	A (74)	16.0	GDSL-like Lipase(AT1G54790)	33686.6
Fmax	6A	A_comp20392_c0_seq1:770	5.50E-04	14571.10	A (47)	T (45)	7.1	UDP-glucuronic acid transporter, (Os02g40030) [*1]/ *OsARF16* (Os02g41800)[*2]	14571.1[*1] /15189.7[*2]
Fmax	5B	B_comp6903_c0_seq5	3.49E-04		T (33)	C (67)	10.7	-	
F/V	1B	B_comp205_c0_seq2:624	1.25E-04	33686.60	C (26)	A (74)	16.0	GDSL-like Lipase (AT1G54790)	33686.6
F/V	6A	A_comp20392_c0_seq1:770	1.43E-03	14571.10	A (47)	T (45)	12.2	UDP-glucuronic acid(Os02g40030) [*1]/ *OsARF16*(Os02g41800) [*2]	14571.1[*1] /15189.7[*2]
F/V	5B	B_comp346_c0_seq1:809	5.76E-04	31799.80	C (8)	G (83)	20.2	-	-

Trait	Linkage group	Marker ID	P value	Pseudomolecule position (kb)	Increasing allele (no. observations)	Decreasing allele (no. observations)	Effect %	Potential candidate gene(s)	Candidate position in pseudomolecule (kb)
Stem hollow area	1B	B_comp2073_c0_seq1:412	4.05E-03	33645.20	G (85)	C (15	18.7	GDSL-like Lipase (AT1G54790)	33686.6
Outer cortex area	6A	A_comp10724_c0_seq1:311	7.90E-04	14590.20	T (48)	C (37)	21.3	-	-
Log10 outer cortex area	6A	A_comp10724_c0_seq1:311	9.11E-04	14590.20	T (41)	C (27)	20.5	-	-
Parenchyma area	5B	B_comp3333_c0_seq2:814	5.75E-05	32478.70	C (27)	T (67)	22.6	-	-
Grain weight	3D	D_comp9141_c0_seq1:355	1.60E-04	31299.10	T (90)	G (5)	31.7	-	-
Main stem threshed weight	6A	A_comp20392_c0_seq1:770	1.40E-04	14571.10	A (47)	T (45)	16.5	UDP-glucuronic acid transporter (Os02g40030)[*1]/ *OsARF16*(Os02g41800)[*2]	14571.1[*1] / 15189.7[*2]
Main stem threshed weight	3D	D_comp1133_c0_seq1:932	8.57E-04	413.20	A (73)	G (13)	19.0	-	-
Main stem threshed weight	5B	B_comp3140_c0_seq1:1427	6.63E-04	22514.20	A (37)	G (38)	23.2	*OsSAUR39* (Os09g37330)[*1]/ *OsSAUR48* (Os09g37430)[*2]	23464.7[*1]/ 23471.9[*2]
Main stem threshed weight	5B	B_comp6903_c0_seq5:522	4.99E-04	32574.30	T (33)	C (67)	13.4	-	-
Plant height	6A	A_comp20392_c0_seq1:770	4.59E-05	14571.10	A (47)	T (45)	18.0	UDP-glucuronic acid transporter, (Os02g40030) [*1]/*OsARF16* (Os02g41800) [*2]	14571.1[*1] / 15189.7[*2]
Plant height	2B	B_comp2616_c0_seq1:705	4.68E-05	43279.60	G (12)	A (87)	25.6	-	-
Plant height	4D	D_comp3497_c0_seq1:195	3.64E-04	34809.90	A (39)	G (53)	21.0	-	-
Plant height	5B	B_comp3140_c0_seq1:1427	5.37E-05	22514.20	A (37)	G (38)	29.0	*OsSAUR39* (Os09g37330)[*1]/ *OsSAUR48* (Os09g37430)[*2]	23464.7[*1]/ 23471.9[*2]

4.3b

Table 4. 4. Most significant GEM associations detected through Associative Transcriptomics in wheat

4.4a.

Trait	Sequencing platform	Linkage group	Marker ID	Marker pseudomolecule position (kb)	Candidate gene pseudomolecule position (kb)	P value	R^2	Trend	Candidate gene
Fmax	Hi-seq	5A	A_comp15078_c0_seq2	31559.40	31559.4	9.57E-08	0.42	+	RNA-binding motif protein (Bradi1g03270.1)
Fmax	GAII	5D	D_comp3235_c0_Seq1	11950.23	11950.23	4.47E-06	0.49	+	omega-3 fatty acid desaturase (Os12g01370)
Fmax	GAII	2B	B_comp4382_c0_seq1	42530.73	42530.73	3.42E-07	0.44	+	60S ribosomal protein L7-2, (Os04g51630)
F/V	Hi-seq	5A	A_comp15078_c0_seq2	31559.40	31559.4	2.05E-06	0.35	+	RNA-binding motif protein (Bradi1g03270.1)
F/V	Hi-seq	2B	B_comp6103_c0_seq16	43949.70	43949.7	2.63E-06	0.35	-	*PEX6* (Os04g52690)
F/V	Hi-seq	1B	B_comp13413_c0_seq4	2241.81	2241.808	3.73E-06	0.33	+	IQ calmodulin-binding motif family protein (Os05g03190)
F/V	Hi-seq	7B	B_comp2656_c0_seq2	30959.96	30959.96	7.88E-06	0.32	-	Uncharacterised (Os06g21600)
F/V	Hi-seq	7D	D_comp81488_c0_seq1	9206.42	9206.42	3.97E-06	0.34	-	No annotation
F/V	GAII	1A (peak)	A_comp58812_c0_seq1	2913.275	-	2.78E-05	0.33	-	No annotation
F/V	GAII	6A	A_comp59576_c0_seq1	20174.33	20174.33	2.20E-05	0.34	-	*CYTOKININ-O-GLUCOSYLTRANSFERASE 2* (Os04g25980)
F/V	GAII	6B	B_comp626_c0_seq1	19190.11	19190.11	7.12E-05	0.3	-	metal tolerance protein C3 (Os01g03914)
MOR	GAII	5B	B_comp6147_c0_seq2	26779.18	26779.18	1.01E-05	0.36	+	lysine-tRNA ligase activitY (Os03g38980)

Trait	Sequencing platform	Linkage group	Marker ID	Marker pseudomolecule position (kb)	Candidate gene pseudomolecule position (kb)	P value	R^2	Trend	Candidate gene
MOR	Hi-seq	7A	A_comp177291_c0_seq1	45368.04	45368.04	1.10E-10	0.55	+	receptor protein kinase *CLAVATA1* precursor (Os06g50340)
MOR	GAII	6A	A_comp38356_c0_seq1*	23560.41	23560.41	3.03E-06	0.39	+	RNA polymerase II (Os02g49150)
MOR	GAII	7B	B_comp10330_c0_seq1*	34413.15	34413.15	3.31E-06	0.39	+	*QUASIMODO 1* (AT3G25140)
MOR	GAII	2D (peak)	D_comp34057_c0_seq1	44955.11	45243	1.25E-05	0.36	+	*COP9 SIGNALOSOME SUBUNIT 5B* (Os04g56070)
MOR	GAII	4B	B_comp2268_c0_seq5	37272.39	37272.39	1.31E-05	0.35	+	40S ribosomal protein SA (Os03g08440)
MOR	GAII	6B	B_comp8659_c0_seq9*	22558.8	22558.8	1.36E-05	0.35	+	*EIF-2B* (Os02g56740)
MOR	GAII	2D	D_comp3514_c0_seq1*	16438.39	16438.39	1.59E-06	0.41	+	serine carboxypeptidase-like precursor (Os07g29620)
MOR	Hi-seq	5A	A_comp214176_c0_seq1	2562.42	2562.418	8.52E-10	0.52	+	serine/threonine-protein kinase *SAPK9* (Os12g39630)
MOR	Hi-seq	5A	A_comp835_c3_seq7	23140.17	23140.17	8.28E-10	0.52	+	aquaporin *PIP2.3* (Os04g44060)
MOR	Hi-seq	7A	A_comp177291_c0_seq1	45368.04	45368.04	1.10E-10	0.55	+	receptor protein kinase *CLAVATA1* precursor (Os06g50340)
MOR	Hi-seq	7A	A_comp835_c1_seq6	46809.51	46809.51	1.60E-09	0.5	+	aquaporin *PIP2.3* (Os04g44060)

4.4b

Trait	Sequencing platform	Linkage group	Marker ID	Marker pseudomolecule position (kb)	Candidate gene pseudomolecule position (kb)	P value	R^2	Trend	Candidate gene
MOR	Hi-seq	1B	B_comp35078_c0_seq1	8881.54	8881.54	5.43E-09	0.48	+	protein kinase *APK1B* (Os03g06330)
MOR	Hi-seq	4B	B_comp16517_c0_seq4	29791.97	29791.97	1.97E-08	0.46	+	nucleic acid binding protein (Os07g48410)
Log_{10} MOR	GAII	6A	A_comp38356_c0_seq1*	23560.41	23560.41	1.69E-06	0.4	+	*RNA POLYMERASE II* (Os02g49150)
Log_{10} MOR	GAII	5B	B_comp6147_c0_seq2	26779.18	26779.18	1.10E-05	0.36	+	lysine-tRNA ligase activitY (Os03g38980)
Log_{10} MOR	GAII	7B	B_comp10330_c0_seq1*	34413.15	34413.15	5.08E-05	0.31	+	*QUASIMODO 1* (AT3G25140)
Log_{10} MOR	GAII	2D	D_comp3514_c0_seq1*	16438.39	16438.39	3.19E-06	0.39	+	serine carboxypeptidase-like precursor (Os07g29620)
Log_{10} MOR	GAII	2D (peak)	D_comp34057_c0_seq1	44955.11	45243	4.30E-05	0.31	+	*COP9 SIGNALOSOME SUBUNIT 5B*(Os04g56070)
MOR	GAII	6B	B_comp8659_c0_seq9*	22558.80	22558.8	8.42E-05	33	+	*EUKARYOTIC TRANSLATION INITIATION FACTOR-2B* (*EIF-2B*) (Os02g56740)
Log_{10} MOR	Hi-seq	5A	A_comp214176_c0_seq1	2562.42	2562.42	2.34E-09	0.29	+	serine/threonine-protein kinase *SAPK9* (Os12g39630)
Log_{10} MOR	Hi-seq	5A	A_comp835_c3_seq7	23140.17	23140.17	2.32E-09	0.23	+	aquaporin *PIP2.3* (Os04g44060)
Log_{10} MOR	Hi-seq	6A	A_comp1067_c0_seq1	1440.76	1440.76	4.98E-08	0.25	+	nodule membrane protein (Os02g03230)
Log_{10} MOR	Hi-seq	7A	A_comp177291_c0_seq1	45368.04	45368.04	6.04E-09	0.25	+	receptor protein kinase *CLAVATA1* precursor (Os06g50340)

4.4c

Trait	Sequencing platform	Linkage group	Marker ID	Marker pseudomolecule position (kb)	Candidate gene pseudomolecule position (kb)	P value	R^2	Trend	Candidate gene
Log_{10} MOR	Hi-seq	7A	A_comp835_c1_seq6	46809.51	46809.51	7.95E-09	0.23	+	aquaporin *PIP2.3* (Os04g44060)
Log_{10} MOR	Hi-seq	1B	B_comp35078_c0_seq1	8881.54	8881.54	1.47E-08	0.22	+	protein kinase *APK1B* (Os03g06330)
Log_{10} MOR	Hi-seq	4B	B_comp16517_c0_seq4	29791.97	29791.97	1.65E-08	0.21	+	nucleic acid binding protein (Os07g48410)
MOE	Hi-seq	7A	A_comp177291_c0_seq1	45368.04	45368.04	1.06E-10	0.55	+	receptor protein kinase *CLAVATA1* precursor (Os06g50340)
MOE	GAII	6A	A_comp38356_c0_seq1*	23560.41	23560.41	4.40E-06	0.38	+	RNA polymerase II (Os02g49150)
MOE	GAII	4B	B_comp2268_c0_seq5	37272.39	37272.39	1.55E-05	0.34	+	40S ribosomal protein SA(Os03g08440)
MOE	GAII	6B	B_comp8659_c0_seq9*	22558.8	22558.8	2.37E-05	0.33	+	*EUKARYOTIC TRANSLATION INITIATION FACTOR-2B* (*EIF-2B*) (Os02g56740)
MOE	GAII	7B	B_comp10330_c0_seq1*	34413.15	34413.15	3.19E-06	0.39	+	*QUASIMODO 1* (AT3G25140)
MOE	GAII	2D (peak)	D_comp17582_c0_seq1	45243.00	45243	1.52E-05	0.35	+	*COP9 SIGNALOSOME SUBUNIT 5B*(Os04g56070)
MOE	GAII	2D	D_comp3514_c0_seq1*	16438.39	16438.39	2.27E-06	0.4	+	serine carboxypeptidase-like precursor (Os07g29620)

4.4d

Trait	Sequencing platform	Linkage group	Marker ID	Marker pseudomolecule position (kb)	Candidate gene pseudomolecule position (kb)	P value	R^2	Trend	Candidate gene
Outer cortex thickness	Hi-seq	1A	A_comp107441_c0_seq1	43463.34	43463.34	6.19E-04	0.2	+	*OsMPK20-5*(Os05g49140)
Outer cortex thickness	Hi-seq	7A	A_comp53514_c0_seq6	45632.35	45632.35	1.03E-04	0.2	-	protein kinase *APK1B* (Os02g44920)
Outer cortex thickness	Hi-seq	2A	A_comp61367_c0_seq2	40807.87	40807.87	2.89E-04	0.22	-	protein kinase C binding (Bradi5g18200.1)
Outer cortex thickness	Hi-seq	2A	A_comp21739_c0_seq1	9101.99	9101.99	7.63E-04	0.19	-	gibberellin receptor GID1L2, putative, expressed (Os07g44890)
Outer cortex thickness	Hi-seq	5A	A_comp8781_c0_seq22	17486.87	17486.87	5.59E-04	0.2	+	choline/ethanolamine kinase, putative, expressed (Os09g26700)
Outer cortex thickness	Hi-seq	4B	B_comp17639_c0_seq13	2207.92	2207.92	9.31E-04	0.19	+	No annotation
Outer cortex thickness	Hi-seq	4B	B_comp2346_c1_seq53	28248.13	28248.13	5.34E-04	0.21	+	Unknown function (Os02g30310)
Outer cortex thickness	Hi-seq	5B	B_comp1173_c0_seq1	684.96	684.96	1.25E-03	0.18	+	synthase, chloroplast precursor (Os12g43130)
Outer cortex thickness	Hi-seq	5B	B_comp750_c0_seq1	23948.31	23948.31	1.61E-03	0.18	+	synthase, chloroplast precursor (Os12g43130)
Outer cortex thickness	Hi-seq	7B	B_comp8132_c0_seq1	36085.83	36085.83	1.23E-04	0.25	-	Unknown expressed protein (Os06g09840)
Outer cortex thickness	Hi-seq	5D	D_comp444927_c0_seq1	27976.49	27976.49	5.02E-04	0.21	-	glucuronosyltransferase activity (Os03g55040)

4.4e

Trait	Sequencing platform	Linkage group	Marker ID	Marker pseudomolecule position (kb)	Candidate gene pseudomolecule position (kb)	P value	R^2	Trend	Candidate gene
Outer cortex thickness	GAII	2A	A_comp31610_c0_seq2	12597.96	12597.96	4.73E-05	0.32	-	*SV2* related protein, putative (Os07g38400)
Outer cortex thickness	GAII	3D	D_comp26499_c0_seq1	2219.89	2219.89	1.96E-04	0.27	+	50S ribosomal protein L14 (Os12g42180)
Log_{10} outer cortex thickness	Hi-seq	1A	A_comp107441_c0_seq1	43463.34	43463.34	1.14E-05	0.36	+	*OsMPK20-5* (Os05g49140)
Log_{10} outer cortex thickness	Hi-seq	7A	A_comp53514_c0_seq6	45632.35	45632.35	6.82E-04	0.19	-	protein kinase *APK1B* (Os02g44920)
Log_{10} outer cortex thickness	Hi-seq	2A	A_comp61367_c0_seq2	40807.87	40807.87	1.77E-04	0.21	-	protein kinase C binding (Bradi5g18200.1)
Log_{10} outer cortex thickness	Hi-seq	4B	B_comp17639_c0_seq13	2207.92	2207.92	8.05E-05	0.28	+	No annotation
Log_{10} outer cortex thickness	Hi-seq	4B	B_comp2346_c1_seq53	28248.13	28248.13	1.42E-04	0.25	+	Unknown function (Os02g30310)
Log_{10} outer cortex thickness	Hi-seq	6B	B_comp3218_c0_seq3	16710.51	16710.51	3.10E-05	0.15	+	*TRANSLATION INITIATION FACTOR IF-2* (Os04g47860)
Log_{10} outer cortex thickness	Hi-seq	2D	D_comp188604_c0_seq1	46374.16	46374.16	2.98E-06	0.32	+	transposon protein (Os06g30700)
Log_{10} outer cortex thickness	GAII	3A	A_comp18677_c0_seq1	5918.39	5918.39	3.12E-04	0.17	-	receptor protein kinase *PERK1* (Os01g12720)
Log_{10} outer cortex thickness	GAII	2B	B_comp8127_c0_seq13	16470.54	16470.54	4.38E-05	0.26	-	transposon protei (Os05g06200)
Log_{10} outer cortex thickness	GAII	7B	B_comp8134_c0_seq3	6565.23	6565.23	1.16E-04	0.14	-	ATP binding protein (Os08g43400)

4.4f

Trait	Sequencing platform	Linkage group	Marker ID	Marker pseudomolecule position (kb)	Candidate gene pseudomolecule position (kb)	P value	R^2	Trend	Candidate gene
Log_{10} outer cortex thickness	GAII	3D	D_comp9058_c0_seq4	14003.87	14003.87	3.91E-04	0.16	-	serine carboxypeptidase 1 precursor (Os10g01110)
Log_{10} outer cortex thickness	GAII	2A	A_comp98579_c0_seq4	13645.51	13645.51	1.12E-04	0.14	-	serine/threonine-protein kinase receptor precursor (Os07g36780)
Stem hollow area	Hi-seq	1A	A_comp15240_c0_seq4	41137.36	41137.36	2.36E-06	0.35	+	Unknown expressed protein (Os05g45400)
Stem hollow area	Hi-seq	7A	A_comp353946_c0_seq2	17693.78	17693.78	1.99E-06	0.35	+	aminotransferase, class III family protein (Os04g52450)
Stem hollow area	GAII	4D	D_comp607_c0_seq1	792.16	792.158	9.78E-06	0.36	-	protein retrotransposon protein, Ty1-copia subclass (Os03g46770)
Parenchyma area	Hi-seq	5A	A_comp74327_c1_seq3	28864.83	28864.83	8.96E-05	0.26	-	transposon protein (Os03g56630)
Parenchyma area	Hi-seq	7A	A_comp113240_c0_seq3	4546.73	4546.73	3.64E-04	0.22	-	protein myb-like protein Os11g06190
Parenchyma area	Hi-seq	5D	D_comp68630_c0_seq2	1269.33	1269.334	3.89E-05	0.28	-	No annotation
Parenchyma area	Hi-seq	5D	D_comp23979_c0_seq19	18121.47	18121.47	6.91E-05	0.26	+	*CESA9* Os09g25490
Parenchyma area	GAII	1A	A_comp104369_c0_seq3	9983.39	9983.39	1.58E-04	0.28	+	SNF2 domain-containing protein (Os10g31970)

4.4g

Trait	Sequencing platform	Linkage group	Marker ID	Marker pseudomolecule position (kb)	Candidate gene pseudomolecule position (kb)	P value	R^2	Trend	Candidate gene
Parenchyma area	GAII	5B	B_comp538_c0_seq1	11845.12	11845.12	7.10E-05	0.3	+	ferritin-1, chloroplast precursor(Os12g01530)
Parenchyma area	GAII	5B	B_comp22750_c0_seq4	32541.8	32541.8	1.19E-04	0.28	+	sterile alpha motif (SAM) domain family protein Os03g63320
Parenchyma area	GAII	7B	B_comp53868_c0_seq1	35702.53	35702.53	3.75E-05	0.32	+	hypothetical protein (s06g10610)
Parenchyma area	GAII	6D	D_comp1335_c0_seq3	8963.41	8963.41	3.75E-05	0.32	+	ATA15 protein (Os05g05600)
Plant height	Hi-seq	5A	A_comp1666_c1_seq1	8933.05	8933.05	3.46E-07	0.39	-	histone deacetylase 10 (Os12g08220)
Plant height	Hi-seq	6A	A_comp1212_c1_seq242	15264.4	15264.4	1.08E-06	0.37	-	expressed protein (Os02g41990)
Plant height	Hi-seq	4B	B_comp10968_c0_seq5	38951.36	38951.36	2.48E-08	0.45	-	expressed protein (Os01g41950)
Plant height	Hi-seq	5B(peak)	B_comp14715_c0_seq3	22557.79	23545.69	7.59E-08	0.43	+	*OsSAUR45* - Auxin-responsive SAUR gene family member (Os09g37400)
Plant height	Hi-seq	6B	B_comp7640_c0_seq5	20753.47	20753.47	1.14E-06	0.37	-	NADPH:quinone reductase activity, zinc ion binding (Os03g01190)
Plant height	GAII	6A	A_comp295970_c0_seq1	15190.72	15190.72	1.39E-06	0.41	-	auxin response factor (*ARF*) (Os02g41800)

4.4h

Trait	Sequencing platform	Linkage group	Marker ID	Marker pseudomolecule position (kb)	Candidate gene pseudomolecule position (kb)	P value	R^2	Trend	Candidate gene
Plant height	GAII	4B	B_comp9780_c0_seq2	27385.53	27385.53	5.38E-07	0.44	+	ATPase activity (Os11g47970)
Plant height	GAII	2D	D_comp41801_c0_seq2	38783.72	38783.72	8.38E-08	0.48	+	No annotation
Main stem grain weight	Hi-seq	3A	A_comp12896_c0_seq1	22637.57	22637.57	4.69E-06	0.33	+	metal-dependent hydrolase (Os01g54960)
Main stem grain weight	Hi-seq	4A	A_comp18136_c2_seq1	-	-	3.36E-06	0.34	+	No annotation
Main stem grain weight	Hi-seq	4B	B_comp14960_c0_seq1	11338.07	11338.07	1.48E-05	0.3	-	verticillium wilt disease resistance protein (Os01g04070)
Main stem grain weight	Hi-seq	4B	B_comp33413_c0_seq3	33716.42	33716.42	1.54E-05	0.3	+	ankyrin repeat family protein(Os03g47693)
Main stem grain weight	Hi-seq	5B	B_comp7082_c0_seq1	9321.774	9321.77	3.03E-05	0.29	-	allophanate hydrolase activity, fatty acid amide hydrolase activity, amidase activity (Os12g07150)
Main stem grain weight	Hi-seq	4D	D_comp150298_c0_seq2	5692.46	5692.46	2.10E-05	0.29	+	resistance protein (Os11g11920)
Main stem grain weight	Hi-seq	5D	D_comp40619_c0_seq1	28149.79	28149.79	2.69E-05	0.32	-	magnesium-dependent protein serine/threonine phosphatase activity (Os03g55320)

4.4i

Trait	Sequencing platform	Linkage group	Marker ID	Marker pseudomolecule position (kb)	Candidate gene pseudomolecule position (kb)	P value	R^2	Trend	Candidate gene
Main stem grain weight	GAII	1A	A_comp72548_c0_seq4	11873.86	11873.86	1.24E-05	0.35	-	arylformamidase activity, kynurenine-oxoglutarate transaminase activity (Os10g34350)
Main stem grain weight	GAII	5A(peak)	A_comp5799_c0_seq9	18777.39	-	1.47E-06	0.29	-	-
Main stem grain weight	GAII	1B	B_comp13452_c0_seq4	23366.92	23366.92	1.36E-05	0.35	-	androgen receptor binding, estrogen receptor binding (Os01g07400)
Main stem grain weight	GAII	2B	B_comp752_c0_seq9	29275.7	29275.7	9.12E-06	0.36	-	No annotation
Main stem grain weight	GAII	4B	B_comp14228_c0_seq1	20502.65	20502.65	1.17E-06	0.41	-	DJ-1 family protein (Os11g37920)
Main stem grain weight	GAII	5B	B_comp391_c0_seq1	26807.77	26807.77	1.82E-05	0.34	-	No annotation
Main stem grain weight	GAII	4D	D_comp130805_c0_seq1	22234.88	22234.88	9.58E-06	0.36	+	Plant viral-response family protein (Os11g40570)
Main stem grain weight	GAII	6D	D_comp637810_c0_seq1	4191.78	4191.775	4.17E-05	0.32	-	STE_PAK_Ste20_STLK.3 (Bradi3g05890.1)
Main stem threshed weight	Hi-seq	2A	A_comp685_c0_seq1	3150.37	3150.373	1.29E-05	0.3	+	ruBisCO large subunit-binding protein subunit alpha, chloroplast precursor (Os12g17910)

4.4j

Trait	Sequencing platform	Linkage group	Marker ID	Marker pseudomolecule position (kb)	Candidate gene pseudomolecule position (kb)	P value	R^2	Trend	Candidate gene
Main stem threshed weight	Hi-seq	2A	A_comp157344_c0_seq1	4359.87	4359.87	2.31E-05	0.29	+	*OsWAK121* - OsWAK receptor-like protein kinase (Os11g47110)
Main stem threshed weight	Hi-seq	2A	A_comp309286_c0_seq1	44434.15	44434.15	3.62E-06	0.34	+	NBS-LRR disease resistance protein (Os06g49380)
Main stem threshed weight	Hi-seq	1B	B_comp9909_c0_seq6	26183.95	26183.95	1.90E-05	0.29	+	CRAL/TRIO domain containing protein (Os05g18294)
Main stem threshed weight	Hi-seq	1B	B_comp13413_c0_seq4	2241.81	2241.81	1.24E-05	0.3	+	IQ calmodulin-binding motif family protein (Os05g03190)
Main stem threshed weight	Hi-seq	2D	D_comp124798_c0_seq1	23797.05	23797.05	1.32E-05	0.3	+	phytochelatin synthetase-like conserved region family protein (Os03g30250)
Main stem threshed weight	Hi-seq	5D	D_comp23979_c0_seq19	18121.47	18121.47	1.07E-06	0.37	+	*CESA9* Os09g25490
Main stem threshed weight	Hi-seq	7D	D_comp521_c0_seq1	2249.02	2249.02	5.40E-06	0.33	+	ruBisCO large subunit-binding protein subunit beta, chloroplast (Os06g02380)
Main stem threshed weight	GAII	7B	B_comp14298_c0_seq4	33125.29	33125.29	1.19E-06	41.85	+	ketose-bisphosphate aldolases family protein (Os06g14740)

4.4k

4.2.2.1 Material strength: MOR and MOE

Through analysing the results obtained for MOE and MOR, it was noticed that very similar SNP and GEM associations for the two traits were obtained. For this reason, to avoid extensive repetition, only the results obtained for MOR will be described here. A summary of the results obtained for each of these material strength traits, can however be found in Tables 4.3 and 4.4. For MOR, given the previously described detection of non-normality, the Associative Transcriptomics analysis was carried out on both raw and transformed data.

For the SNP analysis, three clear SNP association peaks were detected (Figures 4.14 and 4.15). The first of these was detected on chromosome 1B and was seen very clearly in both the analyses carried out on the transformed and on the non-transformed data (Figure 4.14 shows the associations for the non-transformed data). In both cases, the most highly significant marker was B_comp205_c0_seq2:624. In the analysis carried out on non-transformed data, this marker was assigned a P value of 1.03E-05and a trait effect of 17.38%. In the $\log_{10}$ transformed data analysis, B_comp205_c0_seq2:624 was assigned a P value of 4.79E-05 and trait effect of 17.8 %. Using the wheat pseudomolecule, this marker was found to correspond to a gene with high sequence similarity to an Arabidopsis GDSL-like Lipase/Acylhydrolase superfamily gene (AT1G54790). GDSL-like lipases are thought to be involved in the hydrolysis of ester bonds in cell wall xylans and have been found to have xylan acetylase activity (Akoh et al., 2004).

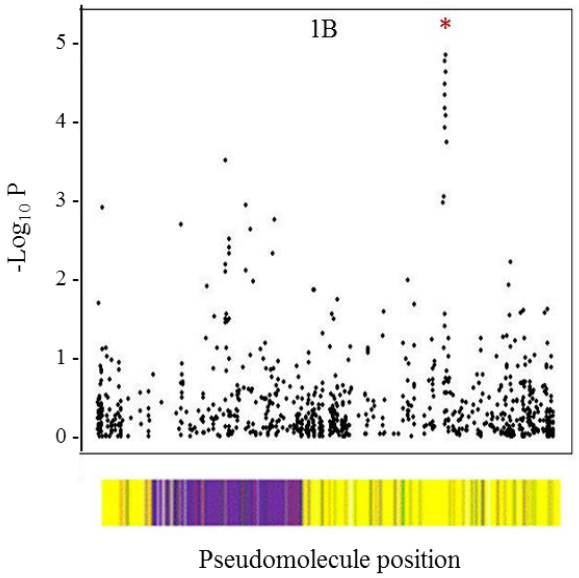

Figure 4. 14. SNP association (marked with a red asterisk) detected on chromosome 1B for MOR. Marker associations are plotted as –log10P against pseudomolecule position.

The second SNP association detected for MOR is on chromosome 2D (marked with a red asterisk), with the top marker being D_comp970_c0_seq1:1030. This marker association was also seen in the results obtained for both the non-transformed (Figure 4.15A) and transformed data. For the non-transformed data, D_comp970_c0_seq1:1030 was assigned a P value of 8.99E-04 and a trait effect of 10.5%. This marker was seen to be slightly more highly significant in the $\log_{10}$ transformed data, with a P value of 5.65E-04 and a trait effect of 11.94 %. This marker is in close proximity to a gene orthologous to a rice acetyl xylan esterase (*AxeA*) (0S04g01980). *AxeA*, is thought to have hydrolase activity, specifically acting on ester bonds in the deacetylation of xylans in the plant cell wall (Alalouf et al., 2011). This is a very similar function to that previously described for the candidate detected on chromosome 1B. Previous work in Arabidopsis has shown that xylan acetylation is an important contributor to stem strength. For example, the *eskimo-1* mutant, which displays reduced xylan acetylation, was seen to exhibit reduced cell wall thickening and significantly weaker stems in comparison to WT (Yuan et al., 2013). Xylan acetylation is considered to be one of the key factors impeding the breakdown of lignocellulosic biomass for ethanol production. This is thought to be due to the acetyl groups blocking the access of cell wall degrading

enzymes to the cellulose and hemicellulose components (Pawar et al., 2013). However, the results presented here suggest that alterations in xylan acetylation may lead to variation in stem mechanical strength.

A third SNP peak for material strength traits was detected on chromosome 2D. This SNP association was seen to be much weaker in comparison to those so far described. In the analysis carried out on the non-transformed data the most highly significant marker was D_comp1058_c0_seq1:1573 (P=1.68E-03, trait effect 18.9 %) (shown in Figure 4.15A marked with a dotted border). In the analysis carried out on the transformed data, this same marker was assigned a P value of 9.23E-03 and a trait effect of 16.21 %. Although this SNP association may itself not be very convincing in terms of the P values obtained, this region was also detected in the GEM analysis in both the transformed and non-transformed data analyses (see Figure 4.15B for non-transformed data). For both transformed and non-transformed data, the most significant GEM marker was D_comp34057_c0_seq1 (P=1.25E-05 for non-transformed and P=4.3E-05 for $\log_{10}$ MOR).

Further assessment of the highly significant GEMs within this region, revealed that in all cases a positive correlation between the transcript abundance (measured as RPKM) of each marker and the material strength trait data was seen. An example of some of these relationships is shown in Figure 4.15B where for non-transformed MOR, the top marker D_comp34057_c0_seq1, was found to explain 36 % of the variation. The identification of a large region of GEMs showing differential expression between accessions may be indicative of a genomic feature such as gene duplication or deletion (Harper et al., 2012).

In addition to this GEM association peak, several individual GEMs were also found to show highly significant associations to material strength. Table 4.4 summarises these results

On chromosome 6A a highly significant GEM, A_comp38356 was detected (P=1.69E-06, R^2=0.40 for $\log_{10}$ MOR and P=3.03E-06, R^2=0.39 for MOR). Transcript abundance for this GEM correlates positively with MOR. This GEM corresponds to a rice orthologue (Os02g49150) which has been described as encoding an RNA pol II. A second GEM association, B_comp10330_c0_seq1, was detected on chromosome 7B (P=3.31E-06, R^2=0.39for MOR and P=5.08E-05, R^2

=0.31for $\log_{10}$MOR). Transcript abundance for this locus correlates positively with material strength. This marker corresponds to an orthologue of Arabidopsis *QUASIMODO 1* (AT3G25140). Mutants defective in this gene exhibit a number of defects including reduced homogalacturonan content in the cell wall and reduced cell adhesion (Bouton et al., 2002). On chromosome 2D, D_comp3514_c0_seq1 was found to be highly associated to material strength. This association was again seen in both the transformed (P=3.19E-06, R^2 = 0.39) and non-transformed (P=1.59E-06, R^2=0.41) analyses. Again, there is a positive relationship between transcript abundance and MOR. This marker corresponds to an orthologue of Arabidopsis *SERINE CARBOXYPEPTIDASE-LIKE 49* (AT3G10410). The Tobacco homolog of this gene, *NtSCP1*, was found to be important for cell elongation and it has been proposed that this gene may target proteins involved in cell wall remodelling (Bienert et al., 2012), making this a very plausible candidate gene for stem mechanical strength. As a final example, on chromosome 6B, B_comp8659_c0_seq9 was detected (P=1.36E-05, R^2=0.35 for MOR and P=8.42E-05, R^2=0.33 for $\log_{10}$MOR). This marker corresponds to a gene which has been described in rice as a translation initiation factor, *EIF-2B* epsilon subunit (Os02g56740). In rice, this gene is thought to have a role in the recruitment of mRNAs and the machinery required for translation. A related protein however, *EIF-5A*, has been found to be involved in a signalling pathway contributing to cell wall integrity and formation (Valentini et al., 2002). It is therefore possible that *EIF-2B* also has a similar, additional function.

To further analyse the individual GEM associations detected, their respective transcript abundances (measured as RPKM) were mapped as traits against the SNP data of the 100 wheat accessions. Interestingly, for each of the described GEMs, a strong SNP peak was detected on chromosome 2D, the same region previously described for MOR in both the SNP and GEM analyses. An example of this can be seen in Figures 4.15 for a single GEM (D_comp3514_c0_seq1) detected on chromosome 2D (Figure 4.15B marked with a red asterisk). Figure 4.15C shows a clear SNP association on chromosome 2D following the mapping of the transcript abundance values for this GEM as a trait against the SNP data. All additional GEMs found to show this relationship with the 2D locus can be seen marked with a red asterisk in Table 4.4. This finding is indicative of an interaction between those genes detected as single marker associations and one or more genes located within the 2D

region. Due to the many genes showing differential expression in the 2D region detected in the GEM analysis for MOR, it is difficult to propose a candidate gene. However, one gene, which corresponds to one of the most highly associating GEM markers (D_comp17582_c0_seq1) within this peak, may be considered a very plausible target. This gene corresponds to an orthologue in rice described as a *COP9 SIGNALOSOME SUBUNIT 5B* (Os04g56070) (labelled in Figure 4.15B). The COP9 signalosome is a multi-protein complex which is known to be involved in protein degradation and has, in plants, been implicated in a number of developmental processes including photomorphogenesis (light-mediated growth), cell cycle progression and gene expression (Nezames and Deng, 2012) . Interestingly, in Fungi, the COP9 signalosome has been implicated in cell wall remodelling. Work conducted by Nahlik et al (2012), found that in the absence of a functional COP9 complex, *Aspergillus nidulans* exhibits altered expression of genes involved in cell wall remodelling (Nahlik et al., 2010). Furthermore, one of the single GEM associations detected for material strength, corresponds to a eukaryotic translation initiation factor (*EIF-2B*) gene. A number of studies have detected interactions with EIF-related genes and the COP9 signalosome (Kim et al., 2001). In addition, a number of single GEM associations which correspond to genes involved in cell wall biosynthesis were also detected. The transcript abundance (measured as RPKM) of these GEMs, when mapped as a trait against the SNP data, also revealed a SNP association in the 2D region. Given this, it is possible that these genes are interacting with the COP9 complex to cause variation in material strength by means of a pathway analogous to that seen in *Aspergillus nidalus*, contributing to cell wall remodelling. To allow for the testing of this candidate gene and further exploration of the detected interaction between this candidate and the many GEMs associating to the 2D locus, it would be interesting to study a *COP9 SUBUNIT 5B* knockout mutant. This could be carried out through the utilisation of the recently developed wheat (Cadenza) TILLING population developed in the lab of Cristobal Uauy, John Innes Centre. Following the detection of mutants within the 2D copy of this gene, it would be possible to a) assess any changes in material strength in comparison to a non-mutagenized Cadenza control and b) assess the expression levels of *EIF-2B* and the cell wall-related genes which, through the GEM analysis were found to interact with the 2D locus. This would allow for clear confirmation of any interactions

present between these loci and the role of the COP9 signalosome in contributing to stem material strength in wheat.

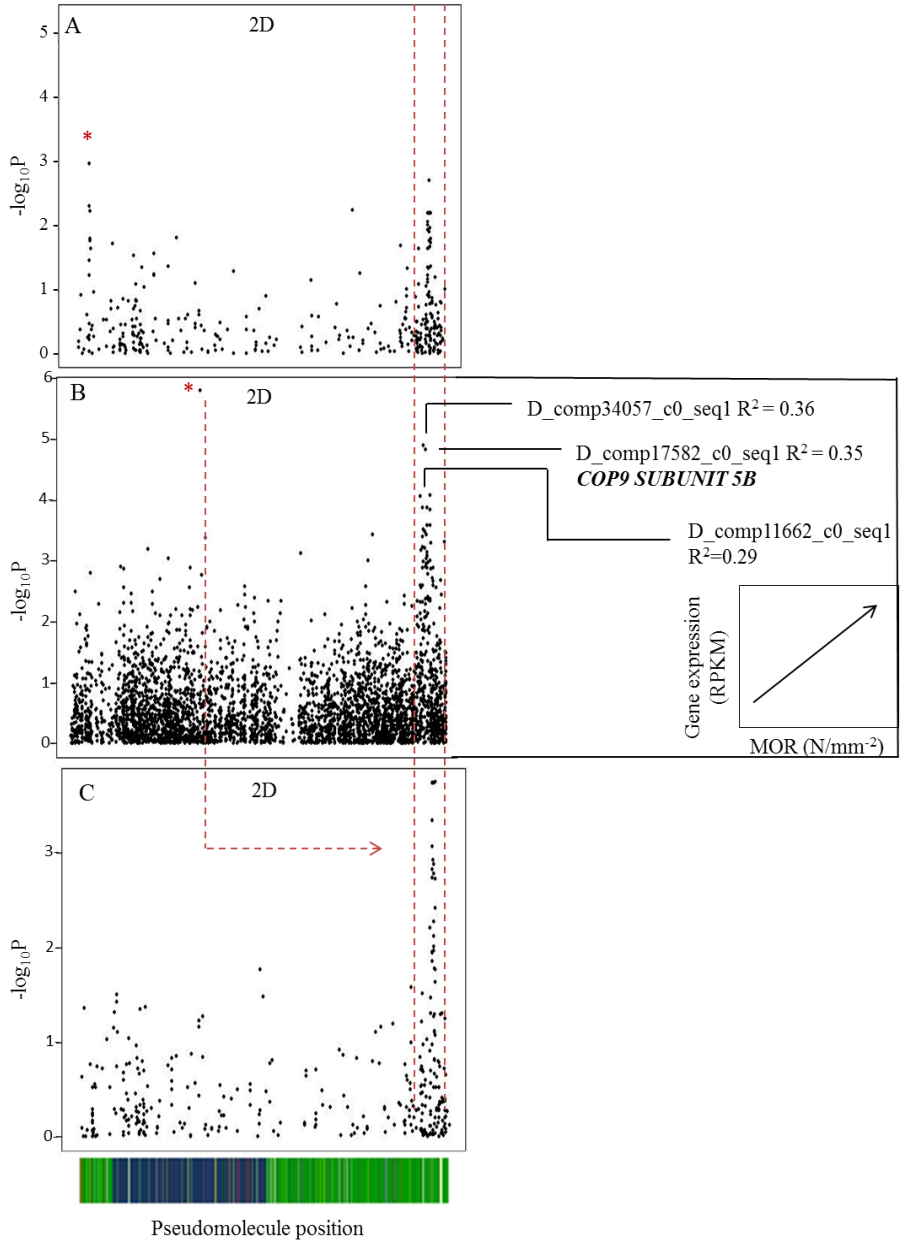

Figure 4. 15. Two SNP associations detected for MOR on chromosome 2D (one marked with a red asterisk (A). The 2D locus marked with a dotted border was also detected in the GEM analysis (transcript abundance in this region correlates positively with MOR) (B). A single GEM association can also be seen on this chromosome (marked with a red asterisk) (B). The transcript abundance values (measured as RPKM) when mapped as a trait against the SNP data, reveal a SNP association on 2D (C). Marker associations are plotted as –log10P against pseudomolecule position.

4.2.2.2 Absolute strength measures: Fmax and F/V

As seen for the material strength traits, very similar results were obtained for the absolute strength traits, Fmax and F/V. The Figures presented here summarise the results obtained for Fmax.

The strongest SNP association was detected on chromosome 1B (Figure 4.16B), a region previously described for the material strength traits. As seen for MOR and MOE, the most significant marker within this association peak is B_comp205_c0_seq2:624 (P=1.84E-05, trait effect 16% for Fmax and P=1.25E-04, trait effect 16% for F/V).

In addition, further minor associations were identified for these absolute strength traits. For example, on chromosome 6A (Figure 4.16A), for both Fmax and F/V, a weak signal was detected with the top marker for both Fmax and F/V being A_comp20392_c0_seq1:770 (P=5.50E-04, trait effect 7.1% for Fmax and 1.43E-03, trait effect 12.2% for F/V). A minor peak was also detected on chromosome 5B for both Fmax and F/V (Figure 4.16C). The top marker for this association for Fmax was B_comp6903_c0_seq5 (P=3.49E-04, trait effect 10.7%). For F/V the most highly associating marker was B_comp346_c0_seq1:809 (P=5.76E-04, trait effect 20.2 %). As described below, the associations detected on 6A and 5B were also detected for stem structural and morphological traits including main stem threshed stem weight and plant height. As described in Section 4.1.2.3, these absolute strength traits were found to correlate significantly with both threshed weight and plant height. The results presented here provide evidence of a genetic component to these correlations. Furthermore, the detection of both common associations and unique associations for the material and absolute strength measures, suggests that the dissection of the genetic components of these related traits has, at least to some extent, been successful.

As can be seen from Table 4.4, GEM associations were also detected for these absolute strength traits. In most cases, the relevance of the genes corresponding to the markers is not clear. One marker, A_comp59576_c0_seq1, detected on chromosome 6A (P=2.20E-05, R^2=0.34) for F/V was however found to be an orthologue of a glycosyltransferase gene in rice (Os04g25980). Not only does this gene have a potential role to play in carbohydrate synthesis, but may also have an

effect on Cytokinin signalling, which is known to be important in stem development (El-Showk et al., 2013). This GEM may therefore represent a good candidate gene for stem mechanical strength.

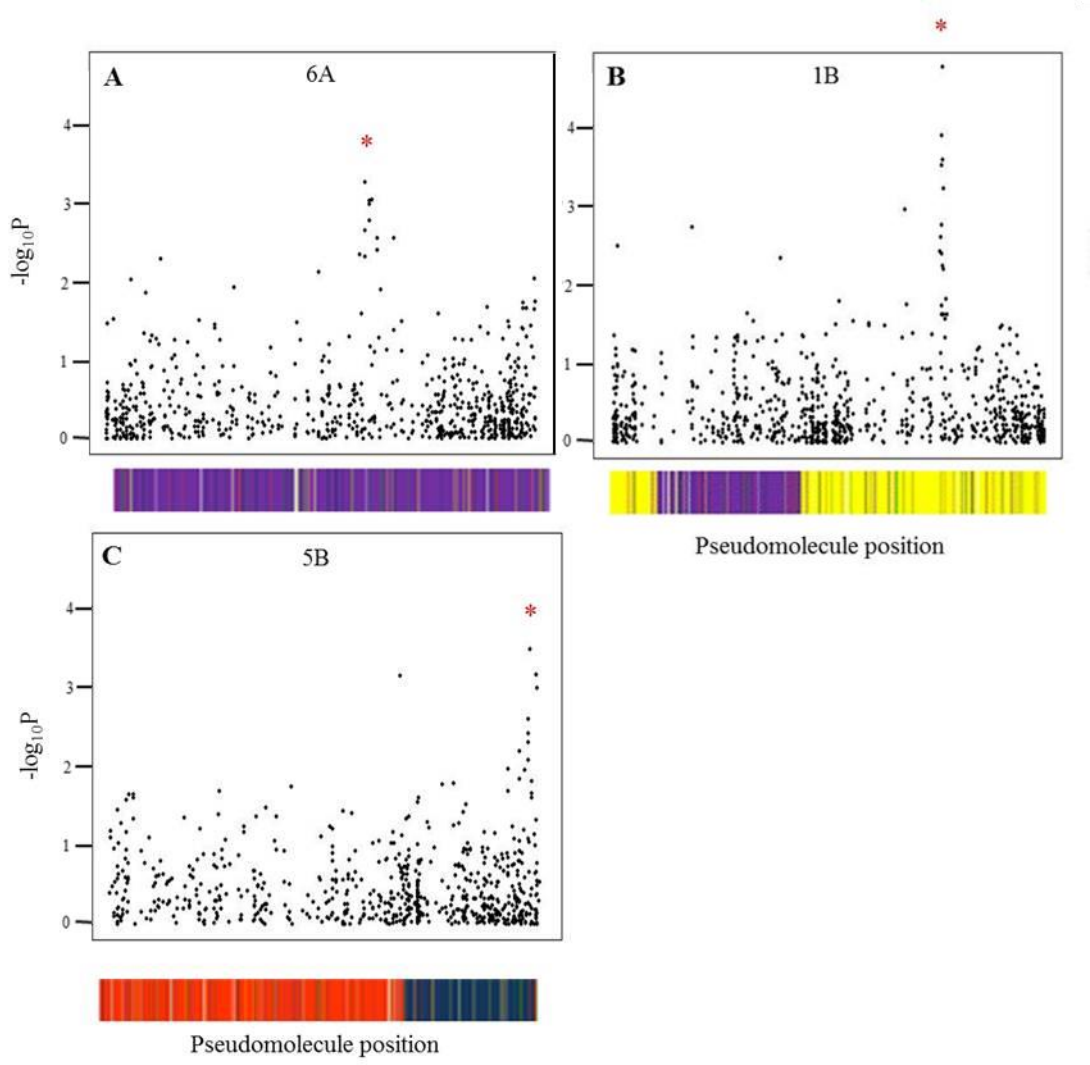

Figure 4. 16. Associative Transcriptomics SNP analysis results for Fmax and F/V (Fmax presented here). SNP associations can be seen marked with red asterisks. Marker associations are plotted as –log10P against pseudomolecule position.

4.2.2.3 Plant height

For plant height, two main SNP associations were identified. These associations were detected on chromosome 6A (Figure 4.17A) and 5B (4.18A). These SNP associations were also found to have coincident GEM association peaks. On chromosome 6A, the most highly significant SNP marker was

A_comp20392_c0_seq1:770 (P=4.59E-05, trait effect 18 %) (Figure 4.17A). This marker corresponds to a gene which in rice has been described as a UDP-glucuronic acid transporter. UDP-glucuronic acid is known to influence height in Arabidopsis (Reboul et al., 2011). However, despite this being a very plausible candidate for plant height, a second candidate was also detected in the GEM analysis for this 6A region (Figure 4.17B) where marker A_comp295970_c0_seq1, corresponds to an orthologue of a rice auxin-response gene, Os02g41800 (*OsARF16*). Transcript abundance (measured as RPKM) for this marker correlate negatively with plant height (R^2=0.41, P=1.39E-06). It is of course possible that both of these candidate genes are involved in the genetic control of plant height.

The signal detected on chromosome 5B may also be auxin-related, where a number of the highly associating GEMs were found to correspond to auxin-responsive genes (Figure 4.18B). One such GEM, B_comp8374_c0_seq1, corresponds to a gene in rice, Os09g3733, encoding an auxin-responsive SAUR, *OsSAUR39*. In addition, B_comp9352_c0_seq3 corresponds to an *OsSAUR45* gene (Os09g37400). The transcript abundances of these GEMs both correlate negatively with plant height with an R^2 of 0.40 (1.09E-06) for B_comp9352_c0_seq3 and an R^2 of 0.37 (P=1.85E-06) for and B_comp8374_c0_seq1. In rice, *OsSAUR39* has been found to cause reduced auxin synthesis and transport. When over-expressed, this led to reduced plant growth (Kant et al., 2009). The finding that the auxin-related GEMs detected for this trait correlate negatively with plant height, agrees with this (Harper et al., Under review).

In addition highly significant single GEM associations were detected for plant height. These can be seen summarise in Table 4.4.

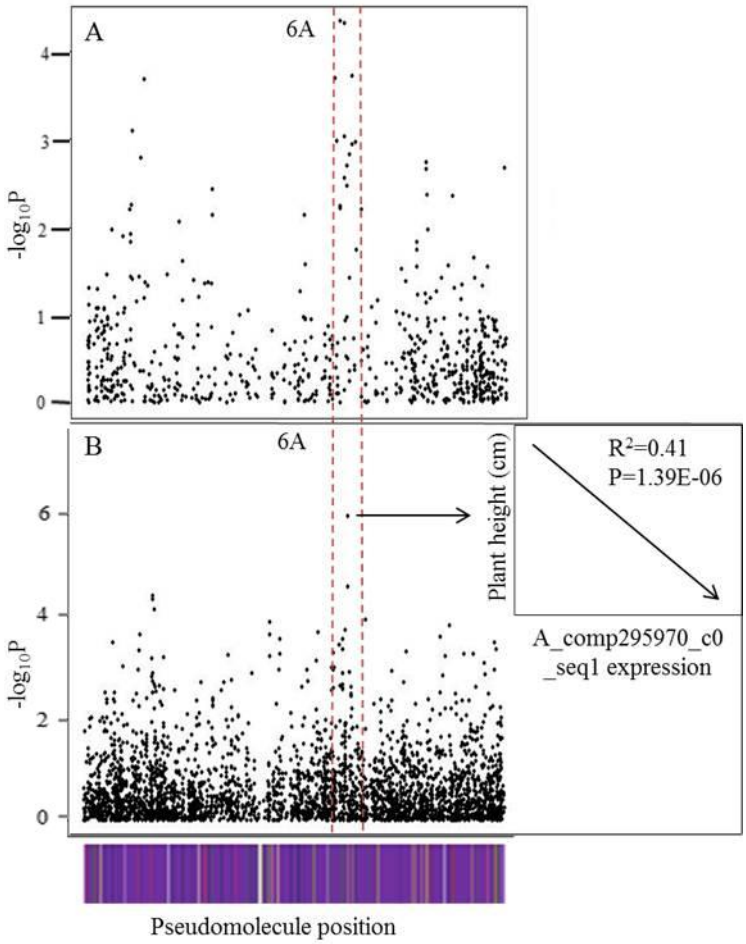

Figure 4. 17. Associative Transcriptomics SNP and GEM analyses results for plant height. Panels A and B show SNP association peaks on 6A and 5B. Panels C and D show corresponding GEM peaks in these regions. Panel E shows a further SNP peak on 2B. SNP and GEM associations can be seen within the dotted border. Marker associations are plotted as –log10P against pseudomolecule position.

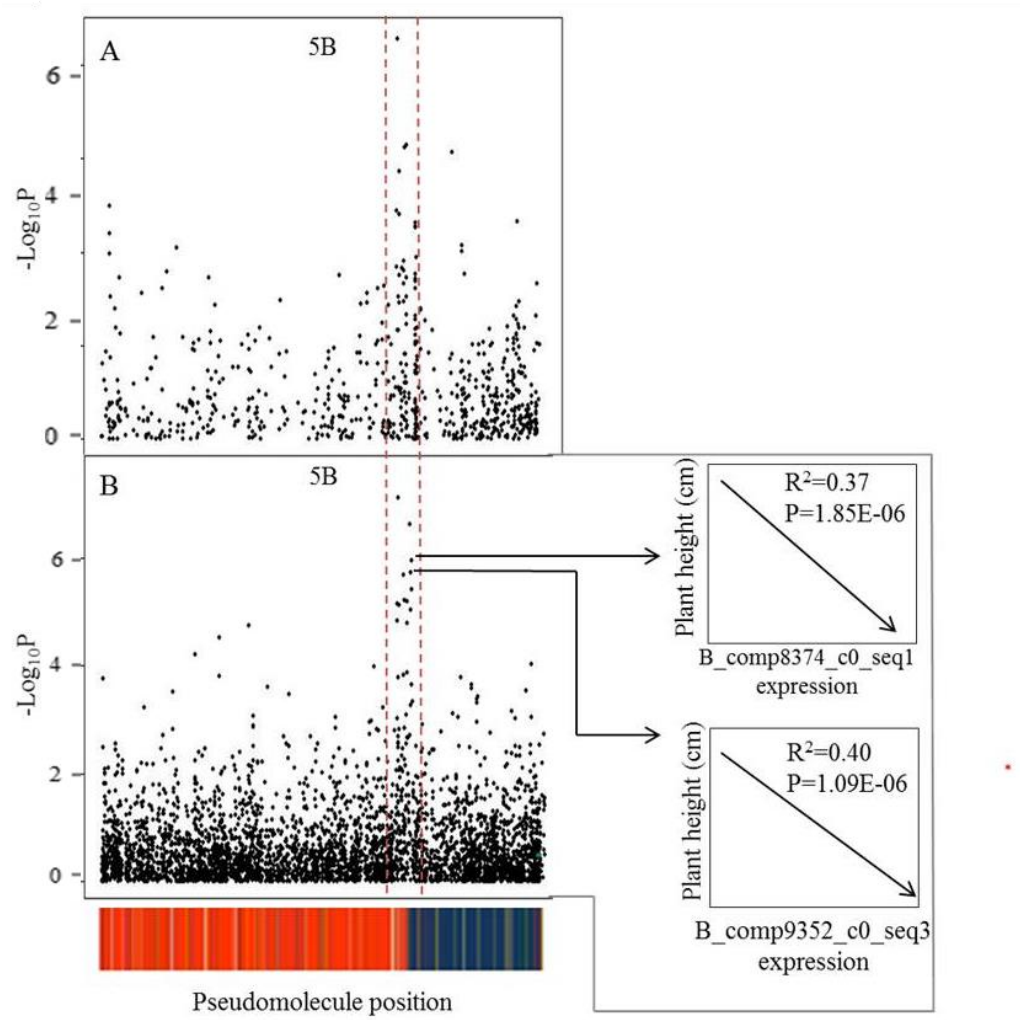

Figure 4. 18. SNP association detected on chromosome 5B (A). This region was also detected in the GEM analysis (B). The transcript abundance of the associating GEMs in this region, correlate negatively with plant height. This includes B_comp8374_c0_seq1 (OsSAUR39) and B_comp9352_c0_seq3 (*OsSAUR45*). The SNP and GEM associating regions can be seen within the dotted border. Marker associations are plotted as –log10P against pseudomolecule position.

4.2.2.4 Main stem threshed weight

For main stem threshed weight, SNP associations were detected on chromosome 6A (Figure 4.19A) and 5B (Figure 4.19B marked with a red asterisk) with the most highly significant markers being A_comp20392_c0_seq1:770 (P value = 1.40E-04, trait effect 16.5 %) and B_comp3140_c0_seq1:1427 (P value = 6.63E-04, trait effect 23.2 %) respectively. Both of these regions were also detected for plant height, a trait which is closely related to stem weight (as seen in Section 4.1). In addition, associations were detected which were not identified for plant height. This included a SNP association on chromosome 3D (Figure 4.19C), with the top marker being D_comp1133_c0_seq1:932 (P=8.57E-04, trait effect 19%). A second association was also detected on chromosome 5B (Figure 4.19B marked with a green asterisk), with the top marker being B_comp6903_c0_seq5:522 (P=4.99E-04, trait effect 13.4 %). No likely candidate genes were detected for these two associating regions. This 5B region was also detected in the SNP analyses carried out for stem absolute strength as was the previously described 6A region.

Significant single GEM associations were also detected for this trait. An example of this is D_comp23979_c0_seq19 detected on chromosome 5D (P=1.07E-06, R^2=0.37). This marker corresponds to an orthologue of a *CESA9* gene in rice, Os09g25490, known to have a role in cellulose biosynthesis. This CESA subunit is however thought to be specific to seed development and therefore its role in determining stem weight (if any) may not be related to the levels of cellulose in the stem material, but may be determined much earlier during seed development.

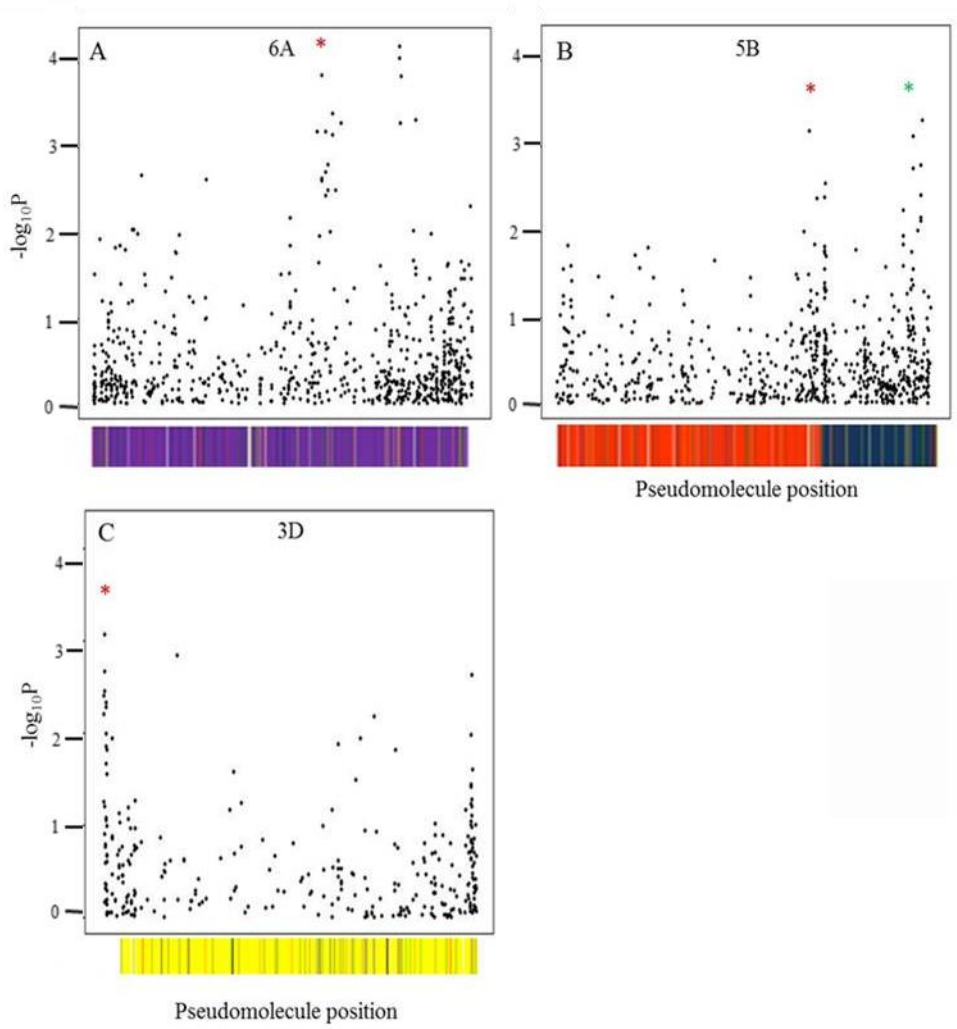

Figure 4. 19. Associative Transcriptomics SNP and GEM results for threshed stem weight. Panels A shows a SNP association on chromosome 6A (marked with red asterisk). Two SNP association peaks were also detected on chromosome 5B (marked with red and green asterisk)(B) and one on chromosome 3B (marked with red asterisk) (C). Marker associations are plotted as –log10P against pseudomolecule position.

4.2.2.5 Grain weight

A very clear SNP association peak was detected for main stem grain weight on chromosome 3D (see Figure 4.20A). The top marker within this peak is D_comp9141_c0_seq:355 (P=1.60E-04, trait effect 31.7 %). In addition, several highly associating GEMs were detected for this trait. One such GEM association, A_comp18136_c2_seq1 (R^2=0.34, P=3.36E-06), can be seen on chromosome 4A (marked with a red asterisk on Figure 4.20B). The transcript abundance for this

marker correlates positively with grain weight. To further assess the role of the differential expression of this locus in contributing to variation in grain weight, the transcript abundance values (measured as RPKM) for A_comp18136_c2_seq1, were mapped as a trait against the SNP data. This analysis revealed a weak signal on chromosome 3D (Figure 4.20C), the same region previously described for grain weight SNP analysis. Unfortunately, no annotation is available for the sequence to which the GEM marker corresponds. It is therefore not possible to propose a biological explanation for the observed interaction between this GEM and the 3D locus for grain weight. Information for all remaining GEM associations can be seen summarised in Table 4.4.

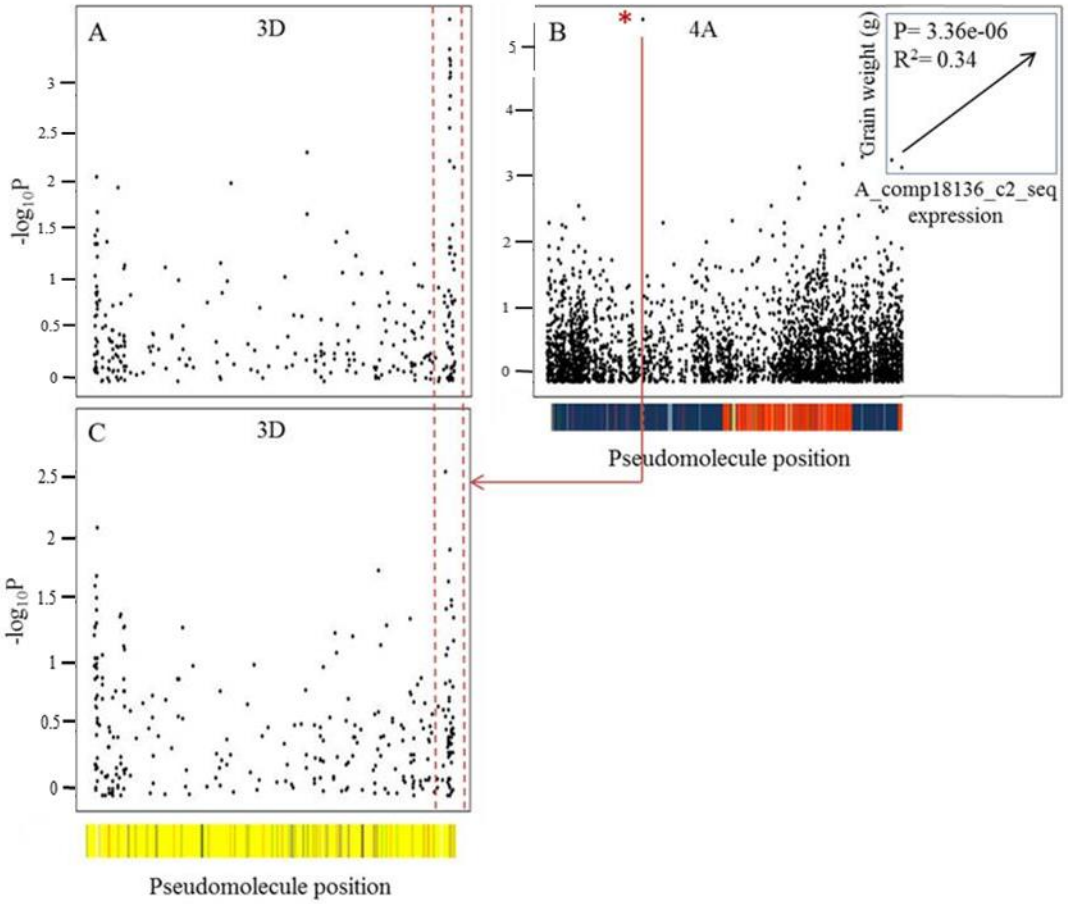

Figure 4. 20. Associative Transcriptomics SNP and GEM results for main stem grain weight. Panels A shows a SNP association peak on chromosome 3D (within dotted border). Panel B shows a highly associating GEM marker, A_comp18136_c2_seq1 on chromosome 4A (marked with red asterisk). Panel C shows the mapping of the RPKM values for this GEM as a trait against the SNP data. This revealed a SNP peak on chromosome 3D (within dotted border). Marker associations are plotted as –log10P against pseudomolecule position.

4.2.2.6 Parenchyma area

For parenchyma area, a single SNP association was detected on chromosome 5B (Figure 4.21A). The most significant marker within this association peak is, B_comp3333_c0_seq2:814 (P=5.75E-05, trait effect 22.6 %). This region was also detected in the SNP analyses carried out for main stem threshed weight and stem absolute strength. Several highly associating GEM markers were also detected for this trait. An example of this can be seen plotted in Figure 4.21B where a GEM

association was detected on chromosome 5B in the same region identified in the SNP analysis (marked with a red asterisk in Figure 4.21B). The most significant GEM marker within this region is B_comp22750_c0_seq4 (P=1.19E-04, R^2=0.28). The transcript abundance levels for this locus correlate positively with parenchyma area. This GEM corresponds to a gene of unknown function. An additional GEM, D_comp23979_c0_seq19, was detected on chromosome 5D (P= 6.91E-05, R^2=0.26), which corresponds to an orthologue of a rice *CESA9* gene (Os09g25490). The transcript abundance levels for this locus correlate positively with this trait. This marker was also detected in the GEM analysis for main stem threshed weight.

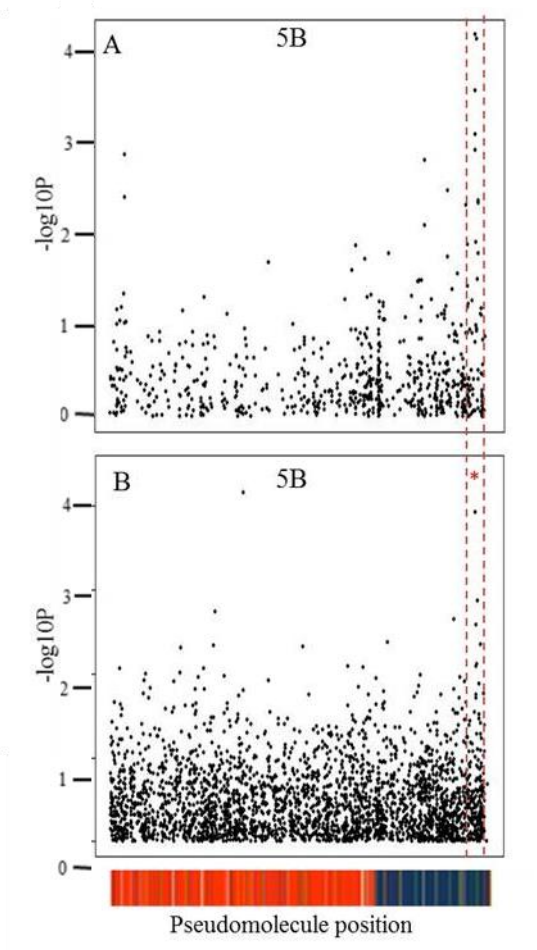

Figure 4. 21. Associative Transcriptomics SNP and GEM results for parenchyma area. Panel A shows a SNP association on 5B. Panel B shows a GEM association in the same region. The region of interest can be seen within the dotted border. Marker associations are plotted as –log10P against pseudomolecule position.

4.2.2.7 Stem hollow area

Very few SNP associations were detected for stem hollow area. A weak signal was seen on chromosome 1B (Figure 4.22), with the top marker being B_comp2073_c0_seq1:412 (P=4.05E-03, trait effect 18.7 %). This region (although the most highly significant marker is different), was detected for both the absolute and material strength traits. Due to the high level of noise in the regress plots, it is difficult to determine which GEM associations to focus on. GEM markers of potential interest can however be seen in Table 4.4.

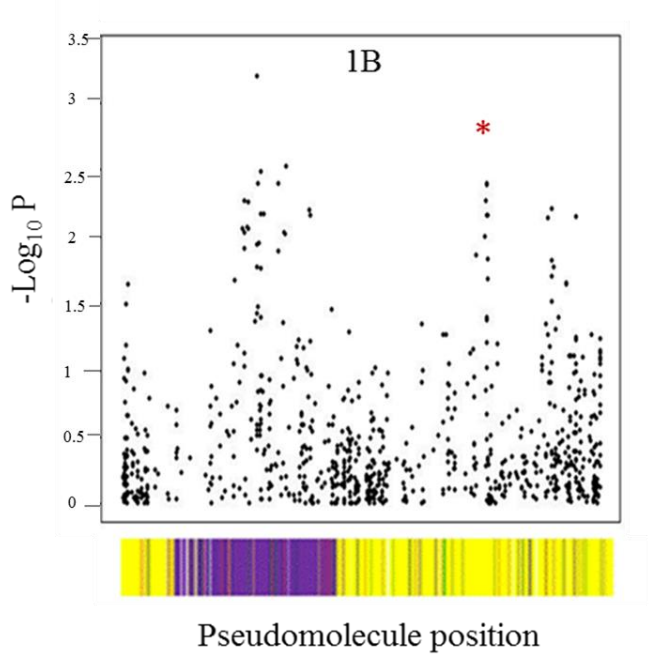

Figure 4. 22. A SNP association (marked with a red asterisk) detected on chromosome 1B for stem hollow area. Marker associations are plotted as –log10P against pseudomolecule position.

4.2.2.8 Outer cortex thickness

As previously discussed, stem outer cortex was seen to display non-normal trait residuals. The Associative Transcriptomics analysis was carried out on both non-transformed and log_{10} transformed data for this trait. For both datasets, only a single SNP association was identified. This association was on chromosome 6A, in the same region detected for plant height and addition traits, including Fmax. In the analysis carried out on the non-transformed data, the most significant marker was

A_comp10724_c0_seq1:311 (P=7.9E-04, trait effect 21.3%). In the analysis carried out on log_{10} transformed trait values, again, A_comp10724_c0_seq1:311, was seen to be the most significant marker (P=9.11E-04, trait effect 20.5%). (Figure 4.23 shows the results of the non-transformed data).

Highly associating GEM markers were also detected. One GEM, A_comp21739_c0_seq1 (P=7.63E-04, R^2=0.19), detected on chromosome 2A, corresponds to an orthologue in rice encoding a gibberellins receptor *GID1L2* (Os07g44890). This may be considered as a plausible candidate for stem outer cortex thickness, given the known role of gibberellins in plant stem development (Xu et al., 1997). A second GEM, D_comp444927_c0_seq1 (P=5.02E-04, R^2=0.21), detected on chromosome 5D, corresponds to a gene orthologous to a glycosyltransferase gene in rice (Os03g55040). Several glycosyltransferase mutants have been described in Arabidopsis. A common phenotype seen in these mutants is reduced secondary cell wall thickening (Lee et al., 2010) which could alter overall outer cortex thickness. Both of these GEM associations were seen only in the analysis carried out on the non-transformed data. Table 4.4 provides a summary of additional GEM associations detected for both the non-transformed and transformed data for this trait.

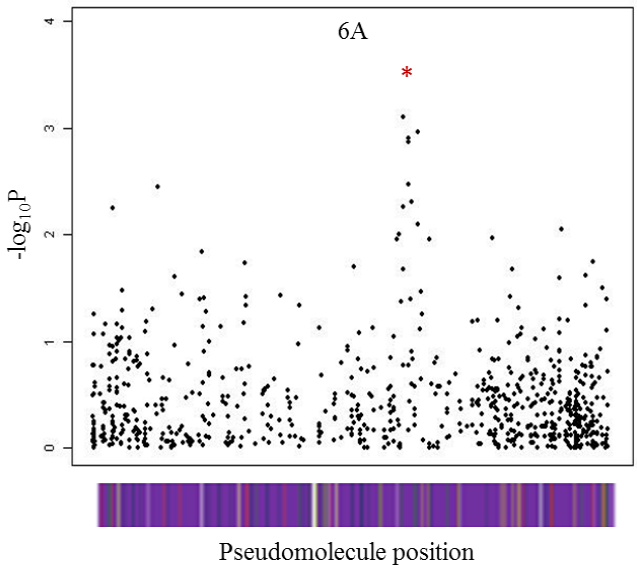

Figure 4. 23. A SNP association (marked with a red asterisk) detected on chromosome 6A for stem outer cortex thickness. Marker associations are plotted as –log10P against pseudomolecule position.

4.2.3 Discussion

Through Associative Transcriptomics several very interesting associations have been detected in both the SNP and GEM analyses. In addition, owing to the sequence similarity to species for which genome annotations are available, it has been possible to propose several potential candidate genes for the stem mechanical and related traits analysed in wheat. In some cases, SNP variation within a potential candidate gene was found to associate directly with the trait of interest. An example of this was seen for the 1B association detected for the material strength traits, where a gene involved in xylan acetylation was found to contain the most highly significant SNP for this association. In other cases, plausible candidate genes were found in close proximity to the most highly associating SNP marker. For example, the SNP marker, D_comp970_c0_seq1:1030, also detected for the material strength traits, was found in the pseudomolecule to be just 6 unigene models away from a potential candidate gene (described in rice as an acetyl xylan esterase). It is of course also possible that the genes proposed as candidates are in fact not the causal genes responsible for the observed associations. This illustrates a great limitation of such analyses, where we are potentially limited by the knowledge of gene function so far achieved in the scientific community. There may in fact be an additional candidate gene located in closer proximity to the most highly associating marker, the potential role in mechanical strength of which has not yet been described. Although not essential for marker assisted selection, it is of course useful to know the causal variation underlying any detected marker associations. For example, it may, through breeding, be realised that the selection of a given gene, may have undesirable pleiotropic effects. With knowledge of the causal variation being selected for, it would be possible for the breeder to evaluate any potential pleiotropy and to decide whether or not the developed marker is appropriate for their breeding aims.

In addition to the results obtained through the SNP analysis, many candidate genes were also identified in the GEM analysis. These associations were seen in two different forms. Firstly, a number of single GEM associations were seen. Given that these represent an association between the expression of a single gene and the target trait, these marker associations are expected to be causative. In some cases, these single GEM associations have identified genes which, based on previously described function in other plant species, can be plausibly implicated in the control of stem

mechanical strength. An example of this can be seen with the GEM detected for the material strength traits which corresponds to an orthologue of the *GLYCOSYLTRANSFERASE 8* gene in rice. The Arabidopsis homolog of this gene, *QUA1,* is, as previously mentioned, known to be involved in cell wall biosynthesis. Another example can be seen with the identification of auxin response genes associating with plant height. In addition to these GEMs found to correspond to genes which have clear potential roles in determining the trait variation, many GEMs not previously identified as candidates for the relevant traits were also detected. These may therefore represent novel regulators for the target traits.

In addition to the single GEM associations, GEM peaks were also identified. These signals represent a more widespread pattern of differential expression across a given genomic region. Given the large number of genes which may contribute to this signal, this type of signal may be indicative of a deletion or duplication of a region containing multiple genes (Harper et al., 2012) . These GEM association peaks may also be the result of other mechanisms known to alter the gene expression level of linked genes, such as DNA methylation. An example of this type of signal was seen on chromosome 2D for material strength.

One very exciting aspect of the described association analysis, was the identification of possible interactions between loci. An example of this was seen for the material strength traits, where when mapped as a trait in the SNP analysis, the expression levels of a number of single associating GEMs were found to map to a region on chromosome 2D, already detected in the SNP and GEM analyses of the material strength traits. The same was also seen for grain weight where the expression values of a single GEM association detected on chromosome 4A was found to map back to a region on chromosome 3D, which was picked up in the SNP analysis for this trait. In the case of the result obtained for material strength, this analysis allowed for the detection of a potential cell wall remodelling pathway which may be important in determining the material strength of wheat stems.

The inclusion of several traits found to correlate to mechanical strength in the association analysis has proven very valuable. In several cases, common loci have been identified for traits which were also found to be correlated. An example of this can be seen with the many associations detected for a specific region on

chromosome 6A. This region was detected for several traits, including plant height, outer cortex thickness, main stem threshed weight and the absolute strength traits Fmax and F/V. For the majority of these traits, the most highly significant marker was found to be A_comp20392_c0_seq1:770 and in all cases the increasing allele was found to be "A" and the decreasing allele "T". As described in Section 1.2.2, reducing plant has been a commonly employed strategy for reducing lodging susceptibility. Given that this locus was found to associate to absolute strength (as well as structural traits positively correlating to strength), it may be that selecting for the increasing allele at this locus, would result in tall plants which are less susceptible to lodging due to an increase in outer cortex thickness, stem density and improved absolute strength. This would of course depend on which factor is contributing most significantly to lodging in the accession of interest. It is however encouraging to see that loci which are more specifically related to mechanical strength and do not seem to contribute to plant height were also identified, suggesting that it is possible that these potentially antagonistic traits can uncoupled for the effective improvement of lodging resistance in wheat.

For plant height, very clear marker associations were identified. However, it would have been expected, that loci in close proximity to Rht-genes (dwarfing genes which have been introduced to many wheat accessions through breeding) would have been identified for this trait. However, within the panel of wheat accessions included in the Associative Transcriptomics analysis, *Rht-B1* is known to be segregating in only a few accessions. Furthermore no SNPs were identified in the region of *Rht-D1*. Given this, it is not surprising that associations for these expected loci were not seen (Harper et al., Under review). Despite this, it has been possible to detect what seems to be a second level of regulation for plant height, which may not have been detected if variation relating to the Rht-loci had been detected.

In Section 4.1, when exploring the relationships between stem mechanical strength and chemical composition, very few significant correlations were seen between the chemical composition traits and material strength. As previously discussed, this may indicate that it is variations in more fine-scale cell wall attributes (which were not screened in the FTIR analysis) that contribute to the observed variation in stem mechanical strength in wheat. The results obtained through Associative Transcriptomics support this hypothesis. Through Associative Transcriptomics, a

number of genes involved in cell wall acetylation were identified for material strength. Cell wall acetylation is thought to be important in binding the various cell wall components together into a complex and more structurally reinforced matrix (Busse-Wicher et al., 2014). As previously discussed, genes controlling cell wall acetylation have also recently been proposed as potential targets for improving cell wall recalcitrance for lignocellulosic ethanol production (Pawar et al., 2013). However, the data presented here suggests that there may be a trade-off between biomass processibility and stem mechanical strength. Given this potential conflict, (following validation of these genes as causal loci) it would valuable to assess the saccharification efficiency of straw taken from wheat accessions showing marker variation in close linkage with these cell wall acetylation genes.

Overall, the results presented here illustrate that there is great power to detect potentially important variation in wheat through the utilisation of this novel mapping approach, Associative Transcriptomics. The combining of both GEMs and SNPs allows for a great level of information to be gained and promotes a much deeper understanding regarding the potential mechanisms underlying the detected marker associations.

Many studies in the literature report the detection of markers which may be of use in the breeding of crops with improved agronomic performance. However, few studies show evidence of marker durability in trait selection. The following section will describe a marker validation experiment which aimed to test the durability of a subset of markers, detected through Associative Transcriptomics, in selecting for stem mechanical strength.

4.3 WHEAT MARKER VALIDATION STUDY

Through Associative Transcriptomics several loci showing significant associations to the mechanical strength and related traits were uncovered in wheat. As described previously, it is important to test the durability of such markers for their use in marker-assisted breeding. To achieve this in the present study, wheat accessions of previously unknown genotype (grown as part of the WAGTAIL panel which consists of elite winter wheat breeding material) were screened for variation at a subset of marker loci detected through Associative Transcriptomics for the material strength trait, MOR. Following mechanical testing, it is then possible to assess marker-trait

segregation patterns to determine whether or not the results obtained corroborate with those seen in the Associative Transcriptomics analysis.

4.3.1 Methods

4.3.1.1 Plant material and harvest

96 wheat accessions were harvested from the WAGTAIL panel grown at KWS, Thriplow in the summer of 2013. For each accession, five plants were harvested. Harvesting was carried out as described in Section 4.1.1.1. Any plants noticed to exhibit stem damage were omitted from further analyses.

4.3.1.2 Leaf sample collection and DNA extraction

Prior to harvesting, green leaf samples were collected from all 96 WAGTAIL accessions. On collection, the leaf tissue was transferred directly into a 96 tube rack and their position within this rack recorded on a 96 well grid sheet. As described for the *B. napus* marker validation study, this information was retained by a work colleague to ensure sample anonymity. For all subsequent genotyping work, these accessions were identified by their 96-well plate position. Only following mechanical testing were the genotypes and accession names matched. Samples were kept on dry ice throughout the sample collection and transportation back to JIC. DNA extraction was performed on the collected leaf tissue as described in section 2.4.1.

4.3.1.3 Primer design for marker validation in wheat

In order to obtain sufficient information required for high-quality primer design, two main sequences were used. To allow for the detection of the target SNP, transcriptome sequence for at least two wheat accessions expected to vary at the target marker loci were obtained. Given that these sequences represent only the transcribed gene regions, it was also important to gain information regarding the full genomic sequence. This would allow for the detection of any introns within the sequence. This is important to ensure that the primers are not designed across intron-exon boundaries which can reduce primer efficiency. To enable this, the transcriptome sequence of the target region was analysed in BLAST against the wheat (IWGSC) pseudomolecule V3.3 which can be found at: http://n61225.nbi.ac.uk/blast.html. The BLAST output provides information regarding the pseudomolecule position of the sequences producing high BLAST scores. This information allowed for the retrieval of the genomic sequence by imputing the pseudomolecule co-ordinates into the "extract segment" tool which can be found at: http://n61225.nbi.bbsrc.ac.uk/wheat.html. In doing this, it was also possible to detect homoeologous and paralogous sequences present within the wheat genome. To identify primer regions, all sequences were aligned using AlignX.

Unlike the marker validation experiment previously described for *B. napus*, which was based on results obtained from a hemi-SNP association analysis (where SNPs are not genome-assigned), the wheat markers being targeted in this experiment had been successfully assigned to specific genomes. Given this, it was possible to design genome-specific primers. Although as seen for the *B. napus* analysis, it is possible to develop successful marker assays without genome-specificity, the amplification of multiple similar, but not identical gene copies can cause problems. For example, the presence of an insertion or deletion (InDel) can cause sequence misalignment. In addition, there may in some cases be the added complication of interhomoeologue polymorphisms (IHPs). This can make it impossible to distinguish between genotypes through sequencing, if conserved primers are used. An example of this is illustrated in Figure 4.24A, where in addition to a SNP between accessions residing in the D genome, there is also variation between the A and C homoeologues. Although this type of variation is expected to be rare, it is an important genomic feature to consider when designing primers in polyploid species. To avoid this

complication, primers were designed in regions found to vary between the target genome and the two non-target homoeologues. This variation, which may be in the form of a SNP for example, is used as the final base of the primer sequence. To further enhance the specificity of the primers, a single mismatched base was incorporated at the penultimate base of the primer sequence (Figure 4.24B) (Huang and Brule-Babel, 2010). This method was also used in the *B. napus* marker validation study to allow for selection against paralogous sequences. Primer quality was assessed using NetPrimer. Following the successful design of high quality primer sequences, all primers were ordered online from Sigma Genosys.

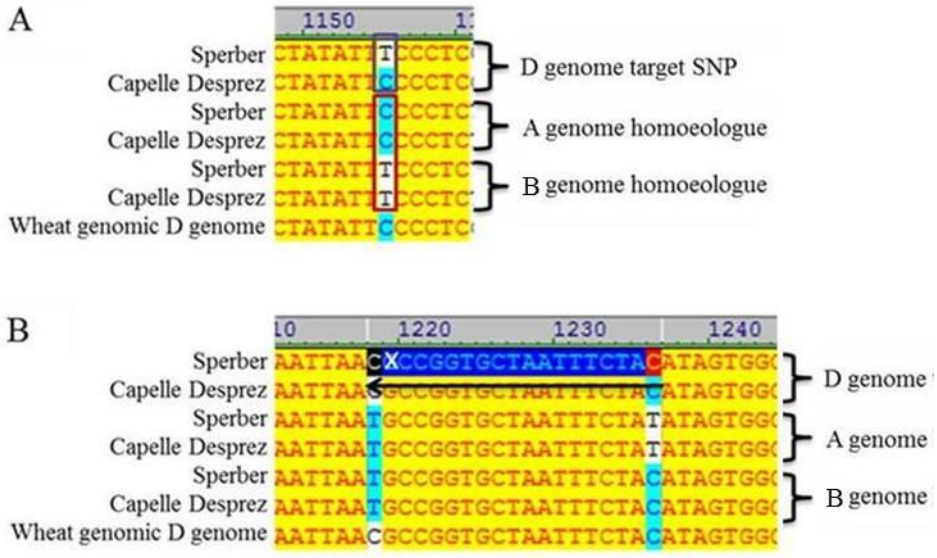

Figure 4. 24. Method used in designing genome specific primers in the wheat marker validation study (B). This was particularly important in cases where the region of target variation (here shown in a black rectangle (A)) was at a locus which also contained an interhomoeologous polymorphism (IHP) which is here shown in a red rectangle (A).

4.3.1.4 Genotyping WAGTAIL wheat accessions for marker variation

All primer combinations were initially tested through gradient PCR. Following this, standard PCR was performed (using the optimised PCR conditions determined through gradient PCR) on six accessions expected, based on the Associative Transcriptomics analysis, to exhibit variation at each of the target loci (three accessions expected to carry the increasing allele and three expected to carry the decreasing allele). Through sequencing, the presence of the expected variation and

the suitability of the markers in screening were assessed. Following confirmation that the marker assays were working effectively, all 96 wheat accessions taken from the WAGTAIL panel were screened for marker variation in the same way. The PCR and sequencing methods used are described in Section 2.4.

4.3.1.5 Screening WAGTAIL wheat accessions for mechanical strength and evaluating marker-trait segregation patterns

A subset of 30 accessions (including representative genotypes for each marker) harvested from the WAGTAIL panel were mechanically tested and processed as described in Section 4.1.1.2.4. The relationship between the uncovered marker variation and the mechanical strength data was then explored using ANOVA. As described previously for *B. napus*, this approach tested the following null hypothesis: There is no difference in trait value between accessions carrying the increasing and decreasing allele for marker *x*. Rejection of this hypothesis suggests that the markers identified through Associative Transcriptomics have trait selection capability.

4.3.2 Results

4.3.2.1 Testing marker assays and genotyping WAGTAIL accessions for marker variation

To assess the durability of the marker associations detected in the wheat Associative Transcriptomics analysis for trait selection, a subset of markers were chosen for inclusion within this predictive breeding study. These markers were B_comp205_c0_seq2:624, D_comp970_c0_seq1:1030 and D_comp1058_c0_seq1:1573, each of which were found to associate with the material strength trait, MOR. Through BLAST analysis, for each of the three markers, as expected, 3 homoeologous regions were detected, one on each of the A, B and D genomes. No highly significant paralogous sequences were identified. Using the mismatch primer-design method described, genome specificity for each of these marker loci was achieved. This was particularly important for D_comp970_c0_seq1:1030, due to the presence of an IHP, which as previously discussed, can complicate genotyping. This can be seen illustrated in Figure 4.24. In this example, the targeted marker variation relates to just the D genome (Figure 4.24A), where the segregation of a “T” allele and a “C” allele (shown in a black rectangle) can be seen. However, the screening of this marker variation using conserved primers would result in the concealment of this marker variation due to

the presence of an additional “T” allele in the B genome and an additional “C” allele in the A genome in each accession (leading to the presence of both “T” and “C” signals for all accessions in the sequencing reads). This IHP can be seen marked with a red rectangle (Figure 4.24A). This complication was avoided through designing genome-specific primers (See Figure 4.24B).

To assess the suitability of the designed marker assays in amplifying the target marker loci, PCR and sequencing was carried out on accessions of known genotype (included in the Associative Transcriptomics analysis) expected to vary at the targeted marker loci. This was carried out as described in Section 4.3.1.4. Figure 4.25 shows an example of this for D_comp1058_c0_seq1:1573. Figure 4.25A shows the gradient PCR used to determine the appropriate annealing temperature of the primers used. The expected annealing temperature calculated using Net Primer for the reverse primer was 65 °C. For the reverse primer we expected an annealing temperature of 58 °C. To assess the ideal annealing temperature of this primer combination, a gradient of 55-67°C (3ºC below the lowest tm +12ºC) was run with 30seconds extension time (according to the alignments made we expect a PCR product of approximately 280bp was expected and therefore 30second extension is more than adequate). As can be seen, good PCR amplification was achieved at the lowest annealing temperature 55ºC. This annealing temperature was then used to achieve amplification in six wheat accessions from the wheat Associative Transcriptomics panel expected to vary at the target loci (Figure 4.25B shows an example of two accessions showing marker variation at the transcriptome level). Through capillary sequencing this target variation was confirmed. This can be seen from Figure 4.25C and D where a comparison between a Clustal alignment of the transcriptome sequence and the genomic sequence obtained through capillary sequence can be made.

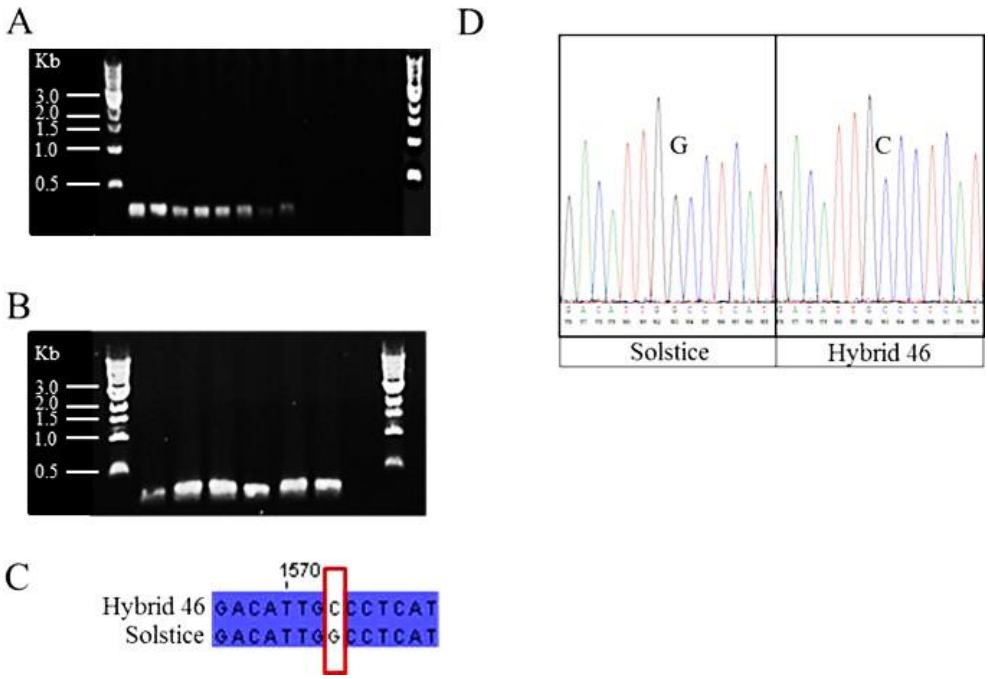

Figure 4. 25. Successful confirmation of target allelic variation for wheat marker D_comp1058_c0_seq1:1573. Through gradient PCR (A) (with a Tm° range of 55-67 in this example), an appropriate annealing temperature was determined (55°C in this example). This allowed for the successful amplification of the target region through standard PCR on six accessions expected to exhibit the target marker variation as determined through transcriptome sequencing (B) (1 Kb ladder used). Through capillary sequencing, it was then possible to compare what would be expected according to the transcriptome data (C) and what was seen in the sequence trace files at the genomic level (D).

Despite this method working very effectively for D_comp1058_c0_seq1:1573, the sequencing results for the remaining two markers were continually inconclusive, despite testing multiple primer combinations. For D_comp970_c0_seq1:1030, despite selecting against homoeologues (which in this case was especially important due to the presence of the previously described IHP), no variation could be detected at the target loci. This suggests that, a non-target homoeologous copy of this region may have been (for some reason not successfully accounted for) preferentially amplified. For B_comp205_c0_seq2:624, the sequence trace files consistently showed "mixed" reads, making it difficult to accurately determine the marker genotype. This may have been due to the presence of unaccounted for similar sequences in the wheat genome. Due to time constraints, alternative markers were

selected for the association peaks to which B_comp205_c0_seq2:624 and D_comp970_c0_seq1:1030 correspond. The first marker, found to be the second most highly associating marker for MOR in the association peak detected on 1B was B_comp2391_c0_seq1:284. This marker was found to have a P value of 1.24E-05 and a trait effect of 20.5 %. Through BLAST analysis, this new target region was found again to have three homoeologous copies. This was accounted for in the primer design, as described for D_comp1058_c0_seq_1:1573 and through PCR and sequencing, the target variation at this alternative locus was revealed.

The alternative marker chosen for the 2D association (in place of D_comp970_c0_seq1:1030) was D_comp19374_c0_seq1:702. This marker was found to have a P value of 5.18E-03 and a trait effect of 9.1 % for the material strength trait, MOR. Unlike, the previously described loci, this marker-containing region was found to have just two highly similar copies, with no A-genome homoeologue detected. Again, through PCR and sequencing of this alternative marker, it was possible to reveal the target variation.

The marker assays used in the successful amplification of the desired target variation for three of the association peaks detected through Associative Transcriptomics are summarised in Table 4.5. This Table provides a summary of the primers used, including the primer sequences; expected product size and the optimal annealing temperatures used to achieve PCR amplification. This Table also gives information regarding the specificity of the primers used.

Following the confirmation that each marker assay was suitable for use in screening the target variation at the three marker loci, each assay was used in determining the variation present within the panel of 96 wheat accessions from the WAGTAIL panel. In all cases this was successfully achieved. Table 4.6 gives an overview of the variation uncovered for each of the three marker loci included in this study (see Supplementary Data file 4.3a for genotype information across accessions). As can be seen, for the marker, D_comp1058_c0_seq1:1573, only one accession was found to carry the increasing allele. Given this, it was not possible to accurately test the durability of this marker for trait selection within this panel of wheat accessions. This marker was omitted from all further analyses. Slightly more variation was detected for B_comp2391_c0_seq1:284, with 9 accessions carrying the increasing

allele and 80 carrying the decreasing allele. However, again, the decreasing allele is markedly more common. A much more promising level of allelic variation was however detected for D_comp19374_c0_seq1:702, where the number of increasing and decreasing alleles detected was 59 and 28 respectively.

Table 4. 5. Primer assays shown to be successful in amplifying target sequences in wheat marker validation study.

Marker ID	Increasing allele	Decreasing allele	Forward primer	reverse primer	Homoeologues selected against	Tm°	Product size (bp)
D_comp1058_c0_seq1:1573	C	G	GGAGGTCATCATACAAGGTTA	GAGGCTCGAGCTAACCAACC	A and B	55	245
D_comp19374_c0_seq1:702	T	C	GCTGCACCATCTCCTATCAC	CCTTGAGTGCGCAGATTATG	B	55	380
B_comp2391_c0_seq1:284	C	T	GAGACGAGCAACAGCATGC	GGAAACGGTATCAAGACATATGT	A and D	50	310

Table 4. 6. Marker variation uncovered across 96 wheat accessions grown as part of the WAGTIAL panel at KWS, Thriplow.

Marker allele	D_comp1058_c0_seq1:1573	D_comp19374_c0_seq1:702	B_comp2391_c0_seq1:284
Increasing	1	59	9
Decreasing	91	28	80
Sequencing failures	4	9	7

4.3.2.2 Mechanical testing of WAGTAIL wheat accessions and assessing marker-trait segregation patterns

Based on the results obtained through Associative Transcriptomics, it would be expected, that accessions carrying the increasing allele for associating markers, would exhibit higher material strength than those carrying the decreasing allele. It would also, based on the allele effects obtained through Associative Transcriptomics, be expected that a greater difference in MOR will be seen between accessions carrying the increasing and decreasing alleles for B_comp2391_c0_seq1:284 (found to have a trait effect of 20.5%), in comparison to D_comp19374:c0_seq1:702 (found to have a trait effect of 9.1%).

To assess these predictions, 30 ASSYST accessions (see Supplementary Data file 4.3a for a list of the chosen accessions) were mechanically tested according to the methods described in Section 4.1.1.2.4 and the trait values obtained analysed through ANOVA using the methods described in Section 4.3.1.5. As seen in the wheat Associative Transcriptomics panel, non-normal trait variance was seen for MOR. Following a $\log_{10}$ transformation, great normality was achieved. The transformed trait values were used in the ANOVA analyses. Figure 4.26 provides an overview of the mean trait values observed for each marker. For B_comp2391_c0_seq1:284, accessions carrying the increasing allele exhibit a mean MOR of 10.35 N/mm^{-2} and those carrying the decreasing allele have a mean MOR of 5.40 N/mm^{-2} . Through ANOVA, this difference was found to be statistically significant (P=0.006). These findings suggest that this marker has strong trait selection capability for the material strength trait, MOR. This is particularly promising given the relatively low level of variation observed at this locus in the WAGTAIL panel.

The second marker, D_comp19374_c0_seq1:702, was also found to have selection capability for this trait. Accessions carrying the decreasing allele at this locus had a mean MOR of 4.79 N/mm^{-2}. Those carrying the increasing allele had a mean trait value of 7.68 N/mm^{-2}. Through ANOVA, this difference was found to be statistically significant (P=0.046).

As expected, based on the results obtained from the Associative Transcriptomics, higher trait values were observed in accessions carrying the increasing alleles at these marker loci. In addition, as expected, B_comp2391_c0_seq1:284 was seen to

make the greatest contribution to the material strength trait, MOR, in comparison to D_comp19374_c0_seq1:702. These findings agree with the predictions made (see Supplementary Data file 4.3b for trait summary and ANOVA outputs).

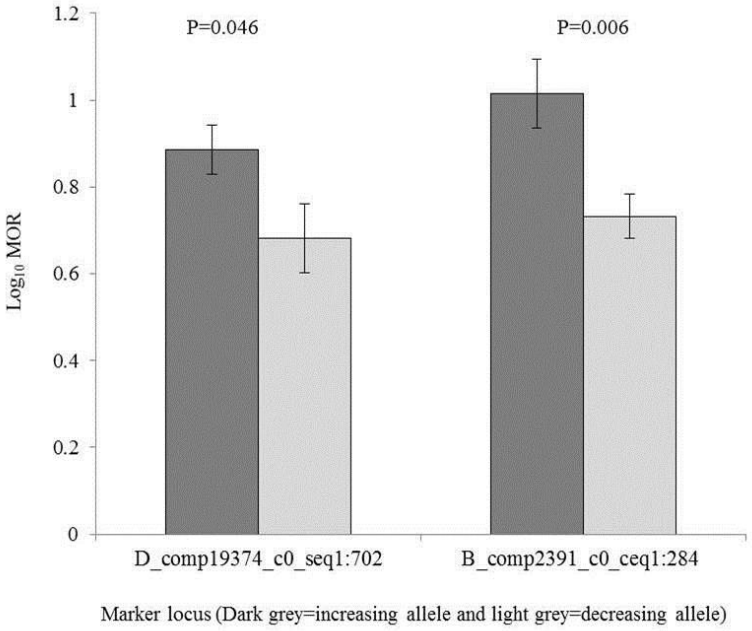

Figure 4. 26. Variation in MOR seen across 30 WAGTAIL accessions grown at KWS, Thriplow for two marker loci, B_comp2391_c0_ceq1:284 and D_comp19374_c0_seq1:702. Error bars represents standard error of the mean as calculated through ANOVA.

4.3.3 Discussion

This section has described the results of an experiment carried out to assess the trait selection capability of markers for stem mechanical strength identified through Associative Transcriptomics. The methods employed have proven effective in uncovering the targeted marker variation in this polyploid species.

Both markers for which a good level of variation was observed in the WAGTAIL panel were found to have statistically significant trait selection ability for MOR. The marker seen to have the greatest trait effect was, as expected, B_comp2391_c0_seq1:284. This agrees with the results obtained from the Associative Transcriptomics and suggests that this method is robust for the detection of durable markers for marker-assisted breeding in wheat. As previously discussed, B_comp2391_c0_seq1:284 and D_comp19374_c0_seq1:702 were not the most highly associating markers detected in their respective association peaks. Following the successful development of marker assays for these more highly associating markers, increased trait selection power may be achieved.

Unfortunately, due to a lack of allelic variation observed in the WAGTAIL panel, it was not possible to explore the durability of the final marker D_comp1058_c0_seq1:1573 in selecting for stem material strength in wheat. It would be interesting to assess the variation at this locus in a larger panel of accessions. The wagtail population is made up of elite UK winter wheat accessions. The panel utilised in the Associative Transcriptomics analysis, although also includes some recent UK winter wheat accessions, includes accessions taken from other parts of Europe, some of which represent very old breeding material. It is therefore possible that the presence of the increasing allele of this marker carries some selective disadvantage which is not desirable in UK elite winter wheat breeding and has been selected against. It may for example, be that this allele, although beneficial for the selection of increased stem mechanical strength, is linked to additional loci which are detrimental for a trait of greater agronomic importance. In this case, it is likely, that through "linkage drag," this variation would have been indirectly selected against. When looking more closely at the accessions included in the Associative Transcriptomics analysis which carry the increasing allele (C) for this marker, it was found that in the majority of cases, it is seen in non-UK, old breeding material. For example, the increasing allele for

D_comp1058_c0_seq1:1573 can be seen in the winter wheat accession “Alba”. This is a Belgian accession which was introduced in 1928. Another example can be seen with “Muck”, a German winter wheat accession introduced in 1962.

The region detected on 2D for which D_comp1058_c0_seq1:1573 was the most highly significant SNP marker, was also detected in the GEM analysis, where a more significant association was observed in comparison to that of the SNP analysis. It may therefore be more appropriate to screen the variation at this locus at the expression level. Unfortunately, the assessment of this level of variation was beyond the scope of the present study, but it may be considered a key area for future investigation.

The approach taken in this marker validation study has proven very powerful. It has shown that it is possible, through Associative Transcriptomics to identify marker associations which can be successfully used in the selection of stem mechanical strength in wheat.

4.4 CHAPTER SUMMARY

This Chapter has described the work carried out to develop an improved understanding of stem mechanical strength in hexaploid wheat. In Section 4.1, stem mechanical and related traits were explored for variation across 100 wheat accessions. Through statistical analyses, it was revealed that in most cases, there was a significant interaction between genotype and year. To allow for an overview of trait variation without the confounding effect of this interaction, REML analysis was used to predict trait means. This analysis revealed that for most traits, a high level of genetic variation exists within this panel, suggesting that there is good potential for the improvement of these traits through breeding.

Through correlation analysis, interesting relationships between stem mechanical strength, plant morphological and stem structural traits were identified. This analysis revealed, that stem weight and stem hollow area make the largest structural contributions to stem absolute strength in wheat. In addition, stem outer cortex and parenchyma area also seem to be of importance. In addition, stem material strength was seen to contribute significantly to overall stem strength. These traits may therefore be seen as key targets for the breeding of improved stem mechanical

strength in wheat. Unlike what has been previously reported in the literature (Kashiwagi et al., 2008), no relationship was observed between stem width and stem mechanical strength

Using data provided by the lab of Professor Keith Waldron, it was possible to explore the relationship between stem mechanical strength and chemical composition for subset of wheat accessions. This analysis revealed that while chemical component traits correlate with structural and absolute strength, very few chemical contributions to material strength were detected. The observed relationships between stem structure and absolute strength are likely to reflect the relative proportions of particular tissue types which themselves vary in chemical composition and also contribute to stem structural and stem absolute strength. For example, a positive correlation was observed between stem outer cortex and absolute strength. The outer cortex will likely have a very specific chemical composition, unique to that tissue type. This composition will be more highly represented in accessions with a thick outer cortex in comparison to those which have a thin stem wall. Given that outer cortex thickness relates to stem strength, the chemical components making up this tissue type will also tend to show a relationship. The lack of correlation observed between stem material strength and chemical composition, may suggest that it is the finer-scale chemical components (which were not incorporated into the developed FTIR models) which are of importance in contributing to the variation in stem material strength in wheat.

Given the low levels of stem lodging observed in the field, and the inherent error in screening this trait naturally under field conditions, a measure of stem lodging risk was obtained experimentally across a subset of 65 wheat accessions. This experiment was carried out using a pulley system, which enabled a score of stem breakage to be obtained under field conditions. Exploring the relationship between this trait and the mechanical strength traits obtained from the three-point bend test experiment, revealed that stems more likely to break in the field, exhibited low stem absolute and material strength *in vitro*. This suggests that the methods employed in the present study have been effective in uncovering important variation for plant standing ability in field conditions. This adds evidence to suggest that understanding the genetic control of these traits may be of key interest for the improvement of lodging resistance in wheat.

To obtain a thorough overview of stem mechanical strength in wheat, all mechanical traits and any additional traits found to correlate with stem strength, were further analysed through Associative Transcriptomics. This analysis was carried out on REML predicted means, allowing for the maximum amount of information to be incorporated into a single analysis. This analysis revealed several interesting loci. For the material strength traits, MOE and MOR, loci in close proximity to genes involved in the degradation of xylan (the major hemicellulosic component in monocot primary and secondary cell walls (Vogel, 2008)) side groups were detected. The first of these candidate genes was detected on chromosome 1B, the orthologue of which has, in rice, been described as a GDSL-like Lipase/Acylhydrolase superfamily protein (AT1G54790). The second candidate gene is an orthologue of a gene which in rice has been described as an Acetyl xylanesterase (AXE) (0S04g01980). These genes are also known to be involved in the hydrolysis of xylan side chains. The detection of these candidate genes in close proximity to associating markers, which have similar biological function, is very interesting and suggests that this process may be a key contributor to the variation in material strength observed between wheat accessions. As previously discussed, this should be considered when aiming to improve feedstock recalcitrance for lignocellulosic ethanol production, where reduced xylan acetylation has been a key focus (Pawar et al., 2013).

In addition to the potential role of xylan side groups in contributing to stem mechanical strength in wheat, the Associative Transcriptomics analysis also revealed a potential importance of the COP9 signalosome. This was proposed based on the identification of differential expression of the *COP9 SUBUNIT 5B* detected in the GEM analyses for MOE and MOR. Additional evidence was gained following the finding that the transcript abundance (measures as RPKM) of a gene, *EIF-2B*, which has in previous studies been found to interact with the COP9 signalosome, were seen to map to the genomic region in which the gene encoding the *COP9 SUBUNIT 5B* resides. In addition to *EIF-2B*, the expression of genes implicated in cell wall biogenesis, also mapped to this region of chromosome 2D when mapped against the SNP data. In fungi, it is thought that the COP9 signalosome is involved in a pathway which promotes cell wall degradation for cell growth (Nahlik et al., 2010). The findings presented in the present study, suggests that there may be a process analogous to this in plants. Given that the expression of the cell wall related genes

and *EIF-2B* correlate positively with that of the *COP9 SUBUNIT 5B*, and given the widely accepted role of the COP9 signalosome in targeting proteins for degradation, it is possible that the COP9 complex is involved in targeting suppressors of cell wall-related genes for degradation. This finding is very interesting and may be indicative of a novel pathway for cell wall biogenesis in plants which has a potentially important role to play in determining stem material strength in wheat. As previously discussed, it would be interesting to assess the expression levels of the cell wall-related genes and *EIF-2B* in a wheat mutant defective in *COP9 SUBUNIT 5B* function. Reduced expression levels in this mutant would provide further evidence of an interaction between these loci. It may also be of interest to assess stem mechanical strength and cell wall composition in this mutant. This could, as previously mentioned, be achieved through TILLING. Functional validation of the proposed xylan-related genes would also be of interest.

In addition to the SNP and GEM associations detected through Associative Transcriptomics for the stem mechanical traits, a number of loci were also identified for stem structural traits which were found to correlate with stem strength. The inclusion of multiple traits which may each contribute to stem strength in wheat, may through breeding allow for the development of a wheat ideotype for lodging resistance.

As described in Section 4.2.2, one aspect of association studies which of current debate, is the use of transformed data. This was explored in the present study through analysing both the raw trait means and the $\log_{10}$ transformed means for traits which exhibited non-normal trait variance. For the stem material strength trait, MOR, despite some fluctuations in P values, very little difference was seen between the analyses carried out on the raw trait means and the $\log_{10}$ values. For stem outer cortex thickness, although the results obtained from the SNP analysis were consistent between the analyses carried out on the raw trait means and the $\log_{10}$ values, some differences were seen in the GEM analysis. In the GEM analysis carried out on raw trait means, some very plausible candidate genes were identified. The fact that these were lost following a $\log_{10}$ transformation, may suggest that the manipulation of these data in this way, resulted in the loss of important biological information. To assess this further, additional experimentation would be required.

However, at this point, it seems most appropriate to concentrate on associations seen across both data sets.

In Section 4.3.2, the results obtained from a marker validation study were described. This experiment was carried out to assess the durability of a subset of markers, detected through Associative Transcriptomics, in selecting for the material strength trait, MOR. MOR was selected for this study due to the strong correlation observed in the stem lodging experiment (see Section 4.1.2.2). The results of this marker validation experiment were very promising. Although for one of the chosen markers (D_comp1058_c0_seq1:1573) it was not possible to assess durability due to low levels of variation between WAGTAIL accessions at this locus, two additional marker loci were found to have significant selection capability for MOR. These markers were D_comp19374_c0_seq1: 702 and B_comp2391_c0_seq1:284. The finding that these are able to select for material strength, as predicted through Associative Transcriptomics, is very promising. This suggests that the methods employed in the present study have been effective in identifying robust markers which may prove valuable in crop improvement. Although based on the results obtained from the field lodging experiment conducted it would be expected that the selection of increase MOR will increase plant standing ability, it will be important to assess the effectiveness of these markers in selecting for wheat accessions with improved resistance to naturally occurring lodging in the field.

Chapter 5. General Discussion

This thesis explores the genetic control of stem mechanical strength in *B. napus* and wheat. This involved us of the recently developed genetic mapping method, Associative Transcriptomics (Harper et al., 2012). The use of transcriptome sequence information allows variation across many accessions to be identified with reduced time and cost in comparison to genomic sequencing. This is particularly advantageous when working with crop species such as *B. napus* and wheat, which have large, polyploid genomes (as transcriptome sequencing focuses on coding sequence only, and thus reduces complexity), and when the approach taken requires sequence information from many individuals. In addition, the use of transcriptome sequence allows variation at both the sequence level and the expression level to be explored. Taken together, these data can be used to not only identify QTL for important agronomic traits, but to aid the formation of testable hypotheses regarding the underlying biology of the detected marker associations. Furthermore, unlike QTL analysis using bi-parental mapping populations, for example, Associative Transcriptomics makes use of a much broader range of genetic diversity. This not only improves mapping resolution (owing to a greater number of historical recombination events since the most recent common ancestor for a diversity collection compared to a bi-parental mapping population), but also provides a more complete overview of the level of allelic variation available for exploitation through breeding.

In the present study, for both wheat and *B. napus*, a high level of variation was observed in the stem mechanical strength traits characterised using the three-point bend test. A high level of variation was also identified for a number of manually measured stem structural and morphological traits. Trait correlations revealed which stem structural and morphological components may be of greatest importance when breeding for improved stem mechanical strength in these crops. Interestingly, the relative importance of individual traits varied greatly between *B. napus* and wheat. In *B. napus,* significant correlations suggested that much of the variation in stem absolute strength could be attributed to structural properties such as stem diameter and second moment of area. In wheat however, the contributions of these traits were modest and no relationship was observed between stem thickness and absolute strength. The most important structural component in wheat was the degree of stem

hollowness, with a decrease in hollowness contributing positively to stem strength. Stem material strength was also important in wheat whereas in *B. napus* only modest contributions were made by this strength component. Other structural traits, including stem outer cortex thickness, correlated positively with absolute stem strength in both species.

In wheat and *B. napus*, experiments were able to successfully simulate mechanical perturbation in a way reflecting that induced in stem lodging. This was achieved through the use of a pulley system, designed to induce leverage on the standing plant from its upright position. This was an important experiment for two reasons a) the lack of natural lodging observed in the field meant that it was not possible to obtain an accurate score for naturally occurring lodging, and b) to assess whether or not the data obtained from the three point bend test has relevance to the strength components that are important under field conditions. For both wheat and *B. napus*, the results obtained were very promising. In wheat, it was possible to score stem breakage of the plant. This experiment revealed that accessions more likely to suffer stem breakage in the field, also have lower absolute and material strength in the results obtained from the three-point bend test. For *B. napus*, it was not possible to induce stem breakage. However, it was possible to collect data describing the amount of force required to pull the stems from their upright position under field conditions. When comparing these data to those of the three-point bend test, positive correlations were observed with the stem absolute and structural strength traits. The results obtained for both wheat and *B. napus*, suggest that important variation has been identified through the three-point bend test, which relates to variation important under field conditions.

The results obtained from Associative Transcriptomics also support the hypothesis that the genetic components for stem mechanical strength vary greatly between these species. This may not be surprising given the large differences in stem anatomy seen between wheat and *B. napus*, but very clearly shows that knowledge of the genetic control of these traits in one species will not necessarily translate to another. Maybe the most striking distinction in stem form between these crops is that while *B. napus* develops just a single stem, wheat develops many secondary stems, or tillers. In addition, monocot and dicot plants in general display very different tissue patterning in the stem, such as the distribution of vascular tissues throughout the stem. In

dicots, such as *B. napus*, the vascular bundles form radially along the stem perimeter. In contrast, monocots, such as wheat, have interspersed vasculature throughout the stem (Zhong et al., 1999). Previous studies also suggest that cell wall composition varies between monocot and dicot plants. Although there may be exceptions, dicotyledonous plants tend to have more pectin-rich cell walls in comparison to monocotyledonous species. Other cell wall components, such as xyloglucans, are thought to be dicot-specific, while monocot species, in general, have a higher level of cell wall xylan in comparison to dicotyledonous species (Vogel, 2008).

The results obtained through Associative Transcriptomics reflect these important distinctions between monocots and dicots. In *B. napus* several pectin-related candidate genes were identified, while in wheat, genes involved in xylan biosynthesis were described. Some commonalities in genetic control were also identified. A number of genes related to various plant hormones, including auxin and gibberellin were identified as potential candidates. These plant hormones are known to have many roles in plant development (Halliday, 2004). It is therefore possible that (following their validation as causal loci) targeting variation at these loci, which may be beneficial for stem mechanical strength, could have positive or negative pleiotropic effects on additional traits of agronomic importance.

It is clear that there are many differences in the genetic control of stem mechanical strength and related traits between *B. napus* and wheat. However, it is important to realise, that genetic mapping approaches are only able to identify contributing loci in the presence of genetic variation. It is possible that these species have many common genetic controls for these traits, but that they are highly conserved and therefore not detected through mapping. These components, although perhaps the most important contributors (hence the high conservation) are of little use in terms of marker-assisted crop improvement. Based on previous reports in the literature (Jones et al., 2001; Wang et al., 2006; Genet et al., 2007; Gibson, 2012; Zhao et al., 2012), it was expected that a greater number of cellulose and lignin-related genes would have been identified in close proximity to associating markers. A number of the candidate genes proposed are predicted to have roles in determining the more subtle chemistry of the cell wall, rather than the gross levels of a given polysaccharide, for instance. In *B. napus* several genes controlling the methylesterification state of cell wall pectin were identified in close proximity to associating markers. Previous reports in the

literature have shown the importance of pectin methylesterification in cell growth, an important contributor to organ outgrowth and final organ size (Wolf et al., 2009; Peaucelle et al., 2011). Given the importance of stem thickness in contributing to stem mechanical strength in *B. napus*, it is very plausible to suggest that such genes, expected to influence cell size, may be good candidates for the genetic control of stem strength. In addition, in wheat, a number of genes involved in the modification of xylan side chains were proposed as potential candidates. Arabidopsis mutants exhibiting reduced xylan/xylan acetylation, have previously been shown to have reduced stem mechanical strength (Yuan et al., 2013). Recent work suggests that the acetylation of cell wall xylan, provides a way through which xylans can form linkages with neighbouring cellulose microfibrils, and that this could be important in determining the mechanical strength of the cell wall matrix (Busse-Wicher et al., 2014). There is therefore good evidence to suggest that the xylan-related genes identified for stem mechanical strength in the wheat Associative Transcriptomics analyses, are very plausible candidate genes. In wheat, the hypothesis that fine-scale cell wall structure is important in explaining variation in stem strength is supported by the lack of correlations identified between the FT-IR data and stem material strength in the present study. This may suggest that plants are better able to cope with these finer-scale variations in cell wall composition, in comparison to greatly reduced/increased levels of lignin for example, which may carry additional biological penalties.

In addition to the identification of genes with known function in cell wall biosynthesis, the GEM analyses identified many genes which have not yet been implicated in biological processes likely to contribute to stem mechanical strength. For example, in wheat a pathway, analogous to that found in *Aspergillus nidulans* (Nahlik et al., 2010), involving the COP9 signalosome is proposed, which could potentially influence stem mechanical strength via similar pathways involving cell wall remodelling. Furthermore, several genes which have not yet been assigned biological function were identified. This illustrates the potential power of utilising GEMs in genetic mapping studies. GEMs have the power to target causal variation directly. In this way, we are less reliant on assumptions regarding the likely genetic components for a given trait, and are more open to novel discovery. By mapping transcript abundance levels (measured as RPKM) of associating GEMs as a trait

against the SNP data, the potential biological relevance of the observed expression patterns can be further explored. An example of this is again the proposal of the COP9 signalosome as a regulator of material strength in wheat. Prior to this analysis, there was little evidence to suggest that this gene may be a plausible candidate. However, further exploration of interactions between the 2D locus (at which the *COP9 SUBUNIT 5B* resides) and additional loci detected as single GEM associations (including *EIF-2B*, a gene known to interact with the COP9 signalosome), allowed a potential biological explanation for the observed associations to be proposed. Several of the GEMs found to interact with the 2D locus correspond to genes with expected roles in cell wall biosynthesis. Given that the expression of the cell wall related genes and *EIF-2B* correlate positively with that of the *COP9 SUBUNIT 5B*, and given the widely accepted role of the COP9 signalosome in targeting proteins for degradation, it is possible that the COP9 complex is involved in targeting suppressors of cell wall related genes for degradation, which ultimately contributes to stem mechanical strength in wheat.

As previously discussed, crop residues, such as the straw of *B. napus* and wheat, have recently been considered as a source of biomass for lignocellulosic ethanol production. However, the recalcitrance of the cell wall to hydrolysing enzymes, means that high-cost pretreatments are required, making lignocellulosic biomass an expensive fuel source. One way through which saccharification efficiency could be improved, is through altering cell wall composition. Altered cell wall pectin methylesterification state and xylan acetylation have both been linked with improved saccharification efficiency. However, the results presented here suggest that these cell wall components may be important in providing stem mechanical strength. It is therefore possible that an increase in lignocellulosic processibility could come at the cost of reduced agronomic performance. However, in both *B. napus* and wheat, several loci were identified which may be utilised in the marker-assisted breeding of accessions with high stem mechanical strength. Not all of these loci were found to be in close proximity to genes expected to influence cell wall saccharification efficiency. It is therefore possible that with the correct combination of markers, the breeding of accessions with improved saccharification efficiency and good standing ability could be achieved.

As previously mentioned, the detection of plausible candidate genes in close proximity, or detected directly by associating markers, does not prove causality. Although it was not possible to further explore the role of candidate genes in wheat, progress was made in *B. napus* through screening Arabidopsis T-DNA insertion lines for stem mechanical defects. Several of the mutants analysed were found to be defective in the traits to which the marker associations in *B. napus* correspond. For example, mutants defective in a gene with proposed pectin methylesterase function (AT3G12880) exhibit reduced stem structural and absolute strength. Altered stem mechanical strength was also observed in mutants defective for *GAUT5* and *SAUR72* function. This provides further evidence that these genes contribute to stem strength in *B. napus.* However, additional experimental work is required to prove this definitively. This could include transformation of the mutant with a functional copy of the Arabidopsis and brassica gene to assess whether the mechanical defects observed can be rescued. Successful complementation following transformation of the mutant with the orthologous *B. napus* gene would provide direct evidence of conserved gene function between these species. Further phenotypic exploration would also be of interest. This could include a more sensitive measure of pectin methylesterification between the mutants and WT. Given the likely role of the proposed candidate genes in controlling cell expansion, an in depth microscopy screen to assess cell size across the stem transverse would also be useful. It would then be of interest to assess whether common stem cross-sectional phenotypes can be identified in *B. napus* accessions for which marker variation was observed. Finding similar stem phenotypes between mutants defective in gene function and accessions showing variation at marker loci in close proximity to an orthologous gene, would again suggest that the causal gene to which the detected marker associations correspond has been identified.

Further validation of the biological relevance of detected marker associations, may also be acquired by more explicitly seeking causal variation in the proposed candidate genes in individuals found to exhibit marker-trait variation. In doing this, it may be possible to identify sequence variation expected to alter gene expression levels or somehow change protein conformation following translation. This would allow for predictions to be made regarding how such variation could contribute to variation in the trait of interest. Such predictions could then be tested experimentally.

Unfortunately, such in depth validation was beyond the scope of this study, but may be an area for future work.

While, it may be of biological interest to know the causal gene to which marker associations correspond, it is not necessarily important in terms of breeding (although the knowledge of the causal gene may provide information regarding potential pleiotropic effects related to the causal variant). For breeding, it is of more importance that the markers detected through genetic mapping are durable in practice. To assess this, marker validation experiments were conducted for a subset of marker loci identified in wheat and *B. napus* through Associative Transcriptomics. This was achieved through screening *B. napus* and wheat accessions of unknown genotype, for the expected marker variation. Following mechanical testing, marker-trait segregation patterns were evaluated. For wheat, both markers for which variation was identified were shown to have trait selection power. In *B. napus*, one of the three markers included in the study proved durable. This is very encouraging, especially given the complexity of these traits. The markers found to have trait selection capability, were identified by Associative Transcriptomics for material strength and absolute (and structural) strength in wheat and *B. napus* respectively. Although these traits were found to be of most relevance for the respective crop in the lodging simulation experiment previously described, it will be important to assess the effectiveness of these markers in the selection of true stem lodging resistance under field conditions. The methods employed in these marker validation studies, aimed to test the trait selection capability of individual markers for a given trait. Although the methods used were adequate for this purpose, they were not sufficient to predict phenotypic values in the accessions harvested from the ASSYST and WAGTAIL panels. Through Associative Transcriptomics, several associating loci were detected for the mechanical strength traits in wheat and *B. napus.* This included associating variation at both the sequence level and the gene expression level. To allow for accurate trait prediction, it would be necessary to screen accessions for variation at many associating loci in combination. This may also help to explain why two of the markers tested in *B. napus* were found to lack predictive capability. It may be that when considered individually, the true contribution made by these markers is too modest to allow for the identification of marker-trait segregation patterns.

The greater success seen in the wheat marker validation study may also be due to the way in which the trait data for wheat was treated in comparison to that for *B. napus.* For wheat, it was possible to obtain trait data for the majority of the 100 accessions across two years. These data were successfully combined using REML analysis, which enabled the calculation of predicted means. Through the inclusion of a large amount of trait data within a single analysis and through the incorporation of information regarding across year variation, the REML models fitted, allowed for a more accurate estimate of trait genetic effects. In contrast, the Associative Transcriptomics analysis in *B. napus*, was largely based on trait data obtained from a single field trial. Due to field space constraints, plant replication within this trial was relatively low. It is therefore likely that the lack of marker durability observed in the *B. napus* marker validation study, is at least in part due to reduced power caused by low plant and across year replication of this field trial.

The results obtained from the KWS *B. napus* trials, suggest that the stem mechanical strength traits are sensitive to variation in environmental conditions. Given this, it is also possible that those markers found to lack durability for trait selection, may be useful in selecting for the target traits, but that this success will vary depending on the environment. To assess whether or not this is the case, it would be of interest to repeat the marker validation experiment using unknown genotypes grown in further, different environments. It would also be of interest to further assess the validated wheat markers in additional environments.

It is also possible that the SNP associations detected are not indicative of markers linked to causal variation for the target traits, but are instead indicative of false positive associations. In association genetics, one of the key problems is that of spurious associations which may arise for a number of reasons. One potential cause of false positive associations is unaccounted for population structure and relatedness, where the associations detected describe shared ancestry rather than true marker-trait variation. In the present study, a Mixed Linear Model approach was taken. This method is considered superior to the General Linear Model for example, because it allows for an additional level of information regarding population structure and relatedness to be incorporated, and therefore for the rate of this source of false-positive association to be reduced. Although in the future more superior methods promising more effective correction of the confounding effects such as population

structure and relatedness will likely be developed, it seems probable that there is a limit to how much correction can be made before the biologically important associations are lost i.e. the more false positives you remove, the more false-negatives you introduce. However, through careful selection of the association panel utilised, the need for such rigorous correction can be reduced.

In addition to the potentially confounding effects of population structure and relatedness, there is also an issue regarding multiple testing in genome-wide association studies. For every marker included in the analysis, a significance, or P value is assigned, describing the relationship between segregation at that locus and trait variation. For a significance level of 5 %, for a single marker, there is a 1/20 chance that the null hypothesis is falsely rejected and that a false positive association has been identified. Of course, this probability is relative to a single test, and in association studies many thousands of tests are being carried out across all markers included in the analysis. Given this, the probability of identifying a false positive association is greatly increased (Bush and Moore, 2012). To reduce the risk of such errors (often referred to as type I error), many association studies have introduced a significance threshold where only markers over a given threshold are considered to represent genome-wide significance. For example, recent work conducted by Li et al (2014), made use of a Bonferonni correction to allow for the filtering of association signals in a genome wide association study looking at seed traits in *B. napus.* This threshold was calculated by dividing 0.05 (5% significance) by the total number of SNP markers included in their analysis (allowing each test statistic to be considered independent (Bush and Moore, 2012)). This calculation gave a significance threshold of $-\log_{10}P = 5.7$ based on the inclusion of 24,256 SNP markers. Only marker associations with a $-\log_{10}$ P value of ≥ 5.7 were deemed to be of biological significance. In the present study, the Associative Transcriptomics analysis carried out on wheat (as an example) comprised 12,456 SNP markers. Based on the method employed by Li et al (2014), the 5% significance threshold for this data set would be equivalent to a $-\log_{10}$ P value of ≥ 5.39. In the SNP analysis carried out in wheat in the present study, there was insufficient power to achieve this level of significance. One of the most highly associating markers detected for stem mechanical strength for wheat, was seen for the material strength trait, MOR. The most significant marker within this association peak was B_comp205_c0_seq1:624, which was

assigned a $-\log_{10}$ P value of 4.98. A second marker association for MOR, D_comp970_c0_seq1:1030, was assigned a much lower $-\log_{10}$P value of 3.04. These would, according to this Bonferonni significance threshold, not be considered significant. However, both of these markers proved to have trait selection capability when assessed in a completely independent panel of wheat accessions. It therefore seems unlikely that these marker associations were the results of type I error. Of course, in the marker validation study carried out in *B. napus*, two markers lacked predictive capability. It is therefore not possible at this stage to rule out the possibility that these marker associations were the result of such stochastic artefacts. Although in some studies low power may not produce marker associations of very high significance; for example in this study small sample sizes and high trait variances are likely to have been power-limiting factors, with further experimental work, markers can be validated and after which a retrospective significance threshold can be applied. In this way, it becomes possible to more accurately assess which markers may be of greatest biological relevance and may prove most valuable in crop improvement.

In cases where marker associations are of true biological significance, it would be expected that these associations would be seen at increased significance if screened across a larger panel of genotypes. The increased P values achieved through inclusion of a greater number of accessions, would reduce the risk of type I error and would allow for the incorporation of a significance threshold with reduced risk of dismissing important marker variation. For both wheat and *B. napus* the Associative Transcriptomics analyses were carried out on relatively few accessions for this type of study (the previously described Associative Transcriptomics study carried out by Li et al (2014) was based on an associating panel of 472 accessions). Although it is very promising to have identified markers for complex traits which proved durable in a second panel of accessions, such small sample size will undoubtedly limit the power of such a study.

In addition to reducing the risk of type I error, increased sample size has several other potential benefits. Firstly, in this study, loci with minor allele frequencies of <5% were omitted from the analysis. While this is important due to the risk of such variation being the result of sequencing errors for example, it may also result in the loss of biologically important information. The inclusion of a greater number of

accessions, would allow for a more accurate assignment of minor allele frequencies where a) alleles which in a smaller population are present at a frequency of <5%, may in a larger population be seen more frequently, allowing them to be incorporated into the analysis, and b) alleles which in a smaller population may be present at a frequency only marginally greater than that for the minor allele cut-off, and which in the present study could have been the cause of spurious results, may be omitted from the analysis if the frequency is not seen to increase with the inclusion of additional accessions. In this way, a more accurate representation of the biological importance of relatively rare alleles may be realised. Secondly (although this will depend on marker density), with a greater number of accessions, it is possible that mapping resolution may be further improved by the greater number of historical recombination events and therefore the further breakdown of linkage disequilibrium between neighbouring loci.

Despite some limitations, the methods employed in the present study have been effective in revealing important variation for stem mechanical strength in two very different crop species, *B. napus* and wheat. Using the novel genetic mapping method, Associative Transcriptomics, several complex traits of importance to stem mechanical strength and lodging resistance have been successfully mapped. Many potential candidate genes which have not previously been implicated in controlling stem mechanical strength in *B. napus* and wheat have been identified. Further functional analysis of these genes will be an exciting area for future research. The marker validation experiments have shown that the methods employed have power to identify markers of true value to crop improvement. The use of these validated markers in marker-assisted breeding has potential to complement current strategies for the improvement of lodging resistance in these important crops, and could as a result have great implications for maximising future yields.

REFERENCES

Agriculture USDo (2012) World agricultural supply and demand estimates.

Akoh CC, Lee G-C, Liaw Y-C, Huang T-H, Shaw J-F (2004) GDSL family of serine esterases/lipases. Progress in Lipid Research **43:** 534-552

Alalouf O, Balazs Y, Volkinshtein M, Grimpel Y, Shoham G, Shoham Y (2011) A new family of carbohydrate esterases is represented by a GDSL hydrolase/acetylxylan esterase from geobacillus stearothermophilus. Journal of Biological Chemistry **286:** 41993-42001

Angela TW, Oliver V, Ke Jun W, Nortrud GH, Keiko S, Madeleine CR, Geoffrey OW (2001) MOR1 is essential for organizing cortical microtubules in plants. Nature **411:** 610-613

Applied.Biosystems (2002) BigDye™ Terminator v3.1 Ready reaction cycle sequencing kit protocol. .

Armstrong E, Nicol H (1991) Reducing height and lodging in rapeseed with growth regulators. Australian Journal of Experimental Agriculture **31:** 245-250

Atkins IM (1938) A Simplified Method for Testing the Lodging Resistance of Varieties and Strains of Wheat. Agronomy Journal. **30:** 309-313

Atmodjo MA, Sakuragi Y, Zhu X, Burrell AJ, Mohanty SS, Atwood JA, 3rd, Orlando R, Scheller HV, Mohnen D (2011) Galacturonosyltransferase (GAUT)1 and GAUT7 are the core of a plant cell wall pectin biosynthetic homogalacturonan:galacturonosyltransferase complex. Proceedings of the national academy of sciences of the United States of America **108:** 20225-20230

Bancroft I, Morgan C, Fraser F, Higgins J, Wells R, Clissold L, Baker D, Long Y, Meng J, Wang X, Liu S, Trick M (2011) Dissecting the genome of the polyploid crop oilseed rape by transcriptome sequencing. Nature Biotechnology **29:** 762-766

Baylis AD, Wright ITJ (1990) The effects of lodging and a paclobutrazol — chlormequat chloride mixture on the yield and quality of oilseed rape. Annals of Applied Biology **116:** 287-295

Bennett T, Sieberer T, Willett B, Booker J, Luschnig C, Leyser O (2006) The arabidopsis MAX pathway controls shoot branching by regulating auxin transport. Current Biology **16:** 553-563

Berry PM, Spink J (2012) Predicting yield losses caused by lodging in wheat. Field Crops Research **137:** 19-26

Berry PM, Sylvester-Bradley R, Berry S (2007) Ideotype design for lodging-resistant wheat. Euphytica **154:** 165-179

Bienert MD, Delannoy M, Navarre C, Boutry M (2012) NtSCP1 from Tobacco is an extracellular serine carboxypeptidase III that has an impact on cell elongation. Plant Physiology **158:** 1220-1229

Blumenkrantz N, Asboe-Hansen G (1973) New method for quantitative determination of uronic acids. Analytical Biochemistry **54:** 484-489

Borojevic K, Borojevic K (2005) The transfer and history of "reduced height genes" (Rht) in wheat from Japan to Europe. Journal of Heredity **96:** 455-459

Bouton S, Leboeuf E, Mouille G, Leydecker M-T, Talbotec J, Granier F, Lahaye M, Höfte H, Truong H-N (2002) QUASIMODO1 encodes a putative membrane-bound glycosyltransferase required for normal pectin synthesis and cell adhesion in arabidopsis. The Plant Cell **14:** 2577-2590

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics **23:** 2633-2635

Braybrook SA, Peaucelle A (2013) Mechano-chemical aspects of organ formation in arabidopsis thaliana: The relationship between auxin and pectin. Public Library Of Scienece ONE **8:** e57813. doi:10.1371/journal.pone.0057813

Brenchley R, Spannagl M, Pfeifer M, Barker GL, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D, Kay S, Waite D, Trick M, Bancroft I, Gu Y, Huo N, Luo MC, Sehgal S, Gill B, Kianian S, Anderson O, Kersey P, Dvorak J, McCombie WR, Hall A, Mayer KF, Edwards KJ, Bevan MW, Hall N (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature **491:** 705-710

Breuninger H, Lenhard M (2010) Control of tissue and organ growth in plants. Current Topic in Developmental Biology **91:** 185-220

Burton RA, Ma G, Baumann U, Harvey AJ, Shirley NJ, Taylor J, Pettolino F, Bacic A, Beatty M, Simmons CR, Dhugga KS, Rafalski JA, Tingey SV, Fincher GB (2010) A customized gene expression microarray reveals that the brittle stem phenotype fs2 of barley is attributable to a retroelement in the HvCesA4 cellulose synthase gene. Plant Physiology **153:** 1716-1728

Bush WS, Moore JH (2012) Chapter 11: Genome-wide association studies. Public Library of Science Computer Biology **8:** e1002822

Busse-Wicher M, Gomes TCF, Tryfona T, Nikolovski N, Stott K, Grantham NJ, Bolam DN, Skaf MS, Dupree P (2014) The pattern of xylan acetylation suggests xylan may interact with cellulose microfibrils as a two-fold helical screw in the secondary plant cell wall of Arabidopsis thaliana. The Plant Journal **79:**492-506

Caffall KH, Pattathil S, Phillips SE, Hahn MG, Mohnen D (2009) Arabidopsis thaliana T-DNA mutants implicate GAUT genes in the biosynthesis of pectin and xylan in cell walls and seed testa. Molecular Plant **2:** 1000-1014

Cao P-J, Bartley LE, Jung K-H, Ronald PC (2008) Construction of a rice glycosyltransferase phylogenomic database and identification of rice-diverged glycosyltransferases. Molecular Plant **1:** 858-877

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. Multidisciplinary Journal of Microbial Ecology: **6:** 1621-1624

Chantret N, Salse J, Sabot F, Rahman S, Bellec A, Laubin B, Dubois I, Dossat C, Sourdille P, Joudrier P, Gautier MF, Cattolico L, Beckert M, Aubourg S, Weissenbach J, Caboche M, Bernard M, Leroy P, Chalhoub B (2005) Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (Triticum and Aegilops). Plant Cell **17:** 1033-1045

Chen H, Shan Z, Sha A, Wu B, Yang Z, Chen S, Zhou R, Zhou X (2011) Quantitative trait loci analysis of stem strength and related traits in soybean. Euphytica **179:** 485-497

Choulet F, Wicker T, Rustenholz C, Paux E, Salse J, Leroy P, Schlub S, Le Paslier M-C, Magdelenat G, Gonthier C, Couloux A, Budak H, Breen J, Pumphrey M, Liu S, Kong X, Jia J, Gut M, Brunel D, Anderson JA, Gill BS, Appels R, Keller B, Feuillet C (2010) Megabase level sequencing

reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. The Plant Cell Online **22:** 1686-1701

Collard BCY, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences **363:** 557-572

Collins SRA, Wellner KN, Martinez I, Harper AL, Miller C, Bancroft I, Waldron KW (accepted) Variation in the chemical composition of wheat straw: the role of tissue ratio and composition. Biotechnology for Biofuels

Crook MJ, Ennos AR (1994) Stem and root characteristics associated with lodging resistance in four winter wheat cultivars. The Journal of Agricultural Science **123:** 167-174

Donnelly P (2008) Progress and challenges in genome-wide association studies in humans. Nature **456:** 728-731

Downes GM, Turvey ND (1990) Lignification of wood from deformed Pinus radiata. Forest Ecology and Management **37:** 123-130

Dunn GJ, Briggs KG (1989) Variation in culm anatomy among barley cultivars differing in lodging resistance. Canadian Journal of Botany **67:** 1838-1843

Eckardt NA (2008) Role of Xyloglucan in Primary Cell Walls. The Plant Cell **20:** 1421-1422

Eckardt NA (2010) Evolution of domesticated bread wheat. The Plant Cell **22:** 993

El-Showk S, Ruonala R, Helariutta Y (2013) Crossing paths: cytokinin signalling and crosstalk. Development **140:** 1373-1383

Ellegren H (2008) Sequencing goes 454 and takes large-scale genomics into the wild. Molecular Ecology **17:** 1629-1631

Etchevers H (2007) DNA sequencing and quick clean-up.

Fabiana de G, Luisa B, Bruno Silvestre L, Amanda Pereira de S, Paula E, Diego D, Saleh A, Marina I, Marcos B, Juliana A, Gabriela G, Alisdair Robert F, Fernando C, Magdalena R (2013) Galacturonosyltransferase 4 silencing alters pectin composition and carbon partitioning in tomato. Journal of Experimental Botany **64**: 2449-2466

Feng G, Qin Z, Yan J, Zhang X, Hu Y (2011) Arabidopsis ORGAN SIZE RELATED1 regulates organ growth and final organ size in orchestration with ARGOS and ARL. New Phytologist **191:** 635-646

Flint-Garcia SA, Jampatong C, Darrah LL, McMullen MD (2003) Quantitative trait locus analysis of stalk strength in four maize populations. Crop Science. **43:** 13-22

Flintham JE, Borner A, Worland AJ, Gale MD (1997) Optimizing wheat grain yield: effects of Rht (gibberellin-insensitive) dwarfing genes. The Journal of Agricultural Science **128:** 11-25

Francocci F, Bastianelli E, Lionetti V, Ferrari S, De Lorenzo G, Bellincampi D, Cervone F (2013) Analysis of pectin mutants and natural accessions of Arabidopsis highlights the impact of de-methyl-esterified homogalacturonan on tissue saccharification. Biotechnology Biofuels **6:** 163

Garthwaite T, M.R., Hart M (1994) Pesticide usage. survey report 127 Arable farm crops in Britain

Gaut BS, Long AD (2003) The lowdown on linkage desequilibrium. The Plant Cell **15:** 1502-1506

Genet M, Stokes A, Salin F, Mickovski S, Fourcaud T, Dumail J-F, Beek R (2007) The influence of cellulose content on tensile strength in tree roots. *In* A Stokes, I Spanos, J Norris, E Cammeraat, eds, Eco-and Ground Bio-

Engineering: The use of vegetation to improve slope stability, Vol 103. Springer Netherlands, pp 3-11

Gharizadeh B, Herman ZS, Eason RG, Jejelowo O, Pourmand N (2006) Large-scale pyrosequencing of synthetic DNA: A comparison with results from Sanger dideoxy sequencing. ELECTROPHORESIS **27:** 3042-3047

Gibson LJ (2012) The hierarchical structure and mechanics of plant materials. Journal of The Royal Society Interface **12**

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology **29:** 644-652

Grundas S, Skubisz G (2008) Physical properties of cereal grain and rape stem. Research of agriculture and Engineering **54:** 80–90

Gu H, Qi C (2009) QTL analysis of lodging resistance in Brassica napus L. Jiangsu Journal of Agricultural Sciences **25:** 484-489

Guo Y, Qin G, Gu H, Qu L-J (2009) Dof5.6/HCA2, a Dof Transcription Factor gene, regulates interfascicular cambium formation and vascular tissue development in arabidopsis. The Plant Cell **21:** 3518-3534

Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol Biol **57:** 461-485

Hai L, Guo H, Xiao S, Jiang G, Zhang X, Yan C, Xin Z, Jia J (2005) Quantitative trait loci (QTL) of stem strength and related traits in a doubled-haploid population of wheat (Triticum aestivum L.). Euphytica **141:** 1-9

Hall D, Tegström C, Ingvarsson PK (2010) Using association mapping to dissect the genetic basis of complex traits in plants. Briefings in Functional Genomics **9:** 157-165

Halliday KJ (2004) Plant Hormones: The interplay of brassinosteroids and auxin. Current Biology **14:** 1008-1010

Han Y, Agarwal V, Dodd D, Kim J, Bae B, Mackie RI, Nair SK, Cann IKO (2012) Biochemical and structural insights into xylan utilization by the thermophilic bacterium caldanaerobius polysaccharolyticus. Journal of Biological Chemistry **287:** 34946-34960

Harholt J, Suttangkakul A, Vibe Scheller H (2010) Biosynthesis of pectin. Plant Physiology **153:** 384-395

Harper AL, Trick M, Higgins J, Fraser F, Clissold L, Wells R, Hattori C, Werner P, Bancroft I (2012) Associative transcriptomics of traits in the polyploid crop species Brassica napus. Nature Biotechnology **30:** 798-802

Harper AL, Trick M, Miller CN, Clissold L, Werner P, Griffiths S, Bancroft I (Under review) A transcriptomics-based association genetics platforms for crops: hexaploid wheat as an exemplar. Plant Cell

Harris DM, Corbin K, Wang T, Gutierrez R, Bertolo AL, Petti C, Smilgies D-M, Estevez JM, Bonetta D, Urbanowicz BR, Ehrhardt DW, Somerville CR, Rose JKC, Hong M, DeBolt S (2012) Cellulose microfibril crystallinity is reduced by mutating C-terminal transmembrane region residues CESA1A903V and CESA3T942I of cellulose synthase. Proceedings of the National Academy of Sciences **109**:4098-4103

HGCA (2003) PROJECT REPORT No.305. To establish seperate standing power ratings for stem and root lodging in the UK recommended lists for wheat.

Higgins J, Magusin A, Trick M, Fraser F, Bancroft I (2012) Use of mRNA-seq to discriminate contributions to the transcriptome from the constituent genomes of the polyploid crop species Brassica napus. Biomed central Genomics **13:** 247

Hongo S, Sato K, Yokoyama R, Nishitani K (2012) Demethylesterification of the primary wall by PECTIN METHYLESTERASE35 provides mechanical support to the Arabidopsis stem. Plant Cell **24:** 2624-2634

Huang X-Q, Brule-Babel A (2010) Development of genome-specific primers for homoeologous genes in allopolyploid species: the waxy and starch synthase II genes in allohexaploid wheat (Triticum aestivum L.) as examples. Biomed Central Research Notes **3:** 140

Hussey SG, Mizrachi E, Creux NM, Myburg AA (2013) Navigating the transcriptional roadmap regulating plant secondary cell wall deposition. Frontiers in Plant Science **4**

Islam N, Evans EJ (1994) Influence of lodging and nitrogen rate on the yield and yield attributes of oilseed rape (Brassica napus L.). Theoretical and Applied Genetics **88:** 530-534

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

Kaack K, Schwarz K-U (2001) Morphological and mechanical properties of Miscanthus in relation to harvesting, lodging, and growth conditions. Industrial Crops and Products **14:** 145-154

Kabakchiev B, Silverberg MS (2013) Expression quantitative trait loci analysis identifies associations between genotype and gene expression in human intestine. Gastroenterology **144:** 1488-1496

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

Kant S, Bi YM, Zhu T, Rothstein SJ (2009) SAUR39, a small auxin-up RNA gene, acts as a negative regulator of auxin synthesis and transport in rice. Plant Physiology **151:** 691-701

Kashiwagi T, Togawa E, Hirotsu N, Ishimaru K (2008) Improvement of lodging resistance with QTLs for stem diameter in rice (Oryza sativa L.). Theoretical and Applied Genetics **117:** 749-757

Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou, S., Childs KL, Davidson RM, Lin H, Quesada-Ocampo L, Vaillancourt B, Sakai H, Lee SS, Kim JA, Numa H, Itoh T, Buell CR, Matsumoto T (2013) Improvement of the oryza sativa nipponbare reference genome using next generation sequence and optical map data. Rice **6**

Kearsey MJ (1998) The principles of QTL analysis (a minimal mathematics approach). Journal of Experimental Botany **49:** 1619-1623

Keegstra K (2010) Plant Cell Walls. Plant Physiology **154:** 483-486

Kelbert AJ, Spaner D, Briggs KG, King JR (2004) The association of culm anatomy with lodging susceptibility in modern spring wheat genotypes. Euphytica **136:** 211-221

Keller M, Karutz C, Schmid JE, Stamp P, Winzeler M, Keller B, Messmer MM (1999) Quantitative trait loci for lodging resistance in a segregating wheat×spelt population. Theoretical and Applied Genetics **98:** 1171-1182

Kern KA, Ewers FW, Telewski FW, Koehler L (2005) Mechanical perturbation affects conductivity, mechanical properties and aboveground biomass of hybrid poplars. Tree Physiology **25:** 1243-1251

Kim S, Dale BE (2004) Global potential bioethanol production from wasted crops and crop residues. Biomass and Bioenergy **26:** 361-375

Kim T-H, Hofmann K, von Arnim AG, Chamovitz DA (2001) PCI complexes: pretty complex interactions in diverse signaling pathways. Trends in Plant Science **6:** 379-386

Kircher M, Kelso J (2010) High-throughput DNA sequencing – concepts and limitations. BioEssays **32:** 524-536

Kitin P, Voelker SL, Meinzer FC, Beeckman H, Strauss SH, Lachenbruch B (2010) Tyloses and phenolic deposits in xylem vessels impede water transport in low-lignin transgenic poplars: A Study by Cryo-Fluorescence Microscopy. Plant Physiology **154:** 887-898

Ko J-H, Yang SH, Park AH, Lerouxel O, Han K-H (2007) ANAC012, a member of the plant-specific NAC transcription factor family, negatively regulates xylary fiber development in Arabidopsis thaliana. The Plant Journal **50:** 1035-1048

Koerhuis ANM, Hill WG (1996) Predicted response in food conversion ratio for growth by selection on the ratio or on linear component traits, in a (sequential) selection programme. British Poultry Science **37:** 317-327

Kong E, Liu D, Guo X, Yang W, Sun J, Li X, Zhan K, Cui D, Lin J, Zhang A (2013) Anatomical and chemical characteristics associated with lodging resistance in wheat. The Crop Journal **1:** 43-49

Krysan PJ, Young JC, Sussman MR (1999) T-DNA as an Insertional Mutagen in Arabidopsis. The Plant Cell Online **11:** 2283-2290

Kubo M, Udagawa M, Nishikubo N, Horiguchi G, Yamaguchi M, Ito J, Mimura T, Fukuda H, Demura T (2005) Transcription switches for protoxylem and metaxylem vessel formation. Genes and Development **19:** 1855-1860

Lagercrantz U, Lydiate DJ (1996) Comparative genome mapping in Brassica. Genetics **144:** 1903-1910

Lee C, Abdool A, Huang C-H (2009) PCA-based population structure inference with generic clustering algorithms. Biomed Central Bioinformatics **10:** S73

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

Lehner A, Dardelle F, Soret-Morvan O, Lerouge P, Driouich A, Mollet J-C (2010) Pectins in the cell wall of Arabidopsis thaliana pollen tube and pistil. Plant Signaling & Behavior **5:** 1282-1285

Li M (1997) Anatomical and Morphological Characteristics of Maize Genotypes Varying in Resistance to Brittlesnap. University of Nebraska--Lincoln

Liu Q, Talbot M, Llewellyn DJ (2013) Pectin methylesterase and pectin remodelling differ in the fibre walls of two Gossypium species with very different fibre properties. Public Library of Science ONE **8:** e65131. DOI: 10.1371/journal.pone.0065131

Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IA, Zhao M, Ma J, Yu J, Huang S, Wang X, Wang J, Lu K, Fang Z, Bancroft I, Yang TJ, Hu Q, Wang X, Yue Z, Li H, Yang L, Wu J, Zhou Q, Wang W, King GJ, Pires JC, Lu C, Wu Z, Sampath P, Wang Z, Guo H, Pan S, Yang L, Min J, Zhang D, Jin D, Li W, Belcram H, Tu J, Guan M, Qi C, Du D, Li J, Jiang L, Batley J, Sharpe AG, Park BS, Ruperao P, Cheng F, Waminal NE, Huang Y, Dong C, Wang L, Li J, Hu Z, Zhuang M, Huang Y, Huang J, Shi J, Mei D, Liu J, Lee TH, Wang J, Jin H, Li Z, Li X, Zhang J, Xiao L, Zhou Y, Liu Z, Liu X, Qin R, Tang X, Liu W, Wang Y, Zhang Y, Lee J, Kim HH, Denoeud F, Xu X, Liang X, Hua W, Wang X, Wang J, Chalhoub B, Paterson AH (2014) The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. Nature Communications **5:** 3930. doi:10.1038/ncomms4930

Long J-R, Zhao L-J, Liu P-Y, Lu Y, Dvornyk V, Shen H, Liu Y-J, Zhang Y-Y, Xiong D-H, Xiao P, Deng H-W (2004) Patterns of linkage disequilibrium and haplotype distribution in disease candidate genes. BMC Genetics **5:** 11. doi:10.1186/1471-2156-5-11

Ma Y, Mazumdar M (2011) Multivariate meta-analysis: a robust approach based on the theory of U-statistic. Statistics in Medicine **30:** 2911-2929

Mahdavian A, Banakar A, Mohammadi A, Beigi M, Hosseinzadeh B (2012) Modelling of shearing energy of canola stem in quasi-static compressive loading using artificial neural network (ANN). Middle-East Journal of Scientific Research **11:** 374-381

Mandel JR, Nambeesan S, Bowers JE, Marek LF, Ebert D, Rieseberg LH, Knapp SJ, Burke JM (2013) Association mapping and the genomic consequences of selection in sunflower. PLoS Genet **9:** e1003378. DOI: 10.1371/journal.pgen.1003378

Mardis ER (2008) Next-Generation DNA Sequencing Methods. Annual review of genomics and human genetics **9:** 387-402

Mast SW, Moremen KW (2006) Family 47 α-mannosidases in N-glycan processing. *In* F Minoru, ed, Methods in Enzymology. Academic Press **415**: 31-46

May S, Clements D, Bennett M (2002) Finding your knockout. Molecular Biotechnology **20:** 209-221

McCann MC, Roberts K, Carpita NC (2001) Plant cell growth and elongation. *In* eLS. John Wiley & Sons, Ltd

McCurdy DW, Patrick JW, Offler CE (2008) Wall ingrowth formation in transfer cells: novel examples of localized wall deposition in plant cells. Current Opinion in Plant Biology **11:** 653-661

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

Mortimer JC, Miles GP, Brown DM, Zhang Z, Segura MP, Weimar T, Yu X, Seffen KA, Stephens E, Turner SR, Dupree P (2010) Absence of branches from xylan in Arabidopsis gux mutants reveals potential for simplification of lignocellulosic biomass. Proceedings of the National Academy of Sciences **107:** 17409-17414

Müller K, Levesque-Tremblay G, Bartels S, Weitbrecht K, Wormit A, Usadel B, Haughn G, Kermode AR (2013) Demethylesterification of cell wall

pectins in arabidopsis plays a role in seed germination. Plant Physiology **161:** 305-316

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES (2009) Association mapping: critical considerations shift from genotyping to experimental design. The Plant Cell **21**:2194-2202

Nahlik K, Dumkow M, Bayram Ö, Helmstaedt K, Busch S, Valerius O, Gerke J, Hoppert M, Schwier E, Opitz L, Westermann M, Grond S, Feussner K, Goebel C, Kaever A, Meinicke P, Feussner I, Braus GH (2010) The COP9 signalosome mediates transcriptional and metabolic response to hormones, oxidative stress protection and cell wall rearrangement during fungal development. Molecular Microbiology **78:** 964-979

Nezames CD, Deng XW (2012) The COP9 Signalosome: Its regulation of cullin-based E3 ubiquitin ligases and role in photomorphogenesis. Plant Physiology **160:** 38-46

Nieminen KM, Kauppinen L, Helariutta Y (2004) A weed for wood? Arabidopsis as a genetic model for xylem development. Plant Physiology **135:** 653-659

Noguchi T, Fujioka S, Choe S, Takatsuto S, Yoshida S, Yuan H, Feldmann KA, Tax FE (1999) Brassinosteroid-insensitive dwarf mutants of arabidopsis accumulate brassinosteroids. Plant Physiology **121:** 743-752

O'Neill CM, Bancroft I (2000) Comparative physical mapping of segments of the genome of Brassica oleracea var. alboglabra that are homoeologous to sequenced regions of chromosomes 4 and 5 of Arabidopsis thaliana. The Plant Journal **23:** 233-243

Ohta T (1982) Linkage disequilibrium due to random genetic drift in finite subdivided populations. Proceedings of the National Academy of Sciences **79:** 1940-1944

Orfila C, Sorensen SO, Harholt J, Geshi N, Crombie H, Truong HN, Reid JS, Knox JP, Scheller HV (2005) QUASIMODO1 is expressed in vascular tissue of Arabidopsis thaliana inflorescence stems, and affects homogalacturonan and xylan biosynthesis. Planta **222:** 613-622

Pallotta MA, Warner P, Fox RL, Kuchel H, Jefferies SJ, Langridge P (2003) Marker-assisted wheat breeding in the southern region of Australia. Proceedings of the Tenth International Wheat Genetics Symposium**:** 789-791

Park S, Baker J, Himmel M, Parilla P, Johnson D (2010) Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance. Biotechnology for Biofuels **3:** 1-10

Pavlista AD, Hergert GW, Baltensperger DD, Knox S (2010) Reducing height and lodging of winter wheat. **9.** doi:10.1094/CM-2010-0806-01-RS

Pawar PM-A, Koutaniemi S, Tenkanen M, Mellerowicz EJ (2013) Acetylation of woody lignocellulose: significance and regulation. Frontiers in Plant Science **4.** doi: 10.3389/fpls.2013.00118

Payne RW, Murray DA, Harding SA, Baird DB, Soutar DM (2006) GenStat for windows (9th Edition) introduction. VSN International, Hemel Hempstead.

Peaucelle A, Braybrook SA, Le Guillou L, Bron E, Kuhlemeier C, Hofte H (2011) Pectin-induced changes in cell wall mechanics underlie organ initiation in Arabidopsis. Current Biology **21:** 1720-1726

Pelloux J, Rustérucci C, Mellerowicz EJ (2007) New insights into pectin methylesterase structure and function. Trends in Plant Science **12:** 267-277

Peng J, Sun D, Nevo E (2011) Domestication evolution, genetics and genomics in wheat. Molecular Breeding **28:** 281-301

Persson S, Caffall KH, Freshour G, Hilley MT, Bauer S, Poindexter P, Hahn MG, Mohnen D, Somerville C (2007) The Arabidopsis irregular xylem8 mutant Is deficient in glucuronoxylan and homogalacturonan, which are essential for secondary cell wall integrity. The Plant Cell **19:** 237-255

Pfaffelhuber P, Lehnert A, Stephan W (2008) Linkage disequilibrium under genetic hitchhiking in finite populations. Genetics **179:** 527-537

Ping Mu Z-cL, Chun-ping Li, Hong-liang Zhang and Xiang-kun Wang (2004) QTL analysis for lodging resistance in rice using a DH population under lowland and upland cultural conditions. Australian sociaty of agronomy

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics **155:** 945-959

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics **155:** 945-959

Qiu T, Chen Y, Li M, Kong Y, Zhu Y, Han N, Bian H, Zhu M, Wang J (2013) The tissue-specific and developmentally regulated expression patterns of the SAUR41subfamily of SMALL AUXIN UP RNA genes: potential implications. Plant Signaling & Behavior **8:** e25283. doi: 10.4161/psb.25283

Reboul R, Geserick C, Pabst M, Frey B, Wittmann D, Lutz-Meindl U, Leonard R, Tenhaken R (2011) Down-regulation of UDP-glucuronic acid biosynthesis leads to swollen plant cell walls and severe developmental defects associated with changes in pectic polysaccharides. Journal of Biological Chemistry **286:** 39982-39992

Reif J, Liu W, Gowda M, Maurer H, Möhring J, Fischer S, Schechert A, Würschum T (2010) Genetic basis of agronomically important traits in sugar beet investigated with joint linkage association mapping. TAG Theoretical and Applied Genetics **121:** 1489-1499

Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M, Miller N, Mueller LA, Mundodi S, Reiser L, Tacklind J, Weems DC, Wu Y, Xu I, Yoo D, Yoon J, Zhang P (2003) The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. Nucleic Acids Research **31:** 224

Roach MJ, Mokshina NY, Badhan A, Snegireva AV, Hobson N, Deyholos MK, Gorshkova TA (2011) Development of cellulosic secondary walls in flax fibers requires β-galactosidase. Plant Physiology **156:** 1351-1363

Rodríguez-Gacio MdC, Iglesias-Fernández R, Carbonero P, Matilla ÁJ (2012) Softening-up mannan-rich cell walls. Journal of Experimental Botany **63:** 3976-3988

Sanchez P, Nehlin L, Greb T (2012) From thin to thick: major transitions during stem development. Trends in Plant Science **17:** 113-121

Shen X, Ronnegard L (2013) Issues with data transformation in genome-wide association studies for phenotypic variability. F1000Res **2:** 200. (doi: 10.12688/f1000research.2-200.v1)

Sikora P, Chawade A, Larsson M, Olsson J, Olsson O (2011) Mutagenesis as a tool in plant genetics, functional genomics, and breeding. International Journal of Plant Genomics **2011**

Skøt L, Humphreys MO, Armstead I, Heywood S, Skøt KP, Sanderson R, Thomas ID, Chorlton KH, Hamilton NRS (2005) An association mapping

approach to identify flowering time genes in natural populations of <i>Lolium perenne</i> (L.). Molecular Breeding **15:** 233-245

Soto-Cerda BJ, Cloutier S (2012) Association mapping in plant genomes, genetic diversity in plants, Prof. Mahmut Caliskan (Ed.), ISBN: 978-953-51-0185-7, InTech, DOI: 10.5772/33005. Available from: http://www.intechopen.com/books/genetic-diversity-in-plants/association-mapping-in-plant-genomes.

Spatz H, Köhler L, Speck T (1998) Biomechanics and functional anatomy of hollow-stemmed sphenopsids. I. Equisetum giganteum (Equisetaceae). American Journal of Botany **85:** 305-314

Sterling JD, Atmodjo MA, Inwood SE, Kumar Kolli VS, Quigley HF, Hahn MG, Mohnen D (2006) Functional identification of an Arabidopsis pectin biosynthetic homogalacturonan galacturonosyltransferase. Proceedings of the National Academy of Sciences of the United States of America **103:** 5236-5241

Stokes D, Fraser F, Morgan C, O'Neill C, Dreos R, Magusin A, Szalma S, Bancroft I (2010) An association transcriptomics approach to the prediction of hybrid performance. Molecular Breeding **26:** 91-106

Sun Y, Fan X-Y, Cao D-M, Tang W, He K, Zhu J-Y, He J-X, Bai M-Y, Zhu S, Oh E, Patil S, Kim T-W, Ji H, Wong WH, Rhee SY, Wang Z-Y (2010) Integration of brassinosteroid signal transduction with the transcription network for plant growth regulation in arabidopsis. Developmental Cell **19:** 765-777

Szymanska-Chargot M, Zdunek A (2013) Use of FT-IR spectra and PCA to the bulk characterization of cell wall residues of fruits and vegetables along a fraction process. Food Biophysics **8:** 29-42

Tien M, Tu C-PD (1987) Cloning and sequencing of a cDNA for a ligninase from Phanerochaete chrysosporium. Nature **326:** 520-523

Trick M, Long Y, Meng J, Bancroft I (2009) Single nucleotide polymorphism (SNP) discovery in the polyploid Brassica napus using Solexa transcriptome sequencing. Plant Biotechnology Journal **7:** 334-346

Uddin MN, Marshall DR (1989) Effects of dwarfing genes on yield and yield components under irrigated and rainfed conditions in wheat (Triticum aestivum L.). Euphytica **42:** 127-134

Valentini SR, Casolari JM, Oliveira CC, Silver PA, McBride AE (2002) Genetic interactions of yeast eukaryotic translation initiation factor 5A (eIF5A) reveal connections to poly(A)-binding protein and protein kinase C. Signaling. Genetics **160:** 393-405

Van Acker R, Leplé J-C, Aerts D, Storme V, Goeminne G, Ivens B, Légée F, Lapierre C, Piens K, Van Montagu MCE, Santoro N, Foster CE, Ralph J, Soetaert W, Pilate G, Boerjan W (2014) Improved saccharification and ethanol yield from field-grown transgenic poplar deficient in cinnamoyl-CoA reductase. Proceedings of the National Academy of Sciences **111:** 845-850

van Delden SH, Vos J, Ennos AR, Stomph TJ (2010) Analysing lodging of the panicle bearing cereal teff (Eragrostis tef). New Phytologist **186:** 696-707

Vanneste S, Friml J Auxin: A trigger for change in plant development. Cell **136:** 1005-1016

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

Wang J, Zhu J, Huang R, Yang Y (2012) Investigation of cell wall composition related to stem lodging resistance in wheat (Triticum aestivum L.) by FTIR spectroscopy. Plant Signalling and Behaviour **7:** 856-863

Wang J, Zhu J, Lin Q, Li X, Teng N, Li Z, Li B, Zhang A, Lin J (2006) Effects of stem structure and cell wall components on bending strength in wheat. Chinese Science Bulletin **51:** 815-823

Wang M, Yuan D, Gao W, Li Y, Tan J, Zhang X (2013) A comparative genome analysis of PME and PMEI families reveals the evolution of pectin metabolism in plant cell walls. Public Library Of Science ONE **8:** e72082. DOI: 10.1371/journal.pone.0072082

Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F, Huang S, Li X, Hua W, Wang J, Wang X, Freeling M, Pires JC, Paterson AH, Chalhoub B, Wang B, Hayward A, Sharpe AG, Park BS, Weisshaar B, Liu B, Li B, Liu B, Tong C, Song C, Duran C, Peng C, Geng C, Koh C, Lin C, Edwards D, Mu D, Shen D, Soumpourou E, Li F, Fraser F, Conant G, Lassalle G, King GJ, Bonnema G, Tang H, Wang H, Belcram H, Zhou H, Hirakawa H, Abe H, Guo H, Wang H, Jin H, Parkin IA, Batley J, Kim JS, Just J, Li J, Xu J, Deng J, Kim JA, Li J, Yu J, Meng J, Wang J, Min J, Poulain J, Wang J, Hatakeyama K, Wu K, Wang L, Fang L, Trick M, Links MG, Zhao M, Jin M, Ramchiary N, Drou N, Berkman PJ, Cai Q, Huang Q, Li R, Tabata S, Cheng S, Zhang S, Zhang S, Huang S, Sato S, Sun S, Kwon SJ, Choi SR, Lee TH, Fan W, Zhao X, Tan X, Xu X, Wang Y, Qiu Y, Yin Y, Li Y, Du Y, Liao Y, Lim Y, Narusaka Y, Wang Y, Wang Z, Li Z, Wang Z, Xiong Z, Zhang Z (2011) The genome of the mesopolyploid crop species Brassica rapa. Nat Genet **43:** 1035-1039

Wang YH (2008) How effective is T-DNA insertional mutagenesis in Arabidopsis? Journal of Biochemical Technology **1:** 11-20

Weibel RO, Pendleton JW (1964) Effect of artificial lodging on winter wheat grain yield and quality. Agronomy Journal. **56:** 487-488

Wolf S, Mouille G, Pelloux J (2009) Homogalacturonan methyl-esterification and plant development. Molecular Plant **2:** 851-860

Xiao C, Somerville C, Anderson CT (2014) POLYGALACTURONASE INVOLVED IN EXPANSION1 Functions in cell elongation and flower development in arabidopsis. The Plant Cell **26:** 1018-1035

Xu YL, Gage DA, Zeevaart J (1997) Gibberellins and stem growth in arabidopsis thaliana (effects of photoperiod on expression of the GA4 and GA5 loci). Plant Physiology **114:** 1471-1476

Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. Current Opinion in Biotechnology **17:** 155-160

Yuan Y, Teng Q, Zhong R, Ye Z-H (2013) The arabidopsis DUF231 domain-containing protein ESK1 mediates 2-O- and 3-O-acetylation of xylosyl residues in xylan. Plant and Cell Physiology **54:** 1186-1199

Zhang F, Wang Y, Deng H-W (2008) Comparison of population-based association study methods correcting for population stratification. Public Library Of Science ONE **3:** e3392

Zhao D, Han C, Tao J, Wang J, Hao Z, Geng Q, Du B (2012) Effects of inflorescence stem structure and cell wall components on the mechanical

strength of inflorescence stem in herbaceous peony. International Journal of Molecular Science **13:** 4993-5009

Zhong R, Lee C, Zhou J, McCarthy RL, Ye Z-H (2008) A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in arabidopsis. The Plant Cell **20:** 2763-2782

Zhong R, Ripperger A, Ye Z-H (2000) Ectopic deposition of lignin in the pith of stems of two arabidopsis mutants. Plant Physiology **123:** 59-70

Zhong R, Taylor JJ, Ye Z-H (1999) Transformation of the collateral vascular bundles into amphivasal vascular bundles in an arabidopsis mutant. Plant Physiology **120:** 53-64

Zhong R, Ye Z-H (1999) IFL1, a Gene regulating interfascicular fiber differentiation in arabidopsis, encodes a homeodomain-leucine zipper protein. The Plant Cell Online **11:** 2139-2152

Zhong R, Ye Z-H (2001) Alteration of auxin polar transport in the arabidopsis *ifl1* mutants. Plant Physiology **126:** 549-563

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. Plant Genenetics. **1:** 5-20