Characterisation of a Major QTL For Dietary Fibre in Hexaploid Wheat Using the Watkins collection of Landraces

Submitted by

James Alex Brett

University of East Anglia for the degree of Doctor in Philosophy in Crop Genetics

Rothamsted Research and John Innes Centre.

Submitted: September-2024

The work presented within this thesis is my entirely my own work, unless otherwise stated in the body of the text.

Abstract.

Wheat is the third most widely produced crop globally, and the foods made from it are major staples. 70% of bread sales in the UK are for white bread, primarily due to the low cost and palatability compared to whole grain bread. Arabinoxylan (AX) is the primary dietary fibre in wheat and has a complex biosynthesis. Many QTLs have been identified in various crosses, but few have produced molecular markers that predict high AX for breeding programs. The Watkins landrace collection represents a valuable source of novel genetic diversity absent from elite wheat cultivars. Here I describe in this project the phenotyping of total (TOT) and water extractable (WE) arabinoxylan in three Paragon x Watkins recombinant inbred (RIL) populations, grown over three consecutive years. A QTL analysis was performed, identifying a robust QTL on chromosome 1BL, associated with wholegrain WE-AX. Presence and Absence variation of gene content was determined in the 1B QTL using the Wheat10+ pangenome, and the QTL boundary was reduced, using a bulked segregant analysis. Candidate genes were identified, utilising paired end RNA-seq data from Yumai34 and Valoris, mapped to the 1B chromosome of Paragon, where Yumai34 shares 1B high allele with Paragon. A differential gene expression analysis was performed using the DESEQ2 pipeline, and SNP variants were called to determine expression and SNP differences between Yumai34 and Valoris, in comparison with Paragon. Finally, variants were called for each of the 30 Watkins accessions against the Paragon 1B chromosome using raw sequencing reads derived from the Watseq project. KASP markers have been developed for these candidate genes and may provide breeders with more accurate markers for use in breeding programmes, to achieve higher AX commercial wheat varieties. I also describe various putative mechanisms for each candidate, which may affect AX concentration in wheat grain.

Contents

List Of Figures

List of Tables

Acknowledgements

I would like to thank my supervisors, Dr. Alison Lovegrove, Professor. Peter Shewry and Dr. Simon Griffiths for all the support you have given me over the PhD project. I would also like to thank Heather Angus, our technician at Rothamsted Research, for providing essential assistance in sampling flour, 4000 wheat plant leaves and various biochemical assays (often to our dismay!) and to Dr. Anneke Prins and Dr. Ondrej Kosik, for their additional guidance and support. Next, I would thank Dr. Dan Smith, for the tutorship in bioinformatics, for which I am ever grateful. Dan has been instrumental in coaching me in how to produce a majority of the genome and RNA analysis datasets, along with the pangenome pipeline, that we developed. Furthermore, I would like to thank Alison's daughters for their help in harvesting the three years of field trial material grown for this project. The support of these people has been invaluable and helped me develop skills I otherwise may not have been able to learn.

Finally, I would like to acknowledge my friends and family who provided helpful reminders to keep going and persevere, even when things were going wrong, and to give the emotional support needed through this journey. Without all of these people to promote a healthy work-life balance, the PhD may have been a much more difficult endeavour. Thank you to everyone for creating a friendly and enjoyable PhD experience.

Main body word count: 33102 words

Chapter 1
General Introduction and Literature Review.

1.1 Introduction.

Wheat is of immense global economic importance. From 2014-2018, the global production of wheat averaged ~745 million tonnes, according to FAOSTAT (http://www.fao.org/faostat/en/#home), increasing to ~650 million tonnes between 2008 and 2012 (Shewry and Hey, 2015). Wheat remains the third most widely produced crop in the world, after maize and rice at ~760 mt. About 60% of this is consumed directly by humans, with most of the remainder being used for livestock feed and a small amount for industrial uses, such as starch production for distilling (Emily et al., 2013).

Wheat is a vital food source for developing and developed countries alike. It can be processed to produce an array of breads, cakes, pasta and noodles and be used as a food ingredient, contributing to its global popularity. This versatility has contributed to wheat becoming one of the dominant crops worldwide. Moreover, wheat and bread are deeply embedded in culture and religions, especially in Abrahamic religions (Judaism, Islam and Christianity).

1.2 Nutritional quality of wheat

Cereals are rich in many essential nutrients (Table 1), and as an affordable staple food (consumed every day, often more than once a day) are important sources of these nutrients in a balanced diet. According to the 2014 National Diet and Nutrition Survey (NDNS), cereals make up 31% of the total energy intake, 40% of non-starch polysaccharide (NSP) intake and significant proportions of B vitamin and mineral intakes in the UK diet (Bates et al., 2014a). Bread is a staple in most British households, with around 11 million loaves sold every day. Between 60-70% of

bread purchased in the UK is white (https://www.ukflourmillers.org/flourbreadconsumption) which is largely favoured over wholemeal due to affordability and taste. Additionally, white bread has a longer shelf-life than wholemeal bread. Wholemeal flour consists of the starchy endosperm (SE), germ and bran layers, whereas refined white flour consists only of the SE. Although wholegrain products are undoubtedly 'healthier' (McRae, 2017), due to their increased contents of dietary fibre (DF) and other bioactive components, refined grain products, such as those produced from white wheat flour, should not be ignored in terms of their contribution to nutrition and health. Since 2005, it has been recommended in the US that half of the grains consumed should be whole grain (U.S. Department of Health and Human Services and U.S. Department of Agriculture, 2020). Although white flour and its products are lower in fibre and other beneficial components, whole grain/wholemeal products are less palatable to many, and the nutrients contained may also be less bioavailable.

For example, the bran fraction, that is part of wholemeal flour, contains antinutrients such as phytic acid, lectins, tannins, and enzyme inhibitors. About 90% of phytic acid is located in the aleurone, with the remaining 10% in the germ (Dost and Tokul, 2006). Phytic acid has a strong affinity for minerals, including iron and zinc, forming phytate-mineral complexes (chelates) (Gupta, Gangoliya and Singh, 2013) which reduce mineral solubility and hence bioavailability in the human gastro-intestinal tract. The iron and zinc contents in white bread are about 6.7mg/kg and 8.4mg/kg, respectively, while in wholemeal bread they are much higher at 28.2mg/kg and 28.6mg/kg, respectively. Therefore, although white flour has much lower amounts of these minerals, they are more bioavailable because phytates are not present in white flour (Shewry, Pellny and Lovegrove, 2016).

Table 1: Total percentage contribution of cereals and breads to essential nutrient intake in men and woman between ages 19-64 in the UK, taken from Shewry and Hey (2015), which refers to the 2014 NDNS (Bates et al.*, 2014b)*

Nutrient	Total (Men)	Total (Women)	Bread (Men)	Bread (Women)
Energy	31	31	12	10
Protein	23	23	11	11
Carbohydrates	45	45	21	18
NSP	40	37	21	18
Thiamin (B1)	36	35	16	15
Riboflavin (B2)	22	22	5	4
Niacin (eq) (B3)	25	26	11	10
Vitamin B6	16	18	5	4
Folates (B9)	27	27	12	12
Iron	40	38	16	15
Calcium	32	30	19	15
Magnesium	28	28	13	13
Sodium	31	31	19	18
Zinc	25	25	12	12
Copper	34	32	15	14
Selenium	28	27	13	12

1.3 Importance of high-quality bread in our diet.

The recommended daily intake of dietary fibre is 30g/day (SACN), yet in the UK the average fibre intake for adults is ~18g/day (Table 2).

Table 2: Average daily intake of AOAC fibre in people in different age brackets and the proportion meeting daily recommended targets- taken from NDNS RP UK years 7 and 8 (combined) (2014/2015-2015-2016)

	Age (years)								
	1.5-3 Sex-combined	4 -10 m	f	11-18 m	f	19-64 m	f	65+ m	f
AOAC Fibre intake (g/day)	10.3	14.5	13.5	16.4	14.1	20.7	17.4	19.0	16.4
% meeting AOAC recommended fibre intake	10	11	9	6	2	13	4	9	4

Since white flour products are generally preferred, it would be beneficial to produce wheat cultivars with higher DF contents within the starchy endosperm (SE) to deliver increases in DF intake with minimal cost or the need to change diets.

1.4 The Wheat Grain

Grain originates from the fertilization of a monocarpellary ovary within the ear of cereals and is single-seeded fruit known botanically as a caryopsis. The mature wheat grain comprises several tissues, as shown in Figure 1.

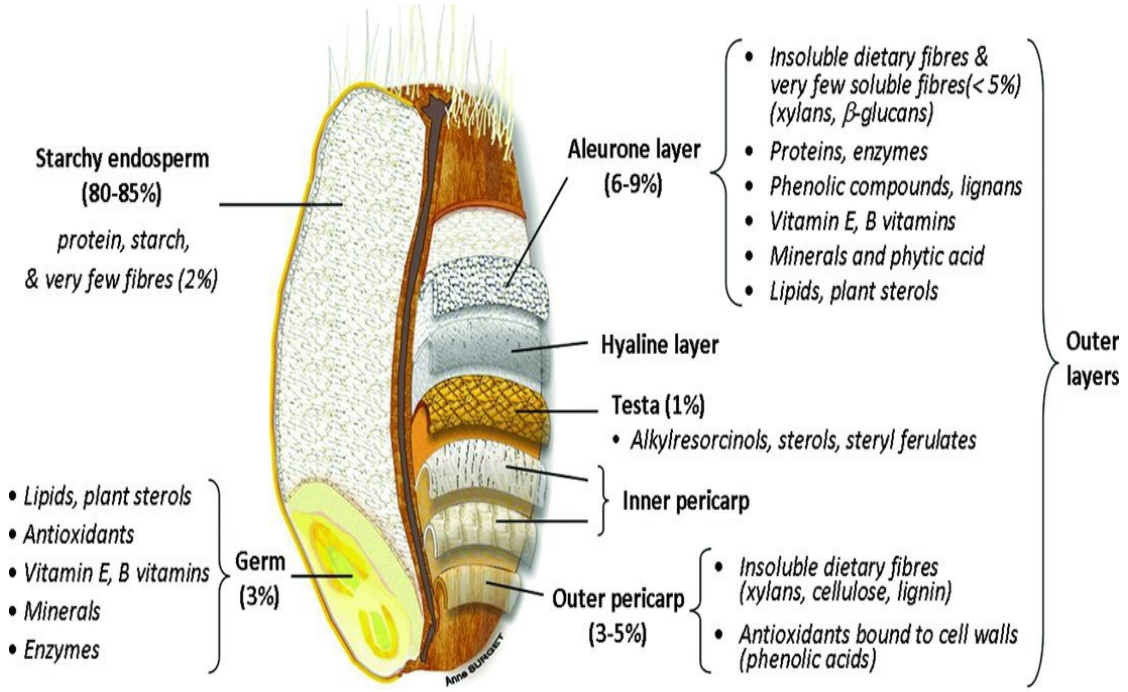


Figure 1: Representative diagram of the structure of wheat caryopsis, including each component of the outer bran layer. The distribution of bioactive compounds, vitamins, minerals, polysaccharides, and protein is labelled alongside the tissue they develop. The image was taken from Onipe et al., (2015)(Onipe, Jideani and Beswa, 2015), which itself was adapted from the original from Surget & Barron (2005)(SURGET and BARRON, 2005) and Brouns et al. (2012)(Brouns et al., 2012).

1.4.1 Embryo

The embryo is located at the dorsal side of the proximal end of the grain and comprises the embryonic axis and a single storage cotyledon (the scutellum) (Fig 1). Although it represents only 3 -4% of the total grain weight it is a rich resource of protein, lipids, fibre, minerals and vitamins (notably B vitamins and Vitamin E).

1.4.2 Endosperm.

The endosperm is the major storage tissue of the grain and the main target for improvement. It accounts for ~85% of grain and comprises two distinct cell types: the starchy endosperm (SE) and a layer of outer cells called the aleurone.

The SE differs markedly from the aleurone and undergoes apoptosis during maturation, although the cells remain intact, The SE gives the white flour fraction on milling and its

composition determines the most important quality criteria for the classification and end use quality of the grain: hardness, protein content and quality and starch content and properties. Soft, low protein wheat is used in noodles, cakes and biscuits whereas hard protein-rich wheat is used for breadmaking. Ultra hard Durum wheat is used for the production of pasta.

The SE is not homogenous and presents gradients in the concentrations of starch, protein, non-starch polysaccharides (NSPs) and other components (Tosi et al., 2018).

1.4.3 Bran

In commercial milling, the bran layer is defined as comprising the outer grain layers (pericarp, testa and hyaline layer), the embryo (germ) and the aleurone, accounting for 19% of the total grain weight (Hossain *et al.*, 2013). The bran, excluding the germ, is referred to as 'outer bran layers' (OBLs) and comprises five different layers, which differ in their contents of bioactive components and nutrients (Fig 1).

The outer bran layers (OBLs) have the most diverse nutritional profile with the outer pericarp in particular being concentrated in cell wall polysaccharides and lignins and containing cell wall-bound phenolic acids with antioxidant activity (Stuper-Szablewska *et al.*, 2019) . The major phenolic acid is ferulic acid, with a lower content of p-coumaric acid, which is concentrated in the aleurone and primarily bound to arabinoxylan.

1.5 Milling.

Wheat grain is usually roller-milled which separates the bran and germ from the starchy endosperm which is recovered as white flour.

Before milling the grain must be tempered to ~15% moisture content (De Brier *et al.*, 2015), which prevents the shattering of the bran to give small fragments which may contaminate the flour fraction. The grain then undergoes passage through up to 4 breaks rollers in a modern

roller mill, and sometimes 5 in a commercial mill, each of which shears more of the starchy endosperm away from the bran. Large SE fragments are then reduced in size using further sets of rollers to give particles 10-300µm which are separated by sieving. Although most products are made from white flour, bran fractions can be combined with the flour with wholemeal being produced by combining all of the fractions.

1.6 Dietary Fibre.

The currently accepted definition of DF is from CODEX Alimentarius Commission (CAC) (2009) and states, "*Dietary fibre means carbohydrate polymers* with *10 or more monomeric units*, *which are not hydrolyzed by the endogenous enzymes in the small intestine of humans and belong to the following categories:*

1. *Edible carbohydrate polymers naturally occurring in the food as consumed.*

2. *Carbohydrate polymers, which have been obtained from food raw material by physical, enzymatic, or chemical means and which have been shown to have a physiological effect of benefit to health as demonstrated by generally accepted scientific evidence to competent authorities,*

3. *Synthetic carbohydrate polymers, which have been shown to have a physiological effect of benefit to health as demonstrated by generally accepted scientific evidence to competent authorities".*

This definition was revised in 2014 to also include carbohydrates with degrees of polymerization (DP) 3-6 (Jones, 2014). A precise definition of DF is crucial for accurate determination and labelling of the DF content of foods.

1.7 Accepted health claims for fibre.

Eleven health claims have been approved by EFSA (European Food Safety Authority) for DF from wheat, barley oats and rye, relating to normal bowel function (including (reducing) transit time and (increasing) bulking) and regulating blood glucose or cholesterol (Table 3). To be considered by the EFSA for an accepted health claim, there must be substantial scientific evidence, so as the scientific consensus regards the health claim as true, and there must be clear human clinical studies showing the effect of a product on the aspect of human health being studied. Furthermore, the UK Scientific Advisory Committee on Nutrition (SACN) carried out an extensive review of the evidence for health benefits from fibre components which reported additional effects (Table 4) which failed to meet the strict criteria required by EFSA or the FDA (US Food and Drug Administration) (which include characterization of the proposed substances, beneficial effect on health and clear cause and effect relationships between the two).

β-glucan has been approved by EFSA for claims that 3g of β-glucan from oat/barley per day lowers serum cholesterol by ~0.27mmol/L, and that 4g/30g per day can reduce postprandial glycaemia, which is considered to reduce the risk of developing type II diabetes or be helpful in managing the condition (Harland, 2014). More recently, arabinoxylans from wheat endosperm have also been approved by EFSA for reducing the spike in serum glucose levels following the consumption of a meal (Kellow and Walker, 2018).

Table 3: EFSA health claims related to cereal fibre intake. From http://ec.europa.eu/nuhclaims/

Nutrient	Total (Men)	Total (Women)	Bread (Men)	Bread (Women)
Energy	31	31	12	10
Protein	23	23	11	11
Carbohydrates	45	45	21	18
NSP	40	37	21	18
Thiamin (B1)	36	35	16	15
Riboflavin (B2)	22	22	5	4
Niacin (eq) (B3)	25	26	11	10
Vitamin B6	16	18	5	4
Folates (B9)	27	27	12	12
Iron	40	38	16	15
Calcium	32	30	19	15
Magnesium	28	28	13	13
Sodium	31	31	19	18
Zinc	25	25	12	12
Copper	34	32	15	14
Selenium	28	27	13	12

Table 4: Summary of the findings from the SACN 2015 report on fibre and cardio-metabolic and colo-rectal health. Taken from Shewry and Hey (2015). 'Total dietary fibre (TDF)', 'soluble' and 'insoluble' include results from other sources besides cereals (e.g., fruit and vegetables). The left shows the health outcome associated with the DF fraction (right) in fruits and cereals. Abbreviations: Cardiovascular disease (CVD), Total dietary fibre (TDF), Low-density lipoproteins (LDL)

Health outcome	Fraction
Cardio-metabolic health	
CVD	TDF, insoluble fiber, soluble fiber, total cereals, wholegrains
Coronary events	TDF, insoluble fiber, cereal fiber, high fiber breakfast cereals
Stroke	TDF, wholegrains
Hypertension	Wholegrains
Blood pressure	*Oat bran/oat or barley β-glucans*
Fasting blood total cholesterol, LDL **cholesterol, triacylglycerol**	*Oat bran/oat or barley β-glucans*
Type 2 diabetes	TDF, insoluble fiber, soluble fiber, cereal fiber, high fiber breakfast cereals, whole grain bread, wholegrains
Colo-rectal health	
Fecal weight	*TDF, wheat fiber, non-wheat cereal fiber*
Intestinal transit time	*TDF, wheat fiber, non-wheat cereal fiber*
Intestinal transit time in patients **with constipation**	*Wheat fiber*
Constipation	*Cereal fiber*
Colo-rectal cancer	TDF, cereal fiber
Colon cancer	TDF, cereal fiber, wholegrains
Rectal cancer	TDF, wholegrains

As of 2012, 'wheat bran fibre' was approved in the EU specifically for reducing intestinal transit time and increasing faecal bulking (Harland, 2015). The mechanism of DF action in modulating health is still not completely understood. However possible mechanisms include a role of soluble fibre in increasing the viscosity within the small intestine, which could slow the absorption of sugars, and increase fermentation in the colon to produce short chain fatty acids which have a range of benefits including reducing the risk of colo-rectal cancer.

1.8 Solubility and The Impact on Health.

DFs are often classified as soluble and insoluble. Cellulose is totally insoluble, whereas others DF components may be partially soluble, for example, mixed-linked-β-glucan (MLG) from barley and oats, or wholly soluble such as AGP and fructans. It is thought that soluble and insoluble DF have different health benefits.

Insoluble fibre may be poorly fermentable by the human colon microbiota (Kelleher *et al.*, 1984; Kaczmarczyk, Miller and Freund, 2012), but is important for increasing faecal bulking and reducing intestinal transit time (Table 4), and potentially improving cardio-metabolic health and type II diabetes management (Table 3), while soluble DF is rapidly fermentable in the colon and has other benefits, for example, on postprandial glucose response (Abutair, Naser and Hamed, 2016; De Carvalho *et al.*, 2017). Soluble types of fibre, such as β-glucan and water-extractable arabinoxylans (WE-AX) form viscous solution which increase the viscosity within the small intestine and reduce the rate of starch digestion and hence the decrease in postprandial blood glucose levels (Chutkan *et al.*, 2012). This effect is supported by studies in animals (Mio *et al.*, 2020) and clinical studies in humans (Ames et al., 2015; Thondre, Shafat and Clegg, 2013). Recently, barley β-glucans and arabinoxlyans have also been shown to contribute to the “second

meal effect" where one meal may have an effect on the glycaemic response after consuming a second meal. Mio *et al.,,* (2023) study showed that the fermentation and subsequent release of short-chain fatty acids from the microbiota in the colon increased secretion of glucagon-like peptide 1 (GLP-1)(Mio *et al.*, 2023). Many GLP-1 receptor agonists (GLP-1 RAs), which stimulate the secretion of GLP-1, are available and have been studied for the management of diabetes. GLP-1 RAs have also been associated with clinical cardiovascular benefits (Nauck *et al.*, 2017) and the stimulation of GLP-1 secretion via the fermentation of dietary fibre may contribute to the beneficial effects on cardiovascular health shown in Table 3.

1.9 Plant cell walls and dietary fibre.

Plant cell walls are complex structures comprising mainly polysaccharides with smaller amounts of proteins and, in some cases, lignin. They vary widely in their structures between species, tissues and developmental stages and may vary with cellular lineages (Keegstra, 2010) to meet the structural and mechanical requirements of the tissues.

Cell walls are categorised into two classes: primary and secondary cell walls.

The synthesis of primary cell walls is associated with cell division and expansion, and they are thin, hydrated structures which are both extensible and strong enough to withstand the tensile forces produced from turgor pressure within the cell. Extensibility of the primary cell wall allows for the enlargement of cells during the growth stages(Hamant and Traas, 2010).

Primary Walls are comprised mainly of cellulose, which consists of unbranched chains of glucose units which form insoluble microfibrils, but also contain other polysaccharides which are broadly classified into two types. Hemicelluloses comprise a range of polysaccharides consisting of hexose and/or pentose sugars while pectins are complex polymers which are rich

in glucuronic acid with three major types: homogalacturonan, rhamnogalacturonan-I and rhamnogalacturonan-II.

Plant cell walls (PCWs) are highly variable in structure and exhibit different NSP profiles across evolutionary clades and families. Most higher plants have Type 1 primary cell walls which comprise cellulose micro-fibrils and chains of xyloglucan embedded in a matrix of pectins. An exception to this is some monocotyledonous plants, which include cultivated cereals. These have Type 2 primary walls in which the major non-cellulose polysaccharides are not pectins and xyloglucan but other hemicelluloses, notably arabinoxylan (AX) and mixed linked β-glucan.

In both type 1 and type 2 walls, pectin polysaccharides are among the earliest components deposited into the cell walls. However, in grasses, pectin polysaccharides are depleted during cell wall maturation.

Secondary cell walls are synthesised once cellular growth has stopped. They are more rigid and require high compressive and tensile strength to withstand mechanical forces that may impact plant tissues. As growth has ceased, these cell walls do not need to be extensible (Speck and Burgert, 2011). Unlike primary walls, they also contain lignin, a complex cross-linked phenolic polymer.

The cell wall composition of wheat varies widely between tissues. The endosperm cells (comprising the starchy endosperm and aleurone layer) have only primary walls which comprise mainly arabinoxylan with smaller amounts of β-glucan, cellulose and mannan. In addition, immunomicroscopy shows the presence of small amounts of xyloglucan, callose and pectins (Tosi *et al.*, 2018b).

The outer layers of the grain (pericarp and testa) are maternal tissues and have similar cell wall compositions to the vegetative tissues, notable high proportions of highly complex glucuronoarabinoxylans (GAX), higher proportions of content of cellulose and the presence of lignin (12% of the dry weight of pericarp tissue). Oxidative cross-linking also enables cross-linking

of the GAX ferulate with lignin forming insoluble complexes which are prevalent in the outer pericarp (Chateigner-Boutin *et al.*, 2018) and results in more rigid cell walls.

The maternal outer layers form the bran fraction together with the aleurone layer which is an endosperm tissue (see above and germ). β-glucan accounts for about 29% of the cell walls of the aleurone, compared to about 20% in the SE, and hence accounts for about 6% of the cell wall composition in bran.

The pericarp and seed coat have thick cell walls, which account for about 46% of their dry weight in the mature grain. Wheat bran is therefore a very rich source of dietary fibre (Maes and Delcour, 2002).

As discussed above, the grain tissues also contain small amounts of other polysaccharides (hemicelluloses and pectins) and two types of polysaccharides which also contribute to the DF fraction: fructans and arabinogalactan peptide (AGP)- though these are not part of the cell wall. AGP comprises a short peptide with three arabinogalactan side chains and accounts for 0.4% dry weight of wheat endosperm (Loosveld, Grobet and Delcour, 1997) and is highly soluble in water. Fructans are polymers of fructose molecules linked together by glycosidic bonds and account for up to 1.7% of the endosperm dry weight, but do not constitute part of the cell wall (Haskå, Nyman and Andersson, 2008)

In terms of range of AX contents in wheat and other cereals, the Healthgrain project performed a screen of multiple wheat varieties and other cereals to determine dietary fibre content. Pertaining to bread wheat, it was established that the cultivar Yumai-34 contained the highest amount of WE-AX amongst the panel, with a dietary fibre content of 1.4% of the dry matter (dm). In contrast, the lowest dietary fibre cultivar, Soissons, contained 0.3%dm of WE-AX (Gebruers *et al.*, 2008). Rye contained more WE-AX than bread wheats; oats; barley; spelt; T. monococcum and T. dicoccum, with WE-AX ranging from 1-1.5% dm in rye. Oats had the lowest

WE-AX distribution, ranging from 0.16-2.1%dm, and barley ranging from 0.15-0.41%dm (Table 5) (Ward et al., 2008). This would suggest that rye may prove to be a powerful target for introgressions into bread wheat to increase the amount of WE-AX available in commercial bread wheat cultivars, and indeed a translocation of the 1R chromosome onto the long arm of chromosome 1B in wheat exists, and has showed marked increase in AX content (Piro, Muylle and Haesaert, 2023).

Table 5: Table showing the range of WE-AX in cereals, and bread wheat cultivars from the Healthgrain project.

Cereal/Cultivar	WE-AX Content (% dry mass)	Notes	Source
Bread Wheat			
- Yumai-34	1.4	Highest among bread wheat cultivars	Gebruers et al., 2008
- Soissons	0.3	Lowest among bread wheat cultivars	Gebruers et al., 2008
Rye	1.0–1.5	Higher than wheat, oats, barley	Ward et al., 2008
Oats	0.16–2.1	Broadest range, but generally low	Ward et al., 2008
Barley	0.15–0.41	Narrow range, low values	Ward et al., 2008

1.10 Arabinoxylans

Arabinoxylans (AX) are a class of polymers that are prevalent in monocot secondary cell walls, and both primary and secondary cell walls of wheat grain. AXs are composed of two pentose sugars (arabinose and xylose), the proportions and distributions of which contribute to variation in properties, such as solubility and viscosity. AX consists of a linear β1-4 linked xylopyranose (Xylp) backbone with some residues either mono-substituted with arabinose at the α-(1,2) position (Figure 2A) or disubstituted at the α-(1,2) & α-(1,3) positions (Figure 2B). Arabinose residues may also be substituted with ferulic acid on the 5 positions of α-(1,2)-linked arabinose residues (Saulnier, Vigouroux and Thibault, 1995; Barron, Surget and Rouau, 2007). AX are the major components in the wheat endosperm cell walls with about 25% usually being soluble.

Although present in similar proportions, the AX present in the aleurone and SE cell walls have different physical properties. AX in the aleurone is highly esterified with a low arabinose:xylose ratio compared to the SE. This low ratio results in increased hydrogen bonding between xylan backbone chains, promoting aggregation. Similarly, ferulic acid (FA) comprises 3.2% AX dry weight in the aleurone (Antoine *et al.*, 2004) compared with 0.0103% in the SE cell walls calculated from Freeman et al., (2017). Oxidative dimerization of ferulic acid groups present on adjacent AX chains results in cross-linking, promoting aggregation. Ester linkage to p-coumaric acid also occurs in the aleurone but not in the starchy endosperm tissues (Barron, Surget and Rouau, 2007).

The AX in bran is highly modified with glucuronic acid, and hence is often called glucuronarabinoxylan, with much lower solubility due mainly to approximately 5x more cross-linking between ferulic acid.

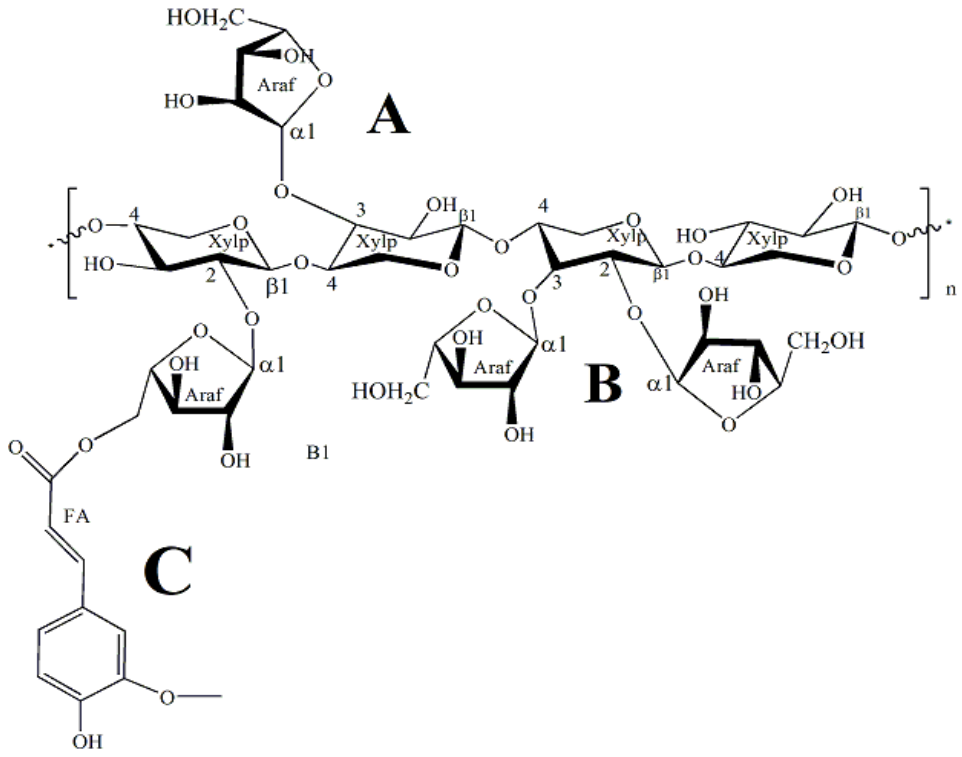


Figure 2: Representative chemical structure of arabinoxylan, including (A) Monosubstitution, (B) disubstitution, (C). feruloylation of mono-substituted Araf. Xylp represents xylopyranose residues, Araf represents arabinofuranose substitution.

1.11 Arabinoxylan biosynthesis

The synthesis of Arabinoxylan is a complex process, which involves multiple gene families for the synthesis of the xylan backbone, arabinose substitutions along the xylan chain, feruloylation and subsequent transportation to the cell wall and a simple schematic of the biosynthetic pathways are shown in Figure 3.

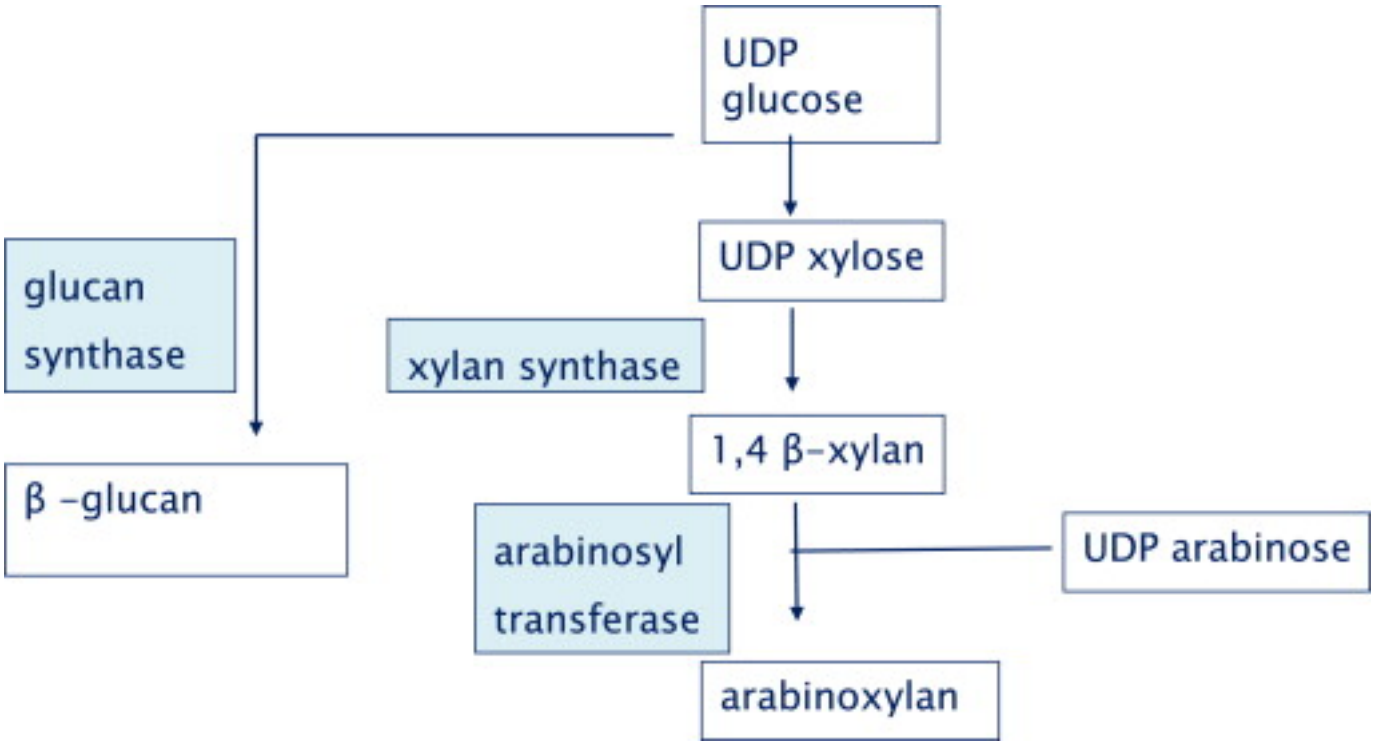


Figure 3: A schematic pathway of the arabinoxylan biosynthetic pathways that are currently known. Taken from(Lafiandra and Shewry, 2014).

AX synthesis takes place in the Golgi body(Carpita and Gibeaut, 1993) and is mediated by membrane channels and UDP-sugar transporters. Different isoforms of enzymes for xylose and arabinose synthesis exist in both the Golgi and cytoplasm. UDP-D-Xylose, which is the precursor of the xylose backbone of AX, is synthesised by UDP-Xyl synthase genes (UXS), via decarboxylation of UDP-glucuronic acid, which occurs in both the cytosol and Golgi (York and O'Neill, 2008). However, a more recent study that targeted UDP-xyl transporter genes (UXT) has shown that only the cytoplasmic isoforms of UXS are required for the synthesis of the xylan

backbone in Arabidopsis (Zhao *et al.*, 2018), in which triple mutants of uxt1, uxt2 and uxt3 had severe effects on secondary cell wall thickening, as well as defective galactoarabinoxylan chains. The arabinose residues that decorate the xylan backbone in arabinoxylans are produced from the conversion of UDP-xylose to UDP-arabinopyranose (Arap) mediated by UDP-xylose epimerase (UXE). Arap is transported from the Golgi into the cytosol where the enzymatic complex of UDP-arabinopyranose mutase (UAM) catalyses the conversion from UDP-Arap into UDP-arabinofuranose (Araf), which is subsequently transported back into the Golgi and substituted onto the xylan backbone (Botticella, Savatin and Sestili, 2021).

The xylan backbone is synthesised from the UDP-sugar precursors by enzymes of two families: gylcosyl transferases (GT), and glycosyl hydrolases (GH). The GT enzymes facilitate the transfer of UDP-Xyl and UDP-Araf monomers to acceptors, to catalyse the formation of glycosidic bonds that link the monomers together. GT43 and GT47 have been shown to be the main components for xylan synthesis(Urbanowicz *et al.*, 2014), while genes belonging to the GT61 family have been associated with the substitution of Araf onto the xylan backbone. These studies support the study of Mitchell et al., (2007), where a novel bioinformatics approach was used to identify potential gene families related to the synthesis of arabinoxylan in grasses. This study used expression sequence tags to detect heavily upregulated genes in grasses when compared to Arabidopsis. Four main gene families were highlighted as the strongest candidates, 3 of which were glycosyltransferase families, GT43, GT47 and GT61, which are all inverting enzymes of the type required for AX synthesis (Mitchell, Dupree and Shewry, 2007). While the GT43 and GT47 gene families are associated with xylan chain synthesis, the GT61 family has been shown to encode xylan arabinosyl transferases (XATs) responsible for the decoration of xylan with arabinofuranose residues in in rice and wheat (Anders et al., 2012).

The xylan backbone is synthesised in two stages, initiation and elongation. Until recently, little was known about the genes responsible for xylan chain production although a β-1,4 xylosyltransferase was identified as the enzyme responsible for xylan chain elongation (Baydoun, Waldron and Brett, 1989; Gibeaut and Carpita, 1990). Recent studies of the moss *Physcomitrella patens* enzymes suggest that β-1,4 xylosyltransferase is encoded by IRX10, which was shown to have the ability to extend xylan chains whereas the products of the IRX9 and IRX14 genes did not show this function. These three genes were initially identified in Arabidopsis and, although they have been subject to many studies, the catalytic activities of IRX9 and IRX14 have not been identified (Ren *et al.*, 2014; Freeman *et al.*, 2016; Zeng *et al.*, 2016). Consequently, it has been suggested that IRX9 and IRX14 are structural components of an enzyme complex. In support of this hypothesis, analysis of knockout mutants of IRX9 of Arabidopsis displayed irregular xylem vessels which lacked xylose (Brown *et al.*, 2005), suggesting a role in xylan synthesis.

Orthologues of IRX10 and IRX9 exist in wheat (*Triticum aestivum*) named TaGT47_2 and TaGT43_2, respectively, where their functions have been confirmed using RNAi silencing. RNAi constructs for GT47_2 also significantly downregulated the IRX10 paralogues GT47_1 and GT47_4. Suppression of GT43_2 resulted in short AX chains, compared to controls, while suppression of TaGT47_2 resulted in longer chains (Lovegrove *et al.*, 2013). It is possible that the xylan synthase activity of GT47_2 is sufficient to produce the shorter chains, but TaGT43_2 is required to organize the xylan synthase complex necessary to produce the longer chains, explaining why long chains exist in the TaGT47_2 deficient transgenic lines. There were 48% and 43% decreases in NSP xylan (which is present almost entirely in the AX backbone) in GT47_2 GT43_2 deficient lines, respectively. Moreover, the transgenic lines had up to 30% increased arabinosylation, with a 40% increase in the xylp:araf ratio. These IRX genes, and their homologues, may represent the core xylan synthesis machinery in plants which may be

influenced by a variety of cofactors, such as transcription factors. However, it has been suggested that the slow rate of xylan synthesis in *in vitro* assays observed in the IRX studies may be due to the requirement for optimal conditions, as present within the cell, and that unidentified members of the GT2 family may play a role in xylan elongation (Bulone, Schwerdt and Fincher, 2019).

1.11.1 Araf substitutions.

Gene families associated with arabinosylation have the greatest expression bias for monocots compared with dicots compared to other GT families (Mitchell, Dupree and Shewry, 2007) and have been experimentally shown to add Araf substitutions to xylan in wheat (Anders et al., 2012). These genes have thus been named xylan arabinosyl transferases (XATs). XATs are present in the Golgi apparatus (Anders et al., 2012)**,** where xylan synthesis predominantly takes place (Rennie and Scheller, 2014). TaXAT1 has been shown by RNAi suppression to be predominantly responsible for monosubstitution of α-(1,3) linked Araf residues on the xylan backbone, reducing mono-substituted Araf by 70-80%, while disubstituted Xylp was not significantly affected. It is also suggested that backbone formation and substitution must take place together, as the lack of monosubstitution did not result in increased unsubstituted Xylp (Anders et al., 2012). Suppression of TaXAT1 gave further evidence for this mechanism as TaXAT1 suppression lowered average AX chain length in wheat endosperm (Freeman *et al.*, 2016).

1.11.2 Feruloylation.

The bioinformatics study of Mitchell *et al., (2007)* also predicted gene members of the PF02458 to be involved in the feruloylation of GAX, under the same assumption that these genes would be significantly more abundant in grass expressed sequence tag (EST) libraries than dicots

(Mitchell, Dupree and Shewry, 2007). Later, Piston *et al.,* (2010) reported evidence for this function, showing that RNAi-induced down-regulation of four arabinoxylan feruloyl transferases (AFTs) in the PF02458 family of BAHD proteins resulted in a significant (19%) decrease in ester-linked cell wall monomers of FA. A significant decrease in FA was also observed in transgenic rice plants transformed with the pAFT-B constructs, which down-regulated the genes; Os05g08640, Os06g39470, Os06g39390 and Os01g09010.

1.11.3 QTL associated with AX in bread wheat.

During the course of the 2011 HEALTHGRAIN project, a panel of elite wheat cultivars were screened for bioactive compounds and DF. From this screening, the Chinese wheat cultivar, known as Yumai-34 was first identified as a particularly high AX cultivar during the HEALTHGRAIN project (Shewry et al., 2011).

Several studies have reported quantitative trait loci (QTL) for arabinoxylan content in wheat grain, although most reported studies have been of wholegrain rather than white flour and few studies aimed to develop molecular markers from the QTL loci for use in breeding. Martinant *et al*., (1998) reported that the content of WE-AX and relative viscosity (RV,) a measure of WE-AX) in aqueous extracts of white flour were highly heritable and identified a QTL accounting for 32-37% of the variability of viscosity on chromosome 1B using a population of 91 doubled haploid (DH) lines from the cross Courtot x Chinese spring (CtxCs). A meta-QTL analysis of AX in white flour, collating data from Martinant et al (1998) with analyses of DHLs or recombinant inbred lines (RILs) from crosses of CtxCs (Perretant *et al.*, no date), DH Arche x Recital (Laperche *et al.*, 2007), Renen x Recital (Quraishi *et al.*, 2009), Valoris x Isengrain and RE006 x CF007 (Charmet *et al.*, 2009), reported 12 QTL, of which 3 were also detected by association analysis (Quraishi *et al*.,2010). Charmet *et al*., (2009) identified a QTL on 1B, which appears to correspond to that described by Martinant *et al*., (1999). Similarly, Lovegrove *et al*., (2020) reported a 1B

QTL for RV and total AX (TOT-AX) in white flour in several crosses with the high-AX genotype Yumai-34 (Yumai34 x Claire, Valoris and Altigo) and developed markers to accurately predict AX content. The 1B QTL identified in these studies may correspond to 1BL QTL more recently identified by Ibba *et al*., (2021) using 125 modern bread wheat lines in a genome wide association study (GWAS). Ibba *et al*., (2021) identified two putative gene candidates in the glycosyl hydrolase (GH) 16 and GT61 families. GH16 enzymes have been implicated to be involved in cell wall modification in grasses (Strohmeier *et al.*, 2009), whilst GT61 enzymes are believed to transfer arabinosyl units to xylan in grasses (Anders et al., 2012 Suliman et al., 2013). However, no functional studies were carried out to identify the roles of either of these genes. Ibba *et al*., (2021) developed markers for the 1BL QTL with higher predictive power than the markers by Lovegrove *et al*., (2020). Perfect markers have not been developed for the 1B QTL and it has been suggested that using a combination of these two markers would be useful when breeding high fibre wheat lines (Ibba *et al.*, 2021). It is therefore crucial to identify the causal gene under the 1B QTL in order to select for AX in wheat breeding.

1.12 The Watkins Collection

Modern wheat varieties have been produced by intensive selection which has led to a loss of diversity compared to older types (Balfourier *et al.*, 2019). Sources of increased variation for wheat breeding are therefore being sought in older types of wheat, including landraces which are types of wheat which were grown before the introduction of scientific breeding. One such collection is the Watkins collection (Wingen *et al.*, 2014a) which consists of wheat cultivars collected by A.E Watkins, in the 1930s, through the London Board of Trade, mainly from countries in Europe, China and the former British Empire (India and parts of Africa). The original collection consisted of over 7000 accessions (Miller *et al.*, 2001) the current collection consists of 826 bread wheat accessions, with a core collection of 119 landraces selected to represent

98% of the genetic diversity (Wingen *et al.*, 2014a). A total of 804 accessions of hexaploid wheat from the Watkins collection were genotyped using the 35k wheat breeders' array (Winfield *et al.*, 2018) containing 35 143 SNP probes (Allen *et al.*, 2017). The accessions were grouped based upon their geographical distribution and compared to modern elite cultivars. The Watkins collection has been shown to contain higher genetic diversity than the Gediflux collection which comprises elite bread wheat lines from Europe and has been shown be a useful source of novel alleles, particularly for local adaptations to environments.

The ~120 accessions in the core Watkins collection and the Gediflux collection were also genotyped using the wheat high-density 820 k Axiom Array. The Gediflux collection constitutes 500 lines with extreme phenotypes and covered more than 5% of total wheat acreage from 1940 onwards. These were collected by John Snape and Tony Worland at the John Innes Centre. This study showed that 32.2% of total 218,106 polymorphic markers were unique to the Watkins collection and did not overlap with the elite Gediflux collection, highlighting the usefulness of the Watkins collection for novel gene discovery.

Together, the Watkins collection and the associated genetic data were therefore used in this project to identify novel QTL, genes, and alleles for the high AX trait. Genetic information on the Watkins collection, provided as part of a collaboration on the Watseq project, with the Agricultural Genomics Institute at Shenzhen (AGIS), China and the John Innes Centre, UK, allowed the identification of markers tightly linked with the previously identified 1BL high AX QTL and for novel QTLs identified using the Paragon x Watkins mapping populations developed at the JIC.

1.13 Aims and objectives.

The project had three aims:

1. To identify new QTLs for high AX in wheat, using three biparental RIL mapping populations derived from crosses between Paragon and Watkins landraces with high AX which were identified in a previous study (unpublished). A robust QTL analysis was performed on two of these populations over three years, and the third for two years, allowing the mapping of genomic loci to arabinoxylan content. These should facilitate the development of new modern varieties by breeders with a high AX content.
2. Fine mapping of the major high AX QTL previously identified on chromosome 1BL in order to reduce the QTL interval, identify a set of candidate genes for the high-AX trait, and design markers with more predictive power than those previously available.
3. To contribute to our knowledge of the control of AX synthesis in wheat by the identification of candidate genes at QTLs for high AX in white flour.

Chapter 2
Methods and Materials.

2.1 Mapping populations

Recombinant inbred lines (RILs) produced at the John Innes Centre, Norwich, UK derived from crosses of the elite spring wheat cultivar 'Paragon' with previously identified high AX lines from the Watkin's core collection. Three F_4 populations were used in this study from Watkins lines; 694,471, 145 and were given the nomenclature 'ParW' followed by the Watkins accession number, e.g., ParW694. Each population contained between 72-82 lines.

2.2 Field Trials

Randomized block design field trials were established in 2018, with the three populations, ParW145, ParW471 and ParW694, harvested in August 2019 with biological repetitions per line included in the design. In 2019 single biological samples for three populations (ParW145, 694 and 471) were grown in field trials. All three populations, with three biological reps for each population, were grown in 2021 (drilling in 2020 was delayed by heavy rainfall). Trials from 2019 were harvested prior to starting the project, but the trials from 2020 and 2021 were monitored, checked for weeds or rogue plants, which were removed by hand. Field plots were routinely checked for maturity and then hand harvested with a pair of scissors.

2.3 Milling

Initially, milling of 5g of grain was performed in a RETSCH centrifugal mill using a 0.25µm screen to produce ~5g wholemeal flour followed by further particle size reduction using a ball mill (Glen Creston) – using five ball bearings in a 5cm diameter canister, for 4 minutes. Care was taken to

randomize the samples, and ten lines were milled per hour to ensure no heat transference from the mill to samples.

2.4 Extraction of soluble arabinoxylan from wholemeal flour.

1g of wholemeal flour from each ParW line was mixed with 4ml of deionized water and incubated for 10 mins on a Spiramix at room temperature. The samples were not balanced before centrifugation, but pairs of samples matched as closely as possible. Following the first centrifugation, the supernatant was decanted to a fresh tube and heated to 100°C for 10 minutes. The sample was then spun a second time in corex tubes at 10,000xg for 10 mins, at 25°C; the supernatant was pipetted into Falcon tubes and was frozen at -20°C. This protocol was developed for the measurement of relative viscosity by capillary. However, the use of an in-line viscometer meant that this protocol could be scaled down to reduce screening times and increase throughput. When used for high performance size exclusion chromatography- multi angle light scattering (HPSEC-MALs), the thawed extract was passed through 0.45µm pore sized disposable filters prior to loading into the autosampler.

2.5 Relative Viscosity.

Relative viscosity was measured at 30°C using an automated viscometer, model: AVS S370 SI Analytics, Germany. The viscometer was fitted with a Micro Ostwald 2ml capacity, 0.43mm diameter capillary. The 'WinVisco' software was used for data collection. 2ml of the extract was carefully pipetted into the capillary for determination of relative viscosity. The viscometer was standardised prior to each run with a water control against which the viscosity was measured (ie. relative viscosity of extract compared to a water control)

2.6 HPSEC-MALs.

To determine the specific viscosity of the water-soluble extracts, an Agilent HPLC (Infinity 1260 series) fitted with a standard autosampler, Infinity 1260 multi-detector suite (refractive index detector), DAD module, and inline Wyatt 'ViscoStar' viscometer and miniDAWN TREOS (MALS detector) was used. The high performance liquid chromatography (HPLC) was equipped with size-exclusion columns; two in series- a guard column (Shodex, OH-pak SB-G), followed by 'SB 806.M HQ' and 'SB 804 HQ'. This combination of columns was able to separate AX from proteins and other water-soluble components in the extract. The eluant was 0.01M $NaNO_3$ containing 0.001% NaN_3. 100µl of filtered water-extract, with a concentration of 250mg/ml, was injected, total runtime per sample was 90 minutes. Silicone covered Agilent HPLC snap-caps were used to prevent evaporation from vials over long runs. Data was collected using the 'Astra, v6' software from Wyatt.

2.7 HPSEC Peak selection and specific viscosity calculation.

The lower and upper elution volume limits of 11-20.5ml were selected as the AX peaks. This was previously confirmed by xylanase and lichenase digestion of extracts (Lovegrove *et al.*, 2013). To calculate the specific viscosity (η_{sp}), a consistent baseline was set for differential pressure (DP) within Astra with an upper limit at 20.5 units- consistent with the maximum elution volume of the AX peak. Values of incremental η_{sp} (here referred to as $i\eta_{sp}$) between 11-20.5ml which formed the peak of the chromatograph (Figure 4) were used to calculate the area under the peak using the equation:

$$A = \frac{inc \times \left(\Sigma i\eta_{sp}\right)}{injection\ volume}$$

where A = area under the chromatogram, '*inc*' = the incremental change in volume between measurements of $i\eta_{sp}$. The area directly translates to the average η_{sp} of AX in the ParW line and is the value that was directly correlated with relative viscosity and the value used to perform QTL analysis. The incremental viscosity values were exported into Microsoft Excel, where the area could be calculated.

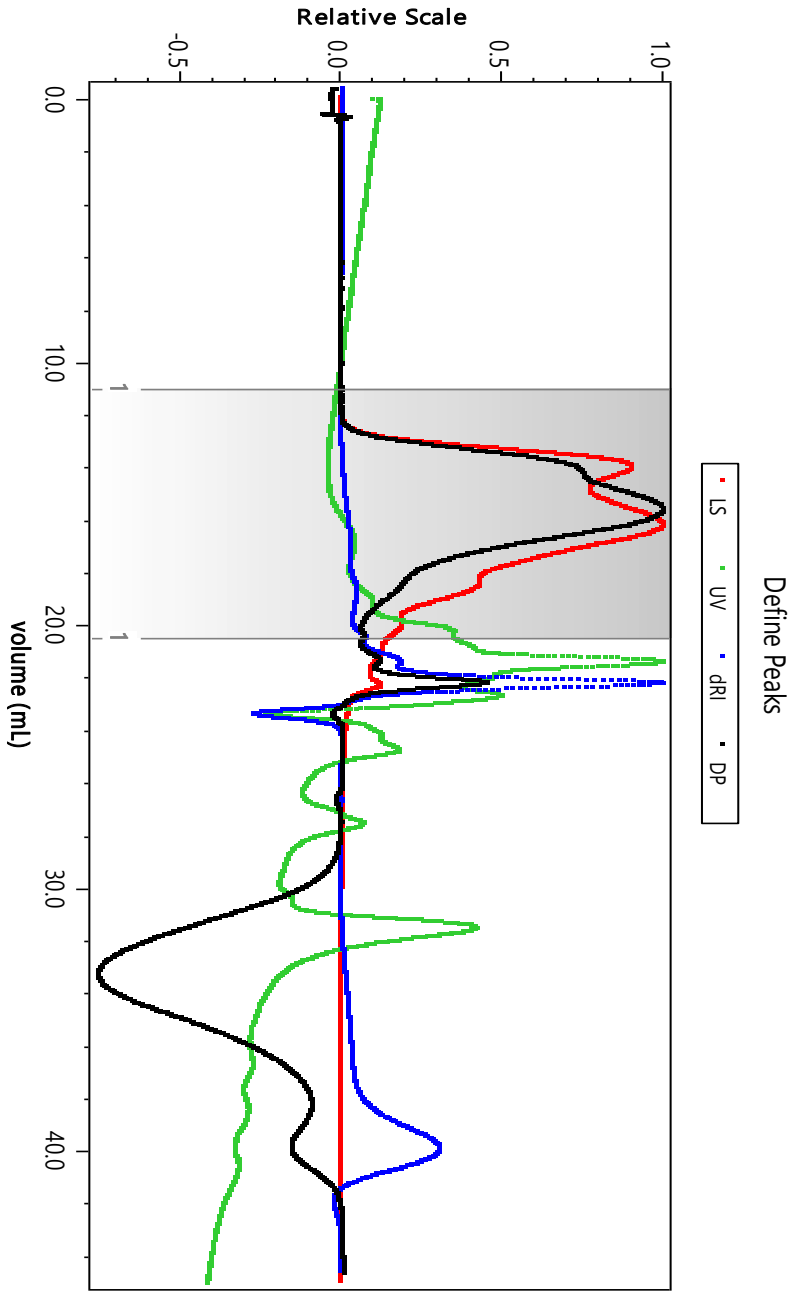


Figure 4: Representative chromatogram of line ParW145_002. The region of 11-20.5ml represents AX in the sample. Differential Pressure (DP) (Black line) represents the change in specific viscosity. Red line shows the change in light scattering. Blue line shows the differential refractive index, and green line shows a change in UV.

2.8 Colorimetric determination of pentosans

Project partners at KU Leuven, Belgium, carried out spectrophotometric measurement of AX for field trial material, using a method adapted from (Douglas, 1981; Finnie et al., 2006; Kiszonas et al., 2012) . A linear fitted calibration curve calculated as ΔA (= A553-A553, blank) as a function of xylose content (mg) was used to estimate WE-AX and TOT-AX content, using the equations:

$$\%\, WE - AX = \frac{\left({}^{\Delta A_{sample}}/_{a}\right) * 0.88}{\left({}^{M_{sample}}/_{5}\right)} * 100$$

$$\%\, TOT - AX = \frac{\left({}^{\Delta A_{sample}}/_{a}\right) * 0.88}{\left({}^{M_{sample}}/_{3}\right)} * 100$$

here ΔA_{sample} is the corrected absorbance of the sample, '*a*' is the slope of the calibration curve, M_{sample} is the sample weight (mg), and is divided by the dilution factor in the equation. 0.88 is a correction for the weight of water incorporated upon acid hydrolysis. This same method was used to measure total pentosans (TOT-AX) in population ParW471 2019, only.

The colorimetric determination of pentosans was also performed at RRes, on the bulked segregant ParW material, collected from the John Innes Centre. For this, I used an alternate, high-throughput methodology developed by CYMMT, Mexico (Hernández-Espinosa *et al.*, 2022). However, to reduce biological variation between samples, 100mg of wholegrain flour was measured, and diluted with 10ml of distilled and deionised water, then vortexed to suspend and rotated for 30 minutes on a Spiramix and then centrifuged at 2,500g for 10 minutes. A 75 µl aliquot of the aqueous fraction was taken and diluted, in a 2ml 'Safelock' Eppendorf tube, with

75 µl distilled and deionised water and then hydrolysed with the extraction solution, comprising of 93% Glacial acetic acid, 1.7% HCL, 0.86% Phloroglucinol, from Sigma Aldrich, and 0.014% glucose. The Eppendorf tubes containing the flour extract and extraction solution were incubated at 99**°C** for 23 minutes, at 700 RPM. Samples were then cooled briefly on ice, before measuring colour intensity at wavelengths 552 and 510 nm on a SPECTRAmax 340PC, by Molecular Devices.

2.9 QTL detection.

An in-house script for the package 'R/qtl, *v1.47-9*' was developed by L. Wingen, (2012) at the JIC (https://zenodo.org/records/12686181). The script performs a two-phase scan. The first scan detects cofactors with a significance (α) of 0.4 and the second scan performs composite interval mapping (CIM) to detect robust QTL at α of 0.05. Genotyping data is freely available for the ParW populations in this project from 'https://designingfuturewheat.org.uk/resources/'. Populations were genotyped using kompetitive allele specific polymerase (KASP) genotyping at Bristol University as part of the Wheat pre-breeding LOLA project. Furthermore, 'CerealsDB' provides a search function for SNP markers, which were anchored on Chinese spring RefSeq v.1.1, which links the relevant page on Ensemble plants. This was used to show the physical positions of QTLs on chromosomes in the RefSeq v1.1 genome. Table 6 shows the number of markers available in each mapping population.

Table 6 :The number of KASP markers available for QTL detection in the three RIL ParW populations. The table shows the maximum number of lines in ParW populations, not accounting for failures in the field.

Genotype/population	No. of KASP Markers	Lines in 2019	Lines in 2020
ParW145	182	82	82
ParW471	281	79	79
ParW694	184	72	72

2.10 Pangenomic diversity of 1B QTL

The wheat10+ panel was designed to capture the genetic variation of hexaploid wheat, to facilitate gene identification and analyses of structural variation of genomic loci (Walkowiak *et al.*, 2020). This Panel consists of cultivars: 'ArinaLFor', 'CDC Stanley', 'CDC Landmark', 'Longreach Lancer', 'Norin61', 'Mace', 'SY Mattis', 'Jagger' and 'Julius'. To determine if Chinese Spring contained all of the endosperm-expressed genes within the 1B QTL for candidate gene identification, genetic diversity across the 1B QTL was assessed using the Wheat10+ Pangenome panel (release 52, annotation V PGSB2.1 for all, and PGSB v 2.2 for CDC_Stanley). Subsequent analyses were performed on the high-performance cluster (HPC) at Rothamsted Research to develop a novel pipeline for candidate gene discovery using the wheat pangenomes for the 1B QTL.

2.10.1 Blast searches

The flanking markers of the 1B QTL (AX-94766122, End: AX-94840083) were queried against the pangenomes, with ncbi-blast+(v2.12.0) and gave the physical position of the QTLs in each cultivar. All sequences were blasted via command line, through the Linux HPC at Rothamsted Research, using in-house Bash scripts. The wheat pangenomes were indexed at Rothamsted Research and used to query sequences. Blastn was performed on flanking Axiom marker sequences, using default parameters to map markers against the wheat pangenomes.

2.10.2 Acquisition of Gene IDs

Gene ids were selected from a SQlite3 database developed by myself, which contained the gene accession tag, chromosomal physical location and the chromosome name. Genes within the 1B QTL region of each cultivar was identified by selecting annotated gene accessions within the physical locations which the flanking markers aligned to in each of the pangenome cultivars.

2.10.3 Proteome sequence extraction across the 1B QTL

The predicted proteome of the 1B QTL region was selected from the published peptide database for each cultivar, available at Ensembl Plants (pangenomes) or NCBI (RefSeq v1.1 and v2.1). This was filtered for primary transcripts, defined as the longest of the transcripts, omitting transcript variants that occur in the gene annotation due to alternative splicing, so as to only find orthologues of the main gene model in each of the pangenome cultivars.

2.10.4 Pairwise gene comparison of pangenome and IWGSC cultivars

The proteome sequences were input into OMA (v2.5.0) to assess pairwise orthologous groups of proteins between pairs of cultivars. This enabled the identification of annotated proteins that were common amongst cultivars, as HOGs (hierarchical orthologous groups), or those that had been deleted or duplicated. This was repeated with nucleotide sequences to compare non-coding elements within the QTLs.

2.10.5 NUCmer

1B QTL regions were extracted from the soft-masked chromosome assemblies at Ensembl plants/NCBI using the python package 'Screed'. An interactive Python3 tool was written to extract the sequence between the bp positions previously identified in the Blastn of flanking markers, for each cultivar. The soft-masked 1B QTL regions were then aligned pairwise using the NUCmer tool (available http://mummer.sourceforge.net/, accessed March 2022). The alignment was parsed through the *delta-filter* tool included, with option '-g'. This filters the "global alignment using length*identity weighted LIS", which leaves the longest alignments. Pairwise dot plots of alignment were drawn with '*mummer-plot*' and compiled with '*Gnuplot'.*

2.10.6 Circos

Common genes were visualised using '*Circos'*- a perl-based application. Links were drawn between HOGs, which correspond to common genes between cultivars.

2.11 RNA-seq mapping of Yumai34x Valoris

Bulked paired-end RNA-seq read data was produced by Mitchell et al., 2023 (bioRxiv preprint doi: https://doi.org/10.1101/2023.03.08.531735) which were composed of 4 groups of 6 replicates. The groups were derived from pooled samples of Yumai-34 x Valoris crosses depending on the allele at the 1B and 6B QTL defined by Lovegrove et al., (2020). For the purposes of studying the 1B QTL, the two groups A and D contained the high 1B allele from Yumai34 and groups B and C contained the low allele from Valoris.

The raw untrimmed reads were mapped to different indexed reference genomes using HISAT2 (v2.2.1) (Kim, Langmead and Salzberg, 2015), using default parameters and output as a BAM file.

2.11.1 Feature Counts

Using the latest version of the annotation files for Chinese spring v1.1 and 2.1, and Lancer, Feature counts were calculated using the FeatureCounts tool (v2.0.0), which calculated the number of reads mapping at annotation features. For the later acquired Paragon 1B assembly (Hall, *et al., early access via the Delivering Sustainable Wheat (DSW) project)*, the 1B chromosome sequence of Paragon was manually inserted, in place of chromosome 1B in Chinese spring 2.1, and Liftoff, available at https://github.com/agshumate/Liftoff, was used to transfer the annotation of Chinese spring v2.1 to the newly inserted Paragon1B chromosome. This newly acquired annotation of Paragon chromosome 1B was used to perform the feature counts of RNA-seq transcript reads

2.11.2 Differential Gene Expression analysis

To calculate the differential gene expression in the Yumai-34 x Valoris RNA-seq data, the feature counts derived for each of the reference genomes underwent normalisation and differential analysis using the standard DESEQ2 pipeline defined by (Love, Huber and Anders, 2014).

2.12 Bulked Segregant analysis

The John Innes Centre hosted a collection of Paragon x Watkins RIL populations which allowed for the selection of 34 genotyped populations, with different Watkins parents to be sampled. Where possible, each of these populations had 30 lines selected, where 15 lines had the 1B high allele determined from Axiom or KASP markers overlapping with the physical region of the 1B QTL detected in Paragon. Similarly, the other 15 lines had the Watkins allele at this same locus. The sampled grains had thousand grain weights measured and for each population, an equal weight of grains was taken from each of the 30 lines. These were pooled together into their respective groups containing the Paragon or Watkins allele. This gave two samples per population which were milled and assayed for WE-AX. A one-tailed t-test was conducted on the WE-AX measurements between the two bulk groups for each population to determine if a statistically significant difference was observed.

2.12.1 Haplotype and Variant calling in the Watkins parents

The haplotypes of the Watkins parents were called automatically by AGIS, China as part of the Watseq collaboration (Cheng *et al.*, 2023), which were called against Chinese spring v1.1. However, to get a finer look at the variants, access was provided to the raw sequencing reads produced from Illumina short-read sequencing with 30x coverage.

I then mapped the sequencing reads to a custom reference genome of Chinese Spring v2.1, with the Paragon 1B chromosome, using bowtie2 (v2.5.2) and the following parameters -p 16 -q -N 0 -X 450 –fr. The bam files were sorted and indexed with Bamtools (v2.5.2).

The variants were called using BCFtools call -snps, as it had previously been shown to offer the highest precision in conjunction with bowtie2 alignments (Cagirici *et al.*, 2021)

2.13 KASP marker design

KASP markers were designed by extracting 100bp sequence from the reference genome either side of the SNP and feeding into polymarker at http://www.polymarker.info/. These sequences were then sent to the JIC genotyping service to synthesise and test on selected lines and populations.

Chapter 3
Determination of AX in the P x W RIL Mapping Populations and QTL Analysis.

3.1 Introduction

Paragon is a UK elite spring wheat classified in the highest quality group for breadmaking (UKFM Group 1) and was selected by the John Innes Centre (Norwich, UK) as the parent to develop ~100 mapping populations of recombinant inbred lines (RILs), by crossing Paragon with different Watkins Landraces.

The Watkins core collection of 120 lines was screened by Reynolds et al., 2018 (unpublished) to determine total and water-extractable arabinoxylan (AX) in wholemeal flour. This allowed the selection of ~40 lines which differed in their AX content. These were grown for three years at Rothamsted Research, from 2012-2014, and white flour analysed for the composition and amount of AX using enzymatic fingerprinting, pentosan analysis and monosaccharide analysis. Based on these results, three Paragon x Watkins RIL populations were selected, with the following Watkins parental accessions: 11900145, 11900471 and 11900694, hereafter referred to as ParW145, ParW471 and ParW694. The three parental accessions were selected based upon having consistently high AX content across the three years. Seeds of the three RIL populations were provided by the JIC to be grown at Rothamsted Research for three years, from 2019-2021, for phenotyping and QTL analysis.

3.1.1 Relative Viscosity

Relative viscosity was used as a screening method for soluble (water-extractable) AX (Lovegrove et al 2020). Relative viscosity measures the viscosity of a solution, relative to the pure solvent, in this case water.

3.1.1.1 Relative viscosity vs Specific Viscosity

Although relative viscosity (η_{rel}) is a very effective proxy for soluble AX content, it is a low-throughput technique. Therefore, to increase the number of samples that could be screened, specific viscosity was determined using an in-line viscometer linked to HP-SEC-MALS chromatography. Specific viscosity is the incremental change in viscosity of a solute measured via the in-line differential viscometer in the HPLC set-up (differential because it compares the flow of eluent to the sample). The relationship between specific viscosity and relative viscosity was first validated, and a linear relationship was established between η_{rel}-1 (relative viscosity minus 1 and specific viscosity η_{sp} (Fig 4), using aqueous extracts produced as described in Chapter 2.

3.1.2 Axiom and KASP Genotyping Arrays

The Axiom 35K Breeders Array is a commercial array used by wheat breeders and researchers, which was developed by Bristol University in collaboration with ThermoFischer Scientific ^{LTD} (Allen *et al.*, 2017). The 35K array was derived from the Axiom 820K array which was originally designed to compare hexaploid wheat to wheat wild relatives. The 35K array is composed of a set of 35,143 markers that are useful to breeders and was used by Allen *et al*. (2017) to screen both elite cultivars of the Gediflux collection and landraces from the Watkins Collection (Wingen *et al.*, 2014b). The Axiom 35K breeders' array has proved to be a useful resource for QTL detection for many traits and genetic linkage maps derived from the array are valuable tools for identifying genomic targets in populations that can be used to design accurate markers for breeding. Since the development of the 35K array, it has been used to successfully identify QTL associated with a variety of traits, including Yellow Rust resistance in durum wheat (Alemui *et al.*, 2021) and spike fertility index (Alonso *et al.*, 2021), height, heading and yield (Farre Martinez *et al.*, 2021) in bread wheat. It has also been used to map traits related to dietary fibre, including relative viscosity, WE-AX and TOT-AX from monosaccharide and pentosan analysis, in white flour

(Lovegrove et al., 2020a). Prior to ParW145, ParW471 and ParW694 being genotyped on the Axiom 35k Breeders' array, a linkage map composed of KASP markers was available. These maps were comparatively low-density, being comprised of <200 markers across the genome, but were still initially viable to use for a preliminary QTL detection, though the increased marker density afforded by the Axiom linkage maps may improve the accuracy of QTL detection and eliminate erroneously detected QTL

3.2 Results

3.3 Determination of the effect of freezing on viscosity.

To determine whether the viscosity of aqueous flour preparations was affected by freezing the samples, aqueous extracts were made from 8 randomly selected samples of different genotypes from the ParW population and analysed for relative viscosity before and after freezing. Figure 5 compares the relative viscosities of the freshly isolated extracts and aliquots of the same extracts which were frozen overnight at -20^{o}C. Paired two-sample t-test for means did show significant differences between the freshly prepared samples and those frozen overnight ($P<0.05$), full P values available in supplementary table 1. Additionally, the frozen samples showed greater standard deviation in the measurements; however, the order of the samples, when plotted as relative viscosity values, lowest to highest, was not affected. It was therefore decided that *all* samples would be frozen prior to assay, as this permitted more samples to be processed at a time and increased the throughput of this screening process.

Furthermore, the use of specific viscosity as a screening methodology allowed the use of smaller flour samples, as the extraction protocol for specific viscosity was reduced to a 0.25x scale of the original (RV) method. Paired two-tail t-tests showed no significant differences between the means of the three extracts at the original scale or 0.25x scale ($p=0.64$).

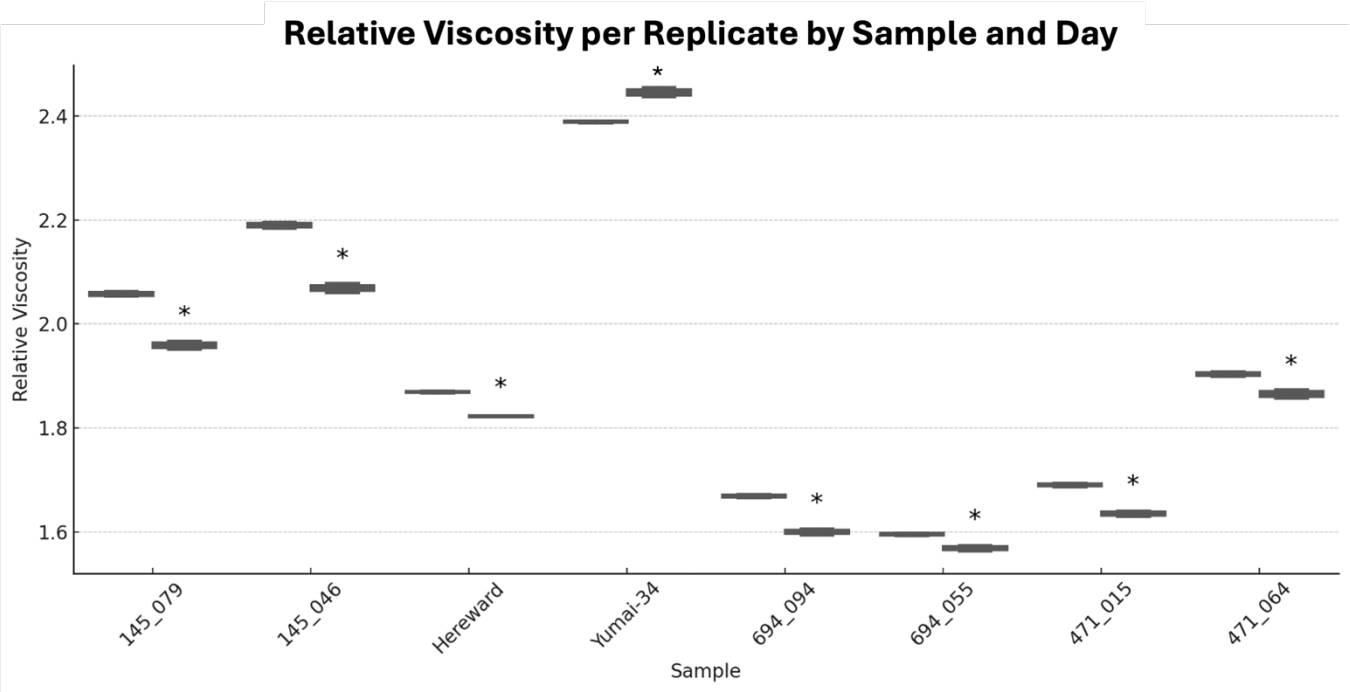
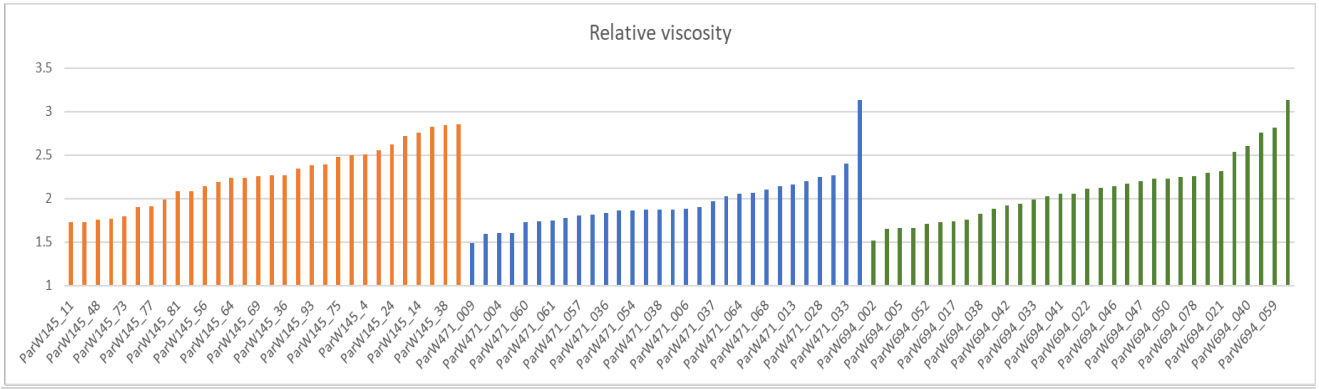


*Figure 5: Boxplots showing the technical replicates of relative viscosity of each sample on day 1, or day 2 (marked with *). Both samples are aliquots of the same sample, where day 1 samples were run immediately, and day 2 samples were immediately frozen overnight and subsequently thawed before analysis.*

3.3.1 Correlating Relative Viscosity with Specific Viscosity

Aqueous extracts were prepared from 93 randomly selected lines, grown in 2019 at RRES, from the three Watkins x Paragon populations (ParW145; ParW471 and ParW694) and used to determine both relative viscosity and specific viscosity as described in Chapter 2, section 2.4. In most cases, samples with high η_{rel} also have high specific viscosity (Figure 6), although this is not always the case. In some cases, η_{rel} relative viscosity value is not reflected in a high η_{sp} specific viscosity value, for example in the ParW694 population (green bars) line 694_060 has the highest η_{rel} measurement, but only the fourth highest η_{sp} (Figure 5). The relationship between η_{rel} and specific viscosity is, however, very strong (Figure 7), with a Pearson Coefficient of 0.8713.

A



B

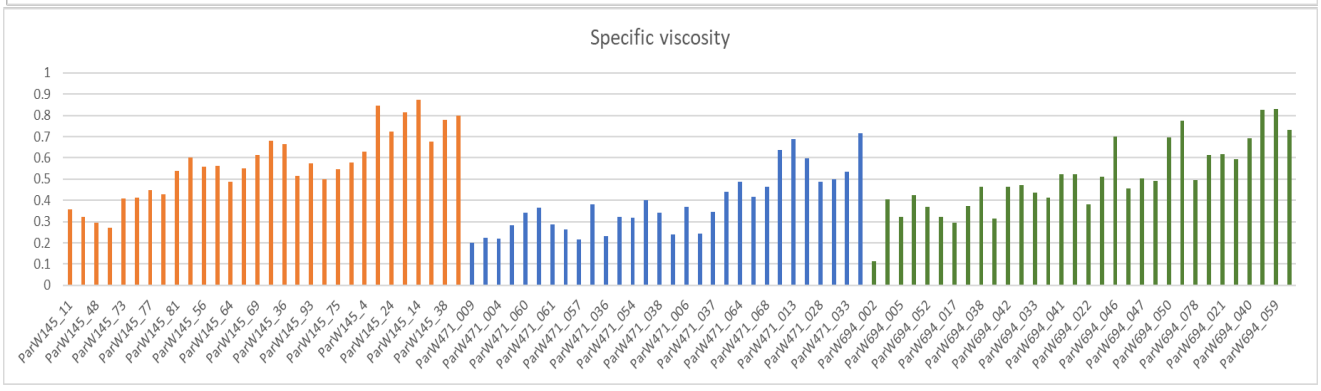


Figure 6: Bar chart showing ηrel (A) of the 93 lines assayed, across the three genotypes ParW145 (Orange) n=30, ParW471 (Blue), n=30, and ParW694 (Green) n= 32, compared with ηsp of the same lines (B). Both are ordered on ascending values of relative viscosity. Replicates per sample are not available for this dataset.

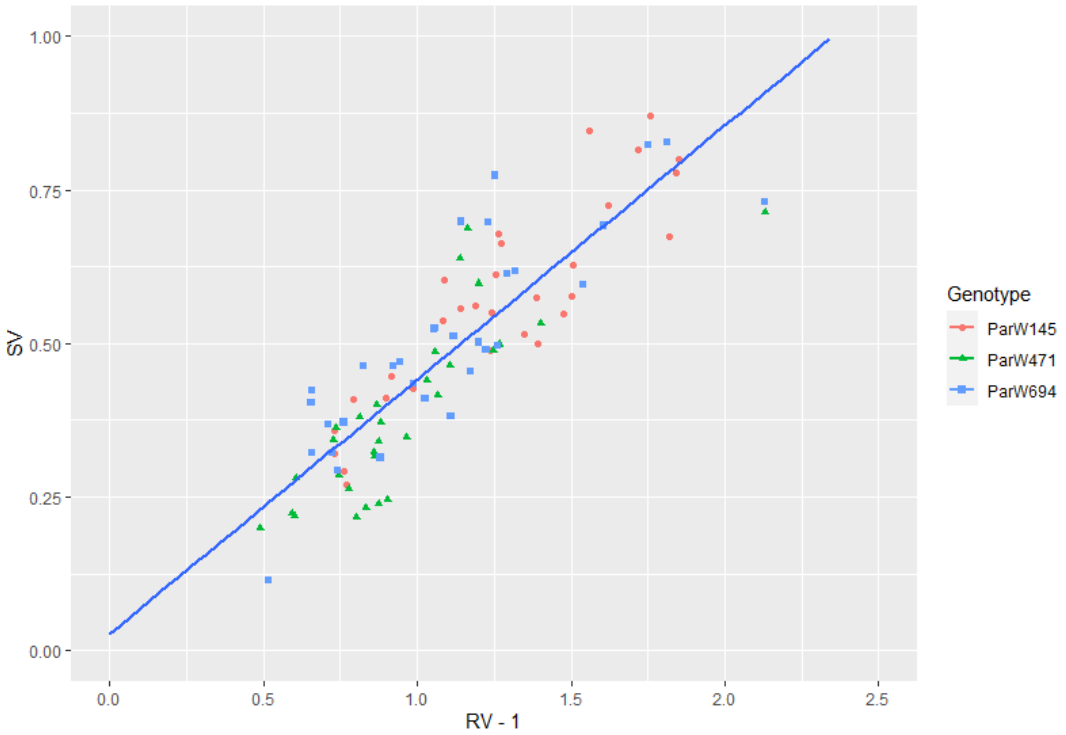


Figure 7: Linear regression of specific viscosity (SV) as a function of relative viscosity minus 1 (RV -1). the regression model is given the equation y=0.4136 + 0.0277. n=93 for the total number of samples. Regression equation calculated as a function of all samples, and not at the population level

75% of the variance is explained by the linear model (R^2=0.759). To keep both η_{rel} and η_{sp} on consistent axes and obtain an accurate measure of the y-intercept, η_{rel}-1 was plotted, as the minimum value of η_{rel} = 1, equal to the viscosity of water. The y-intercept is close to 0, at 0.0277, which could be explained by baseline noise in the HPSEC-MALLS system. ANOVA showed a highly significant association between η_{rel} and η_{sp} (P=0.0001), suggesting specific and relative viscosity are intrinsically linked. Therefore, specific viscosity can be used to estimate relative viscosity within ~0.5 units of η_{rel}, which was sufficient to identify high and low AX lines. The variation between RV and specific viscosity may result from the broad nature of the η_{rel} assay, which

measures the viscosity of a solution as a whole, whereas specific viscosity can be refined to a single component of the solution.

The specific viscosity measurement may represent a better proxy for WE-AX as only the peak relating to AX content is sampled for the measurement. It was therefore decided to use specific viscosity as the screening method for WE-AX content, as it allowed many more samples to be processed at a time, was a more precise measure of soluble AX (as only the peak relating to AX content was integrated for the measurement), and overall was faster to collect. Additionally, information could also be collected related to polymer size during the run. Therefore, specific viscosity was one of the phenotyping methods used to analyse the field trial material.

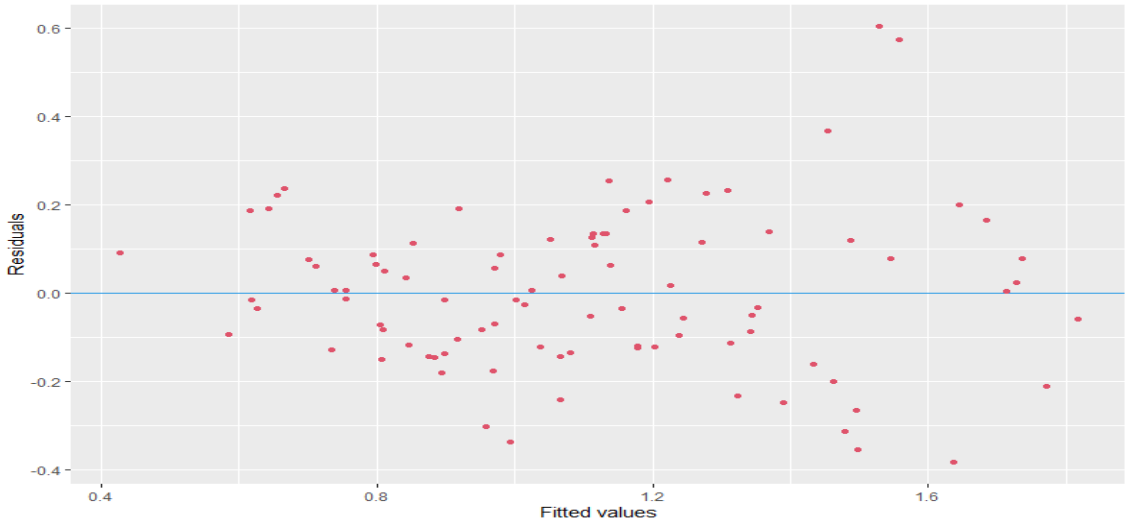


Figure 8: Residuals plotted of scaled version of specific and relative viscosity data against fitted values.

The residuals plot shows no discernible pattern (Figure 8) suggesting that the data meet the assumptions of linear regression, with no systematic error occurring within the data. However, the residual plot in Figure 7 does show that some outliers are present around the 0.6 and -0.4 residual values. To highlight possible outliers, boxplots for η_{rel} and η_{sp} were produced (Figure 9). These show no outliers for η_{sp}, but two outliers (one for each) were detected for η_{rel} for ParW471 and ParW694. The outlier in ParW471 (line 471_030) also displayed the highest η_{sp} value of the

whole population (η_{sp} = 0.71), which was followed by line ParW471_013, which had an η_{rel} of 2.1 and η_{sp} 0.68, suggesting that this outlier is genuine. The difference between the highest η_{rel} and second highest could be explained by components that could increase viscosity in the extract but are not accounted for by η_{sp}, which is centred on the known AX peak in the chromatogram (section 2, Fig 4). Tukey's HSD test shows that ParW145 and ParW694 are not significantly different for both η_{sp} and η_{rel}, (p= 0.2563 and p=0.1418, respectively). Conversely, ParW471 is significantly lower than both ParW145 and ParW694 in η_{sp} (P= 0.0001, and P=0.0111). For η_{rel}, there is a significant difference between ParW471 and ParW145 (P= 0.0018), but not ParW471 and ParW694 (P=0.1488).

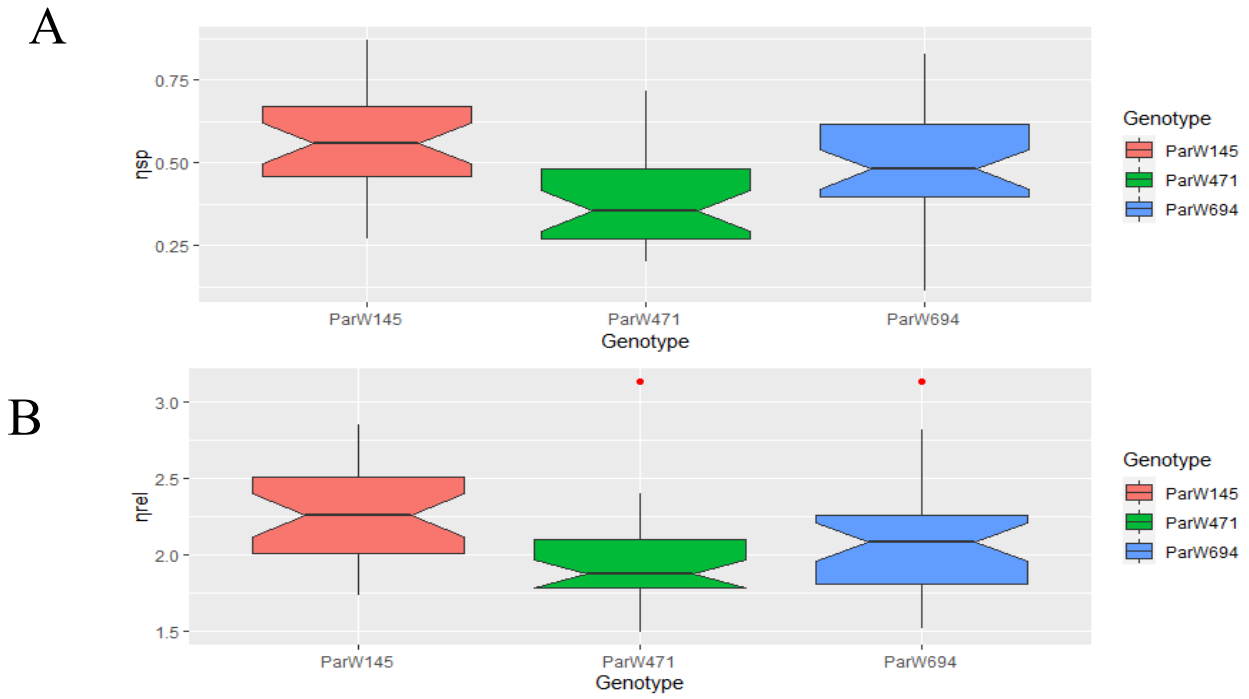


Figure 9: Notched boxplots of specific viscosity η_{sp}, (A) and relative viscosity, η_{rel}, (B) for the three genotypes used in the linear regression model, Parw145 n=30, ParW471 n=32, Parw694 n=30. Both η_{rel} and η_{sp} are calculated as ratios of viscosity, and thus have no units

3.4 Phenotyping using Specific Viscosity(η_{sp})

The remainder of the populations were assayed for η_{sp}, for a total of 186 samples, (single biological replicates) from field trials harvested in 2019 (ParW145, ParW471 and ParW694),

were screened for specific viscosity. There were some clear differences between the populations, as shown in Figure 9. The median for ParW145 differed significantly from ParW471/694 (P<0.001), as shown by the non-overlapping notches on the boxplots, but ParW471 and ParW694 do not differ from each other (P=0.23) (Figure 10). Independent t-tests showed that the mean viscosity of ParW145 was significantly greater than those of ParW694 and ParW471 (p<0.05 and p<0.0001). All three populations showed wide variation between lines, however, which allowed for effective QTL detection. All three populations also showed approximately normal distributions, with slight deviation of the means from the medians in ParW471 and ParW694. No values were excluded from QTL analysis, as they could represent genuine high viscosity lines and have high soluble AX.

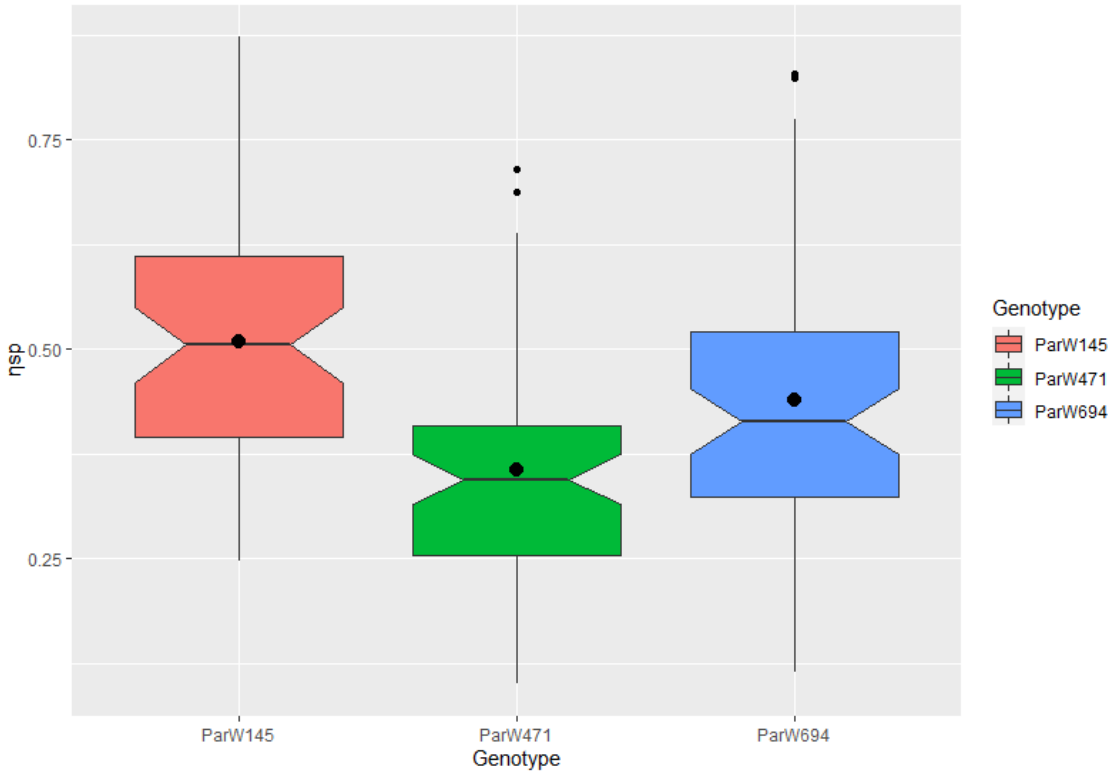


Figure 10: Boxplots showing the distribution of specific viscosity across the three ParW populations in 2019 rep 1. Means are represented by central black dots, and the notches around the medians show confidence intervals of the medians.

3.4.1 Phenotyping for pentosans

Total and water extractable (TOT- and WE-) pentosans were determined on wholegrain flour at KU Leuven, using the phloroglucinol colorimetric method described in section 2.4. WE-pentosans were measured in all three populations (ParW145, ParW471 and ParW694). ParW471 and Par694 populations were screened over three years (2019-2021); ParW145 for 2 years and TOT-pentosans in ParW471 for three years. Following screening for two years of the ParW145 this population was dropped from phenotypic screening as no significant QTL were identified during that time (2019-2020). TOT-pentosans measurements could not be completed for the ParW694 population in 2019 owing to sample loss at KU Leuven.

The assumption that phenotypic data are normally distributed must be met for QTL analysis. To test whether the data were normally distributed, density plots of the pentosan data were made for ParW145, ParW471 and ParW694 populations (Figures 11-13) and a Shapiro-Wilks statistical test of normality was performed on each dataset. The two WE-pentosan datasets ParW145 from 2019 and 2020 showed approximately normal distributions (Figure 11), with P -values of 0.49 and 0.6 respectively.

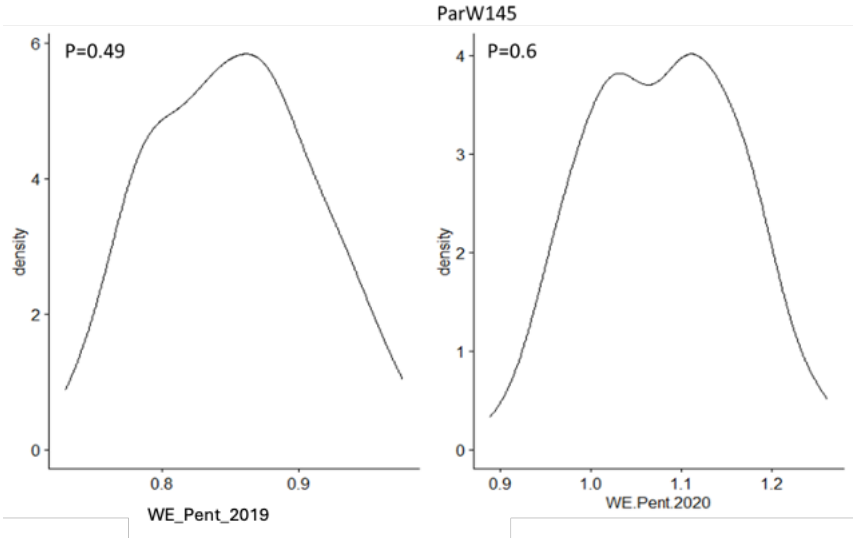


Figure 11: Distribution of mean WE Pentosans in population ParW145 in 2019 and 2020. Only a single biological replicate was available in 2020.

Similarly, the datasets for WE- and TOT-pentosans for years 2019-2021 all showed normal distributions in the ParW471 population. Although a bi-modal distribution may be inferred from the density plot in TOT-pentosans year 2019 and 2020 (Figure 12), the Shapiro-Wilks test of normality showed that both datasets came from normally distributed populations, where P = 0.77 and 0.10 for TOT-pentosans in 2019 and 2020, respectively, and so therefore did not need transforming for a QTL analysis.

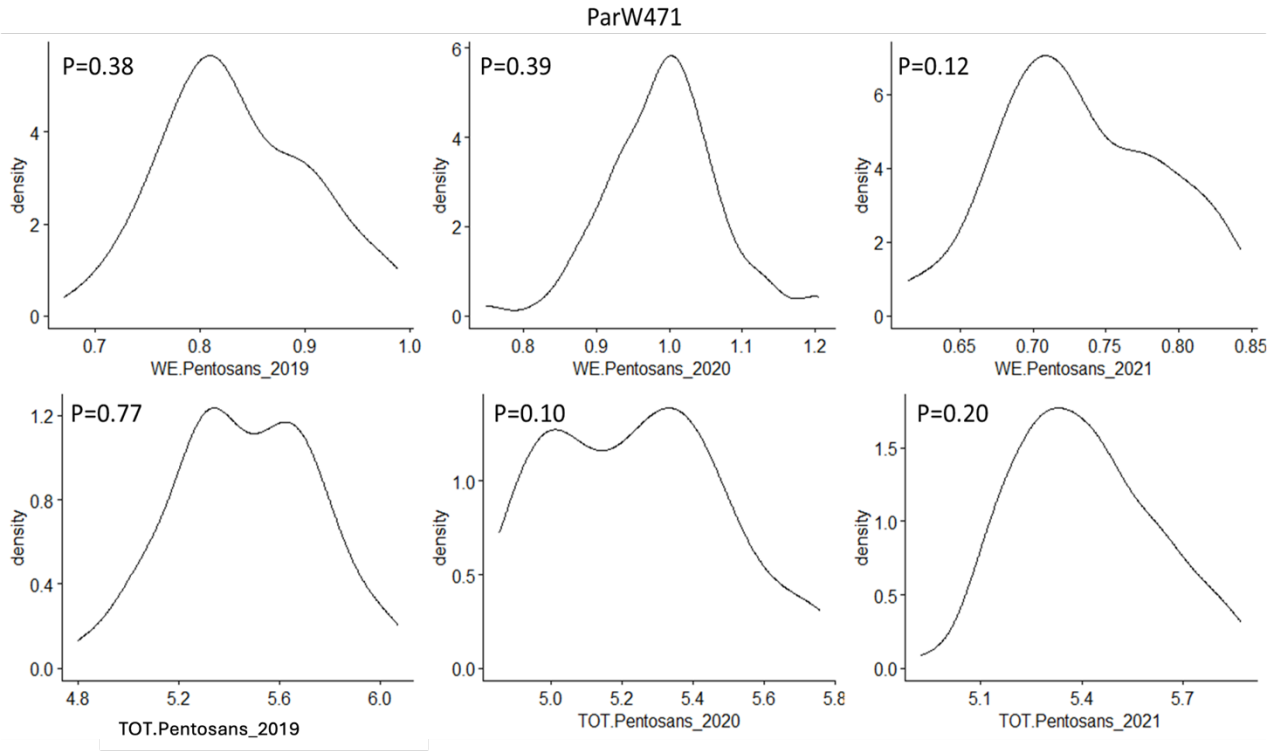


Figure 12 :Distribution of WE-Pentosans (Top) and TOT-Pentosans (Bottom) in the population ParW471 in trial years 2019-2021.

The WE-pentosan datasets from population ParW694 also displayed normal distributions for all three years (2019, 2020 and 2021), with the Shapiro-Wilks test, giving values of P=0.75, 0.61 and 0.58, respectively. Similarly, the two TOT-pentosan datasets from 2020-2021 also met the assumptions of normality, P=0.16 and 0.82 respectively (Figure 13). Because all of the pentosan datasets from the three populations were not significant using the Shapiro-Wilks test, the null hypothesis that these data come from a normally distributed population cannot be rejected, and they are therefore suitable for QTL analysis.

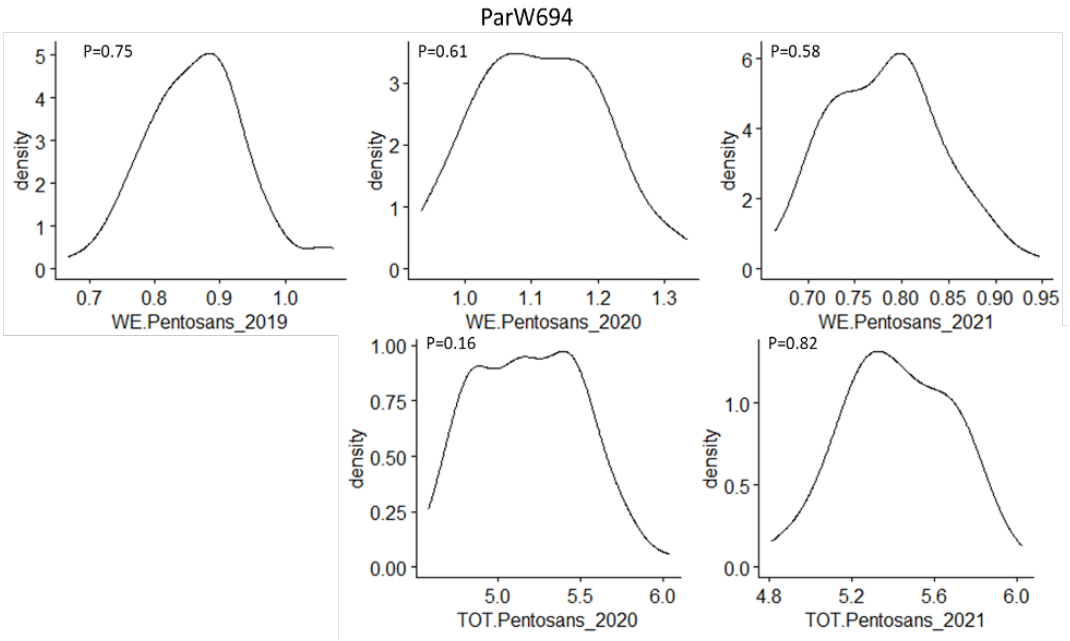


Figure 13: Distribution of WE-Pentosans (Top) and TOT-Pentosans (Bottom) in the population ParW694 in trial years 2019-2021 for WE-AX and 2020-2021 for TOT-AX.

3.4.2 WE-AX vs specific viscosity.

To test whether the specific viscosity measurement and WE-AX (pentosans) measured by the spectrophotometric method gave comparable results in predicting those lines with the highest soluble AX content, the ηsp determined for all samples in the ParW145, Parw471 and ParW694 populations in 2019 was plotted as a function of the WE-AX (2019) (Fig 14), raw data appendix, supplementary table 5). The values for WE-AX deviated substantially from the trend at two points, at 1.08 and 1.09. To determine whether this difference was reproducible, analyses of ηsp and WE-AX in biological replicate samples were carried out. This showed that although replicate 1 of line ParW694_042 had highest content of WE-AX (1.09), the content in replicate 2 was lower (0.88).

The correlation between WE-AX and η_{sp} was rather weak, R=0.45, R^2=0.2025, with little predictive power, and only 20% of the variance of WE-AX being explained by η_{sp}. This contrasts with Lovegrove *et.al.,* (2020), who reported strong correlations between η_{rel} and WE-pentosans.

This suggests that other components contributed to the viscosity of the extracts of wholegrain analysed here, that do not contribute to the absorbance detected during the determination of WE-pentosans via the spectrophotometric method, compared to white flour samples used by Lovegrove *et al.*, (2020). Technical error was accounted for as extracts made from a control flour, Yumai-34 were also included at the start, middle and end of the run to check for changes throughout a chromatography run, which was typically over 12 hr long. When the technical error exceeded 10% on the controls, the results were discarded and the samples rerun, from the frozen stock solution of flour extract.

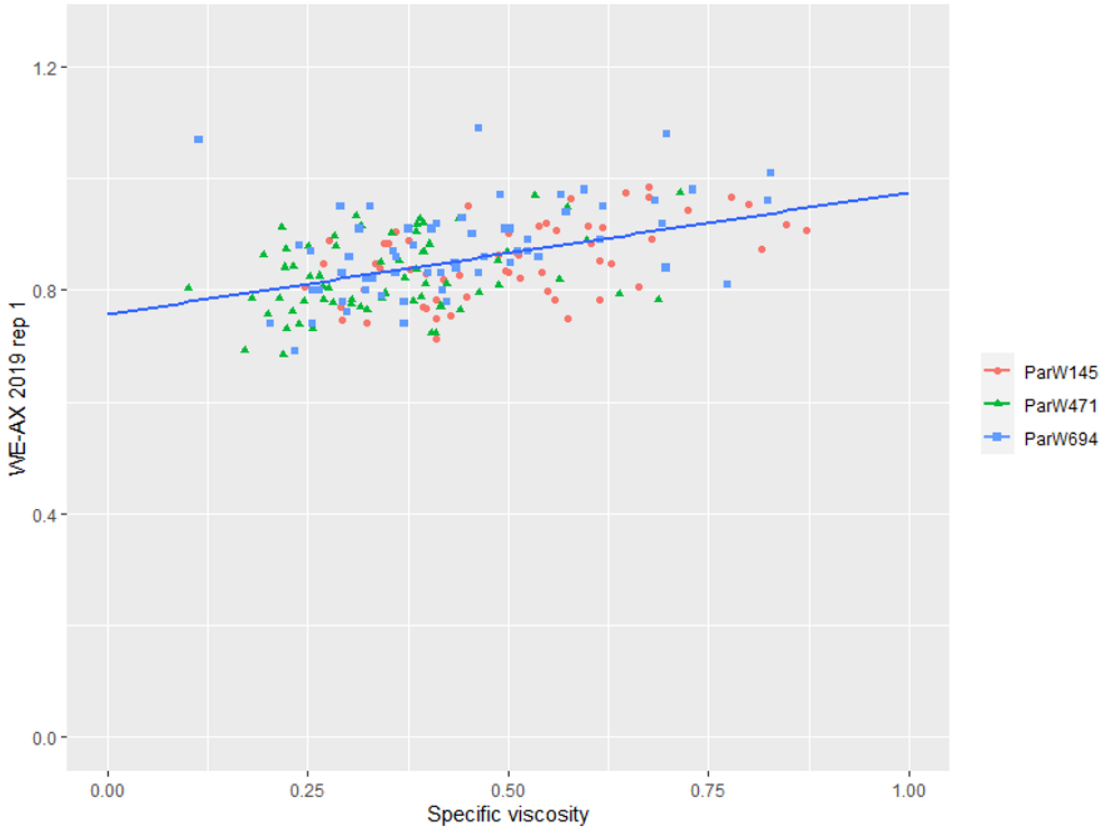


Figure 14 : Linear regression of WE-AX against ηsp for 2019. The trend line is fitted to points as a whole and not individual populations. 'ParW145 red circles; ParW471 green triangles; ParW694 blue squares populations.

The weak correlation between η_{sp} and WE-pentosans from wholemeal flour in these populations indicates that specific viscosity is less good at estimating the WE-AX content and is certainly less accurate than the pentosan assay which are a more direct measure of AX content. Therefore, to

identify QTLs relating to AX content, the samples from 2020-2021 were only phenotyped by pentosan assays.

3.5 Single-pass QTL analysis with KASP genotyping maps.

A preliminary QTL analysis of the three ParW populations for high AX content, using a low-density KASP linkage map consisting of approximately 200 markers across the genome, yielded eight putative QTLs (available in supplementary table 6) mapped to 4 different chromosomes and physical positions were anchored to Chinese Spring Refseq v1.1 (Figure 15). The only QTL that was observed in both years was on chromosome 2B, however, in 2019, it was observed for phenotypes specific viscosity and WE-AX, whilst in 2020, was identified for TOT-AX. The QTL on 1B was detected across two populations in 2019. This QTL appears identical to that described by Lovegrove et al. (2020) using Axiom 35k marker genotyping in 3 different populations from crosses of Yumai-34 with cultivars Valoris, Claire and Altigo.

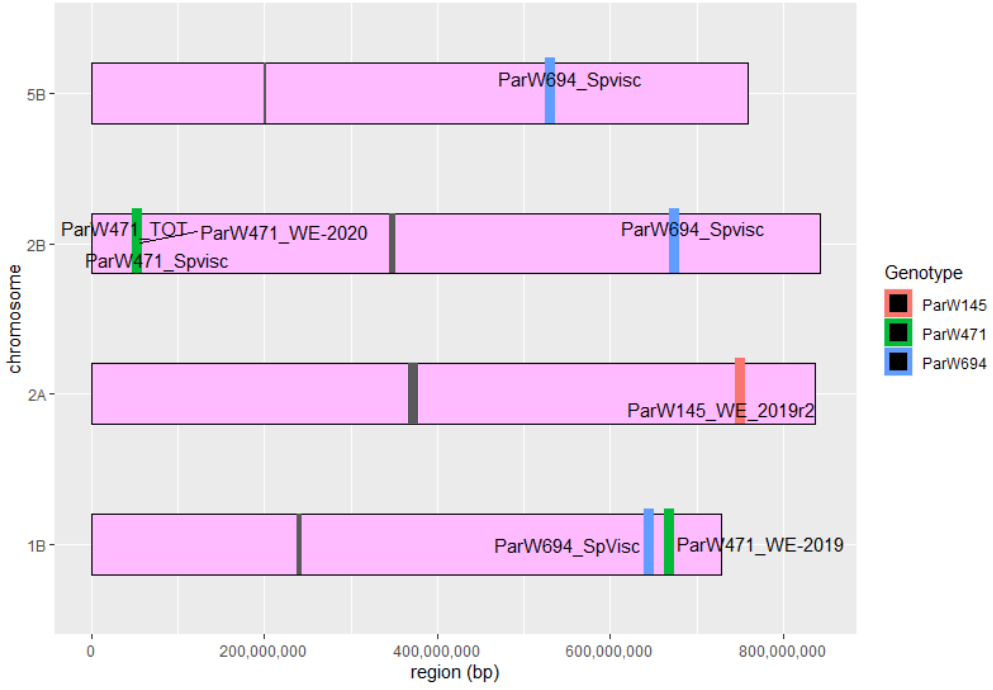


Figure 15: Schematic representation of chromosomes showing the positions of QTLs, anchored to Chinese Spring RefSeqv1.1 for WE-AX, TOT-AX and specific viscosity from 2019 material. Chromosome lengths are in base pairs, and QTL positions are the physical position on the chromosome. The physical location of the centromere is shown by the grey line on each bar.

The locations of the QTLs for WE-AX, as measured by pentosans, in 2019 are presented schematically in Figure 14 for all three of the ParW populations. The average WE-AX content across the biological replicates, within a year, were used to identify QTL.

The 1B QTL showed the highest 'logarithm of odds' (LOD) score, which is a statistical measure of confidence that two genetic loci are inherited together, or genetically linked. A LOD > 3 is typically accepted as the threshold for determining that two loci are linked. The LOD score associated with WE-AX, was determined to be 7.206 in the ParW471 population in 2019 while the corresponding QTL in the ParW694 population displayed a much lower LOD score of 2.19 at this position. Although lower in 694, the high LOD seen in ParW471 and the identification of this QTL in previous studies (Quraishi et al., 2011; Lovegrove, et al., 2020; Ibba et al., 2021) indicates that it is robust. The 2BS QTL associated with specific viscosity in ParW471 is the only 2B QTL above the cut-off (LOD>3), with a LOD of 3.13. Co-located QTLs detected for TOT-AX and WE-AX in ParW471, had LOD scores of 2.49 and 2.83, respectively. The 2BS QTL for specific viscosity accounted for 19% of the variation, and for about 14% and 16% of TOT-AX and WE-AX, respectively. The QTLs mapped on 2BL and 5B in ParW694, and 2A in ParW145 all display LODs <3.0 with the 5B QTL only accounting for 6% of the variation in specific viscosity and the 2A QTL for <17% of variation in WE-AX in ParW694.

3.6 ASMap linkage mapping

Following the initial QTL mapping using the KASP linkage maps the three RIL populations (ParW145, ParW471 and ParW694) were grown by the JIC and DNA was sent for genotyping with the Axiom 35K breeders' array at Bristol University. Linkage mapping was carried out using 'ASMap' in R and a significance level of $P=10^{-11}$ was selected for ParW694 and significance levels of $P=10^{-12}$ were selected for ParW145 and ParW471. These significance levels were chosen

based on the following criteria: the lowest number of mixed chromosomal linkage groups (MCLGs) were produced, all chromosomes were mapped, and the markers were the most evenly distributed. Mixed chromosomal linkage groups may be defined as those in which markers mapped to two chromosomes simultaneously e.g. '5D6B2B' (Figure 16). Chromosome 2B was the largest linkage group in centiMorgans(cM) in ParW694 and ParW471, and the second-largest linkage group in ParW145. The D genome had the fewest markers and chromosome 4D significantly fewer markers than the second-lowest linkage group, excluding MCLGs (7D in both ParW694 and ParW145). This was to be expected, given the low genetic diversity of this genome in hexaploid wheat, with chromosome 4D historically displaying low levels of polymorphism (Chu *et al.*, 2008; Xue *et al.*, 2008). The presence of a large 4D linkage group with few markers in ParW471 indicated the presence of large regions with high recombination frequency occurring in the gaps between markers, but also high genetic conservation in regions where polymorphic SNPs are uncommon.

Previous studies have shown that chromosome 6D generally has low numbers of markers (Xue *et al.*, 2008), and while the genetic distance was slightly greater in 6D than 7D (143.93 cM and 139.29 cM respectively), fewer markers were mapped to 6D (26 and 62 markers respectively). These observations are consistent with published data and show that the new Axiom 35k linkage maps for populations ParW694, ParW471 and ParW145 display the expected marker distribution indicating that few or no errors occurred during the production of the maps. They are therefore suitable for QTL mapping.

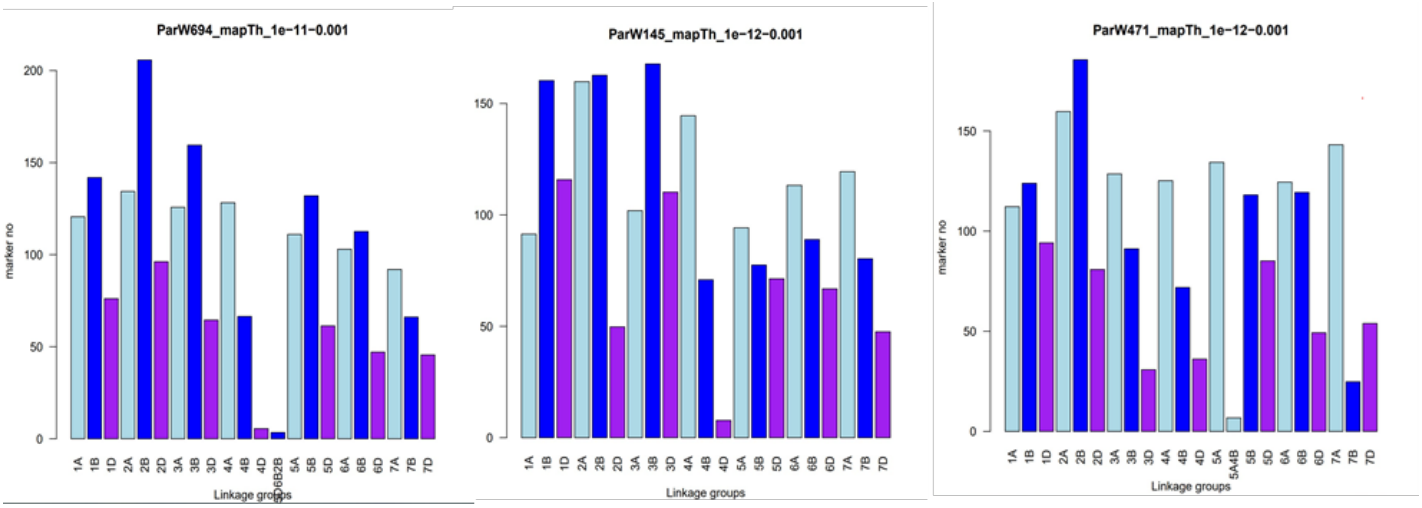


Figure 16: Histogram of the distribution of markers across chromosomes from linkage mapping of Axiom 35k genotyping of ParW145, ParW694 and ParW471. A genome = light blue, B genome = Dark blue, D genome = Pink.

3.7 QTL Detection using the Axiom 35K Breeders Genotyping Array.

Following the preliminary mapping of QTL using low-density KASP linkage maps, QTLs were mapped using the high-density linkage maps derived from the Axiom 35K breeders' array, using the mean values for pentosan content (using the spectrophometric assay) (Supplementary tables 2-4) for all the biological replicates for each year from 2019-2021. A total of 11 QTL were detected across the three populations. Ten QTL were associated with WE-AX (Table 7) and two with TOT-AX (Table 8). Of these QTL, only one was robust enough to appear in successive years. The robust QTL on the long arm of chromosome 1B was detected in the population ParW471 and was associated with WE-pentosans. This QTL corresponds to the 1B QTL detected during the preliminary QTL mapping. However, it was not detected in the population ParW694 even though high density KASP markers were used. This may be due to the 1B allele having a smaller effect in the ParW694 population compared to ParW471, resulting in it not being detected. A

second QTL for WE-pentosans was detected on chromosome 2B in both ParW471 and ParW694 in 2021.

The physical positions of the markers on the Chinese Spring genome are difficult to define for the 2B QTL, as the confidence interval for the loci align to positions at both ends of chromosome 2B and in similar physical positions in both populations. The increasing allele for this QTL is conferred by the Watkins parental lines in both instances, this could suggest a possible ancestral chromosomal rearrangement occurring in W471 and W694. However, this would be difficult to show without genome assemblies for these accessions and would require much more work to identify valid targets for breeders (which is beyond the scope of the present work).

Table 7: QTL table for mean contents of water extractable (WE) pentosans calculated per year. Each marker panel consisted of 6100 markers, the LOD threshold for a significant QTL was 3.2, but only considered as a true QTL if detected across multiple years.

Population	chr	LOD	%var	CIstart	CIend	start marker	end marker	trait	increasing Allele
ParW145	2A	5.432	27.738	119.53	125.43	AX-95096961.6B	AX-94420474.5A	WE_AX 2019	Watkins
ParW145	1A	3	11.9	40.18	59.12	AX-94731769.1A	AX-94432777.1A	WE_AX 2020	Watkins
ParW145	4B	5.5	23.5	9.2	11.37	AX-94902578.4B	AX-94598221.4A	WE_AX 2020	Paragon
ParW471	1B	10.7	42	104.55	108.2	AX-94950893.1B	AX-94519560.1B	WE_AX 2021	Paragon
ParW471	2B	3.4	10.7	36.98	60.32	AX-94882009.2D	AX-95159933.2B	WE_AX 2021	Watkins
ParW471	1B	6.908	33.148	107.8	110.561	AX-94766122.1B	AX-94840083.1B	WE_AX 2019	Paragon
ParW694	3B	2.4	11.4	47.3	52.67	AX-94728322.3B	AX-94630867.3B	WE_AX 2019	Paragon
ParW694	5B	2.8	13.9	63.41	73.04	AX-94768083.5B	AX-94537482.5B	WE_AX 2019	Paragon
ParW694	2A	3.381	20.736	125.38	130.67	AX-94496850.2A	AX-94659499.2A	WE_AX 2020	Paragon
ParW694	2B	4.506	25.039	55.65	61.68	AX-94394763.2B	AX-94516724.2B	WE_AX 2021	Watkins

One QTL detected had a LOD score below the threshold of 3.2. This was located on chromosome 1A in ParW145 and had a LOD score of 3, and two other QTL in ParW694, on chromosomes 3B and 5B had LOD scored of 2.4 and 2.8 respectively. However, it was not detected in any other years or populations. Similarly, the 3 other major QTL, where LOD scores reached above the threshold, on chromosomes 2A and 4B in ParW145, and 2A in ParW694 were not detected again in subsequent years, leaving only 1B and potentially 2B as viable targets for breeding.

Table 8: QTL table for mean contents of Total (TOT) pentosans calculated per year. The same marker panel of 6100 markers were used, and the LOD threshold for significant QTL was set at 3.2, but only considered if robust and appear across successive years.

Population	chr	LOD	%var	pos nearest marker	CIstart	CIend	start marker	end marker	trait	increasing Allele
ParW471	5B	3.805	19.891	43.76	40.37	45.36	BS00023161.5B	BS00084528.5B	TOT_AX 2021	Watkins
ParW694	1A	4.041	22.775	82.76	57	86.56	AX-94480220.1A	AX-94541870.1A	TOT_AX 2021	Watkins

3.8 Discussion.

QTL have been identified in 3 mapping populations of recombinant inbred lines (RILs). These three populations underwent 2 stages of QTL mapping, starting with a preliminary QTL detection using low-density KASP linkage maps and datasets for WE-pentosans, TOT-pentosans, and specific viscosity from two seasons. More precise QTL mapping was then performed using high-density Axiom 35K linkage maps using pentosan data from three growing seasons. The preliminary mapping identified a QTL on 1B for WE-AX in a population of ParW471 grown in 2019 and ParW694 grown in 2019, but not 2020. This QTL is near the KASP marker BS00035268 in ParW471 which has been assigned the Axiom BA code BA00542264. This SNP occurs at a physical distance of 652Mpb on the long arm of chromosome 1B. This QTL occurs in the region between 626-668Mbp and appears to account for 24% of the variation in ParW471 and 14% in ParW694. The physical location of the 1B QTL is close to the physical locations of QTLs described previously (Lovegrove *et al*., 2020; Ibba *et al*., 2021, at 656-668Mbp and 646-654Mbp, respectively) indicating that the three QTLs are probably the same. Furthermore, in agreement with the previous studies, the 1B QTL accounts for 14-24% of the variation. However, the 1B QTL was not detected in the ParW694 population using the high-density Axiom linkage maps, it was calculated that the 1B QTL accounted for 33-44% of the total variation in WE-pentosans in ParW471, which is higher than the values from the preliminary QTL screening and that reported

by Lovegrove *et al., (2020)* and Ibba *et al., (2021)*. It is probable that the expression of the QTL was affected by both the genetic background and environmental conditions for the field trials, which differed between the two studies.

Epistatic interactions between QTLs within the ParW471 population could have also contributed to the greater effect of the 1B QTL on WE-AX, however, analysis of epistasis during the QTL detection yielded no significant epistatic interactions between any loci in any of the populations (data not shown).

To the best of my knowledge, the putative 2BS QTL identified in the preliminary mapping of ParW471, is novel. Although Lovegrove *et al*., (2020) also detected a QTL on 2BS in the population of Yumai-34 x Claire, the location of this (at 249Mbp) indicates that it is not the same as the QTL reported here (at 52Mbp). However, the QTL was not reproducibly detected in successive runs of the QTL detection pipeline, despite running the detection methodology with identical parameters and input data. This would suggest uncertainty in predicting the QTL at this locus with low density linkage maps. This may be due to incorrect reporting of QTL from the software, which may identify 'ghost' QTL that are not real alleles influencing the trait. Furthermore, both parental lines conferred increasing alleles for related traits and, in agreement with the ghost QTL hypothesis, the QTL was not detected in either 2019, 2020 or 2021 field trial years using the high-density linkage maps. However, a second QTL was detected on 2B for WE-pentosans, in ParW471 and ParW694. This may have resulted from a chromosomal rearrangement as markers within the QTL mapped within 9-20 cM but aligned to physical locations of 10-60Mbp and showed similar alignments from both populations. If chromosomal rearrangements are present and overlap with a genuine QTL for AX, this may account for the difficulty in QTL detection using a low-density linkage map that does not cover the locus.

QTL for other crop species have been identified in relation to AX content. In rye, a majority of QTL have been identified from hybrid material between wheat and rye, rather than pure rye

cultivars (Piro, Muylle and Haesaert, 2023).One study, from over 20 years ago, was amongst the first to look at the entire Rye genome, using disomic addition lines between wheat and rye, D genome substitution triticale lines and homozygous wheat translocation lines (Boros *et al.*, 2002). From these lines, it was determined that AX-related QTL lie on the 2R, 5R and 6R chromosomes. A further study, looking at transcriptomic data during grain milk, dough and full kernal stages was also assessed, and candidate genes were identified on rye chromosomes (Kozlova *et al.*, 2022).

The work by Kozlova et al., determined genes related to the IRX9, IRX10, IRX14, XAT1, and BAHD01 genes that were previously known be directly involved in AX synthesis and discussed in section 1.11.

Additionally, a recent genome-wide association study (GWAS) in barley identified 3 loci in the barley genome that are associated with AX content (Hassan *et al.*, 2017). From this, two loci on barley chromosome 2H, and one on 3H were identified, which passed the FDR test of significance. Hassan *et al.,* reported several putative candidate genes that underly these QTL, including those from the GT47, GT61, glycosyl hydrolases (GH) and GT31 gene families, along with two families that contain domains of unknown function (DUF)- DUF 231 and DUF579. The work in both barley and rye, reaffirm our understanding of the role of GT and GH gene families, in AX synthesis, and highlighted two protein domains which may be related to AX biosynthesis. As such, it is important to monitor the presence of these DUF genes in QTL detected in wheat.

Chapter 4
Characterising the 1B QTL for higher AX content in the Pangenome

4.1 Introduction to the Wheat10+ Pangenome Panel.

Chromosome-level assemblies for hexaploid wheat (Chinese Spring RefSeq 1.0) have been available since 2017 (IWGSC, 2018.; Zimin et al. 2017) and have been invaluable for wheat improvement studies. However, these assemblies do not capture the full structural and genetic variation across many wheat genotypes (estimated between 25,000 and 50,000) that can be exploited for crop improvement. For example, in rice, which has a much smaller genome than hexaploid wheat, several studies have shown that the reference assembly 'Nipponbare' lacks key agronomically important genes that were only detected with annotations from the rice pangenome project (McAinsh 2009; Xu et al. 2011; Gamuyao et al. 2012). With this in mind, previous research using the Chinese Spring RefSeq 1.0 for wheat may have missed genes for traits of interest due to their absence from this reference sequence.

More recently, 9 cultivars of hexaploid wheat have been sequenced and full chromosome-level annotated assemblies produced and released, in 2019, using the same 'DeNovoMAGIC v.3.0' assembly pipeline used for Chinese Spring RefSeq1.0 (Walkowiak et al. 2020). These cultivars were collected from various countries across the globe and aim to represent the total allelic diversity across wheat cultivars. Most cultivars in the pangenome panel represent modern cultivars, or Green Revolution cultivars, for example, Norin 61 (Table 9). As such, these cultivars may only represent the total variation seen across modern cultivars and not necessarily represent the wider diversity seen in landrace collections, such as the Watkins collection.

Table 9: The Pangenome cultivars used in the presence/absence analysis in chapter 4, detailing their country of origin and their typical usage and type of wheat they represent.

Cultivar	Country of Origin	Type / Notable Features
Julius	Germany	Modern elite European cultivar
Landmark	Australia	Modern elite cultivar representing Australian breeding
Norin 61	Japan	Source of *Rht1* semi-dwarf gene; key Green Revolution cultivar
Stanley	USA	Hard red winter wheat; represents North American breeding
Mace	Australia	Widely grown Australian cultivar; adapted to dry environments
ArinaLrFor	Switzerland	Disease-resistant winter wheat; donor of *Lr* leaf rust resistance gene
SY Mattis	UK / Germany	Modern high-yielding cultivar; European winter wheat lineage
Lancer	Australia	Modern cultivar adapted to eastern Australia; used for yield and drought tolerance traits

These annotations can therefore be used in conjunction with the newer IWGSC reference genome 2.1 to capture the diversity and allelic variation (including presence or absence) (PAV) across loci of interest. This can prove useful in QTL studies where stable QTL have been detected, but no obvious candidate genes identified. Previous exploration of the 1B QTL, following the initial mapping in the cv. Yumai-34 (Lovegrove et al., 2020) failed to identify the candidate gene by using bulked RNA-seq data from lines with the high and low AX alleles (R. Mitchell, personal communication; T. Pellny *et al., unpublished)*. The sequenced samples consisted of 4 bulked DNA samples, two of which had the high AX Yumai-34 1B allele and two of the low AX 1B allele from the cv. Valoris. This analysis did not yield any candidate genes, and it was suggested that Chinese Spring may have exhibited a translocation or deletion of the underlying gene.

The genome sequences of the parental lines (Yumai-34 and Valoris) and Watkins line 471 were not available, with only fragmented and unassembled genome data for Paragon. It was therefore decided to explore the gene content of the region within the 1B QTL in the wheat pangenome, including PAVs, to identify variants which could relate to the allelic variation. To do this a novel bioinformatics pipeline was developed, utilising the newly annotated wheat10+ Pangenome assemblies, and the newly released IWGSC Refseq 2.1 assembly. This pipeline was

designed to identify candidate genes that are absent from the annotation of Chinese Spring and to characterise any structural variation in the genomic sequences in the region of the 1B QTL.

4.2 Results

4.2.1 Determining the physical location of 1B QTL on Chinese Spring v1.1

The 1B QTL detected in the ParW471 population grown in field trials in both 2019 and 2021 had the highest statistical significance. The peak LOD score for the QTL was at 109cM (Figure 17) with the confidence interval between 108 and 110 cM (Figure 18), between the markers AX-94766122 (Mx) and AX-94840083 (Mz). Mx and Mz are custom abbreviations for the marker at the start of the QTL (Mx) and the end of the QTL (Mz). These markers were therefore used to identify the physical regions of the QTL in all sequenced cultivars, including the Chinese spring RefSeq v1.1 and v2.1 and the Pangenome cultivars.

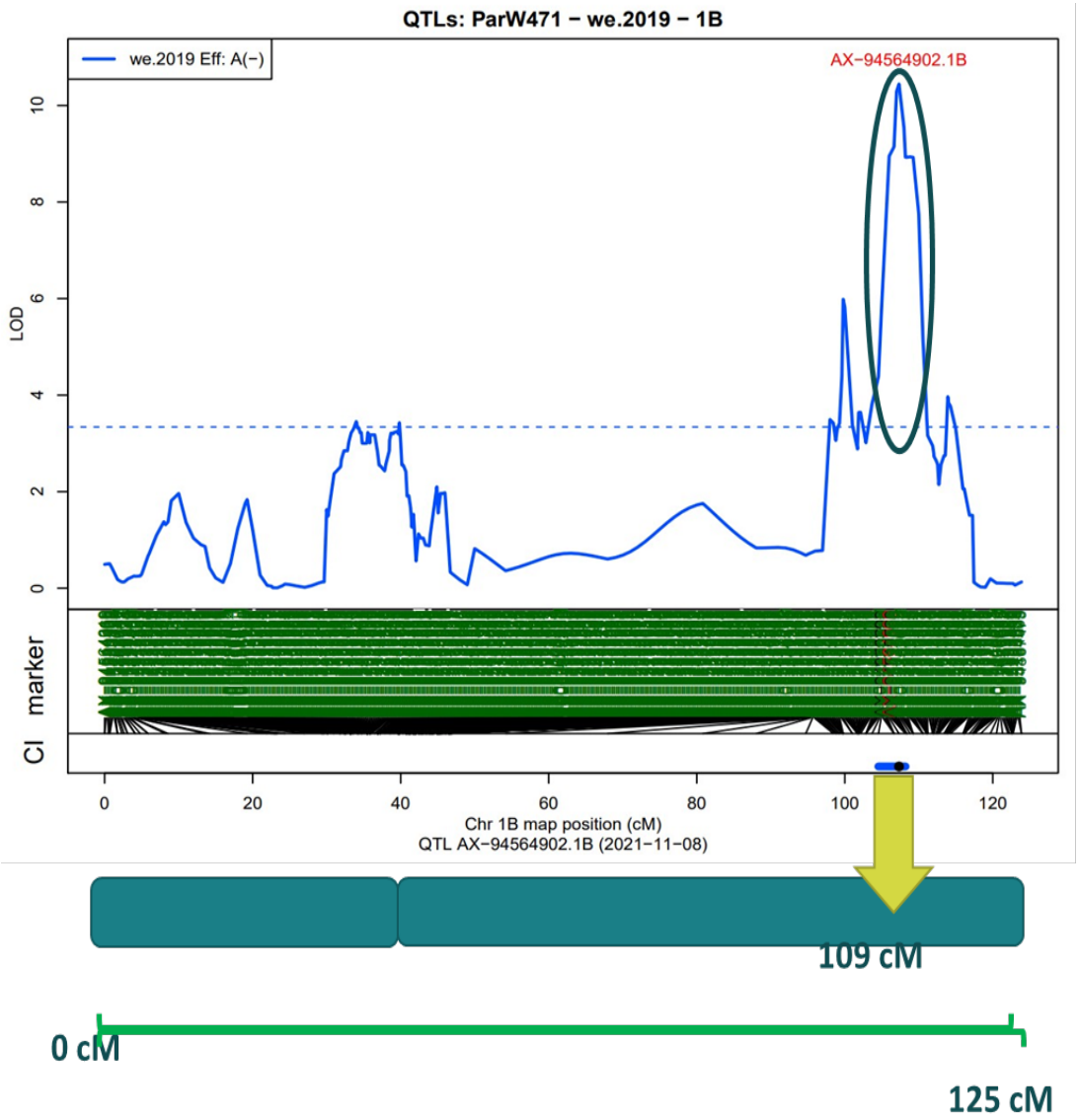


Figure 17: LOD score plot along chromosome 1B of the ParW471 population with a schematic depiction of the chromosome and the relative position of the QTL circled in dark blue

In the Refseqv1.1 version of Chinese spring, the QTL begins at 64.3Mbp and ends at 65.6Mbp. The QTL accounted for between 33 and 42% of the variation in the content of WE-pentosans in the sample sets from the 2019 and 2021 field trials and, had the greatest impact on WE-AX content in ParW471. The increasing allele was found in Paragon. The ParW471 1B QTL overlaps with the QTL found in Yumai-34 x Valoris and Yumai-34 x Altigo by Lovegrove *et al.* (2020) and it is therefore likely that Yumai-34 and Paragon share the same allele at this QTL, which contributes to increased WE and TOT AX content. The 1B QTL identified in this study, in contrast

to Lovegrove *et al*., (2020), only detects an effect on WE-AX, though this may be due to the use of wholegrain flour in the ParW crosses, and the white flour that was assayed by Lovegrove *et al.*

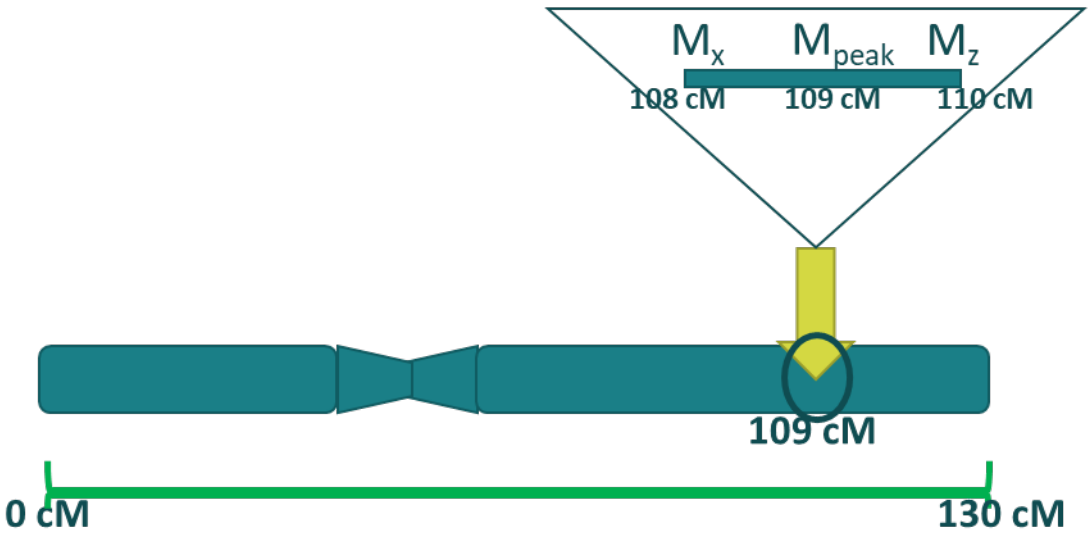


Figure 18: Schematic of the 1B chromosome *and the location of the QTL associated with WE-AX, with its flanking markers (Mx and Mz) shown at their respective position in cM.*

4.2.2 Characterising QTL sizes across the Pangenome

The size of the 1B QTL was determined for each of the cultivars that made up the pangenome as well as IWGSC 1.1 and 2.1 assemblies and is displayed in Figure 19. The QTL was longest in Norin 61 (Norin) and the two IWGSC assemblies, which were about 11-12 Mbp. The QTL lengths in the other seven pangenome cultivars ranged from ~9Mbp to 10.1Mbp.

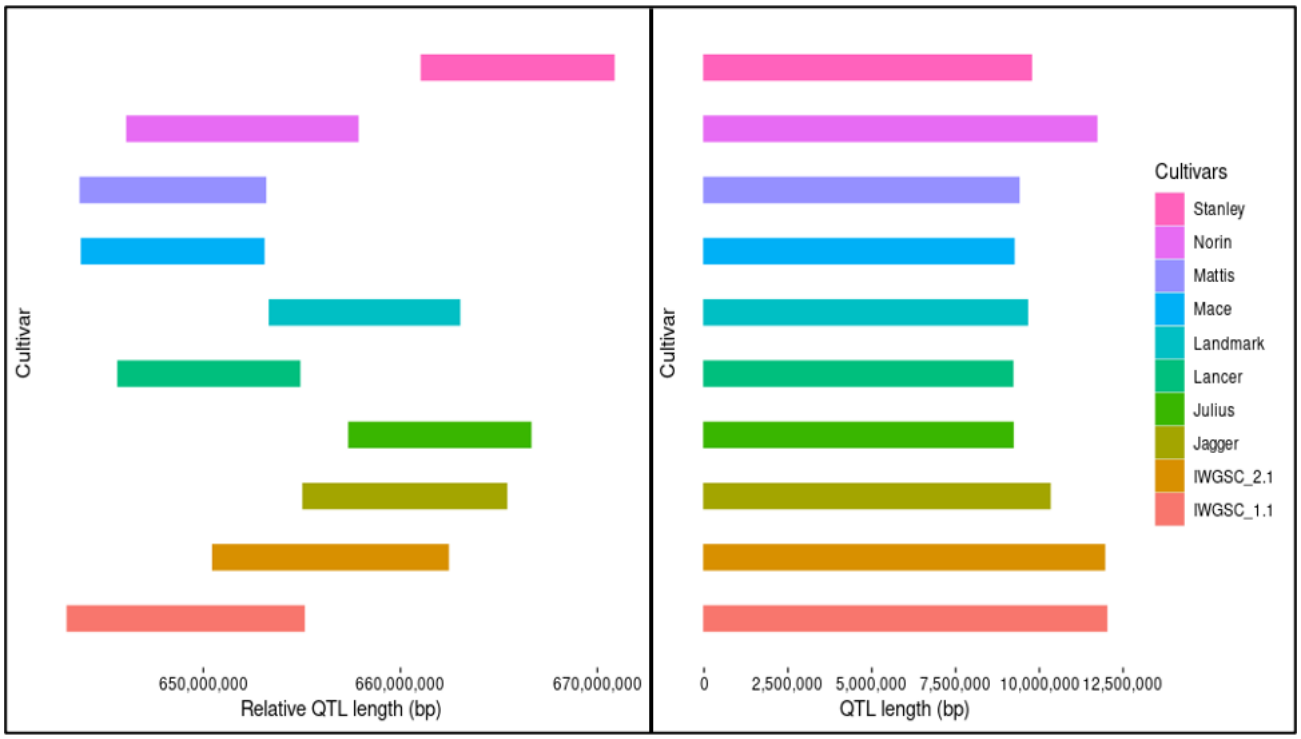


Figure 19: Lengths of the 1B QTL in the pangenomes and the IWGSC 1.1 and 2.1 assemblies. The lefthand panel shows the relative position of the QTL compared to the rest of the chromosome and the righthand panel the absolute QTL lengths.

Variation in gene density across the 1B QTL was observed, with the fewest annotated genes in IWGSC RefSeq 1.1 (only 108 genes). This number was increased to 113 protein coding genes in IWGSC RefSeq 2.1. and the highest gene density was observed in Jagger, which had 155 annotated, protein-coding genes across the 1B QTL.

4.2.3 Detecting Presence/Absence Variation in Genes Across the Pangenome

The available protein sequences were extracted from the annotation files at the Ensembl plants (Pangenomes) and NCBI (IWGSC assemblies) databases, and hierarchical orthologous groups (HOGs) were determined pairwise between the assemblies to show links between orthologous proteins in the cultivars. All of the genes across the panel were found in at least two cultivars, suggesting that the panel had accurately captured the diversity across this region (Figure 20)

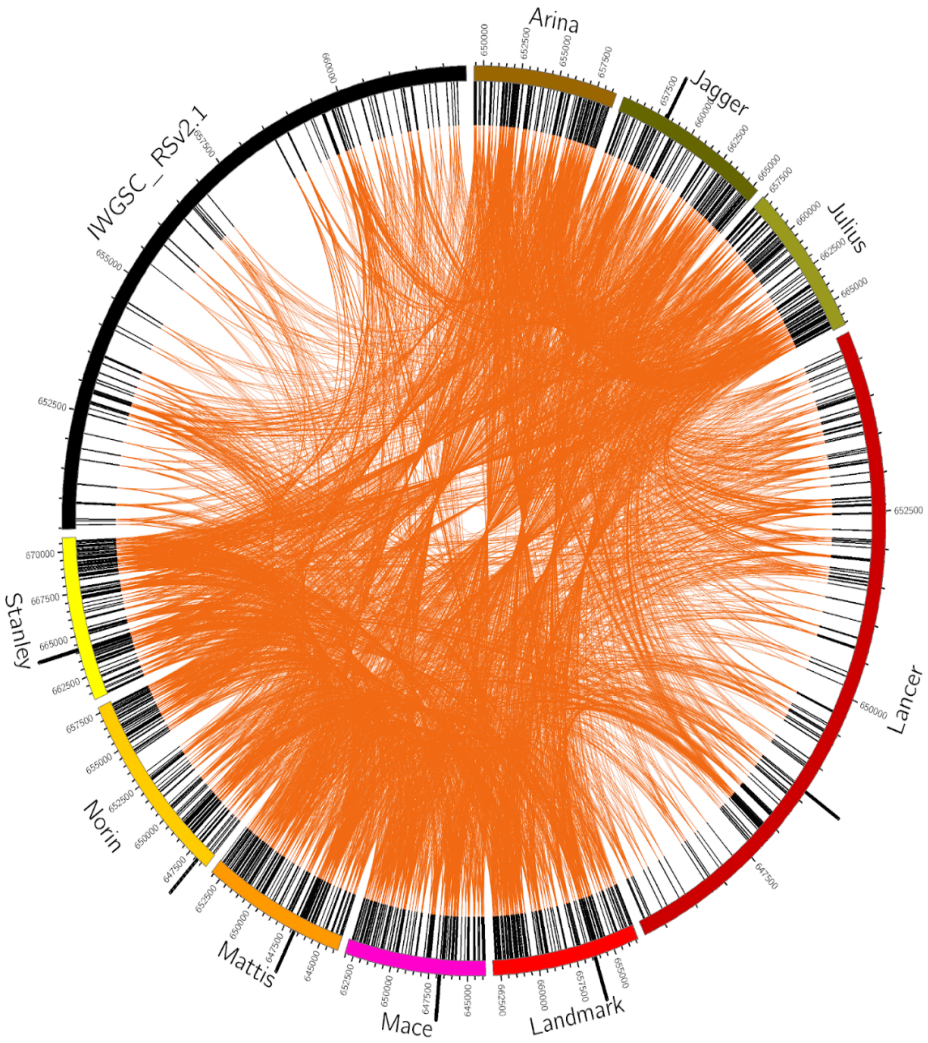


Figure 20: Circos plot indicating all hierarchical orthologous groups (HOGs) between the pangenome cultivars and IWGSC 2.1 across the 1B QTL. Genes are depicted by black ticks on the inside track, connected via links to orthologous genes (orange connections). Colours around edge to differentiate between the QTL boundary across the different cultivars.

Many of the genes that were annotated in the pangenomes were absent from Chinese Spring, indicating that the 1B QTL region may be diverse and candidate genes may not be detected when using Chinese Spring as the only mapping reference. Furthermore, annotation errors may have occurred, giving false annotations across the pangenomes. To explore this possibility, pairwise alignments of the 1B QTL region across the pangenomes and IWGSC were performed to visualise structural variation within the panel (Figures 20-26). These alignments showed IWGSC 2.1 and Norin61 to be most similar, Figure 21 indicating an almost identical alignment. In particular, both cultivars show a large insertion of non-coding sequence when aligned with all the wheat10+ genomes except for Norin 61 (See Figure 20, right).

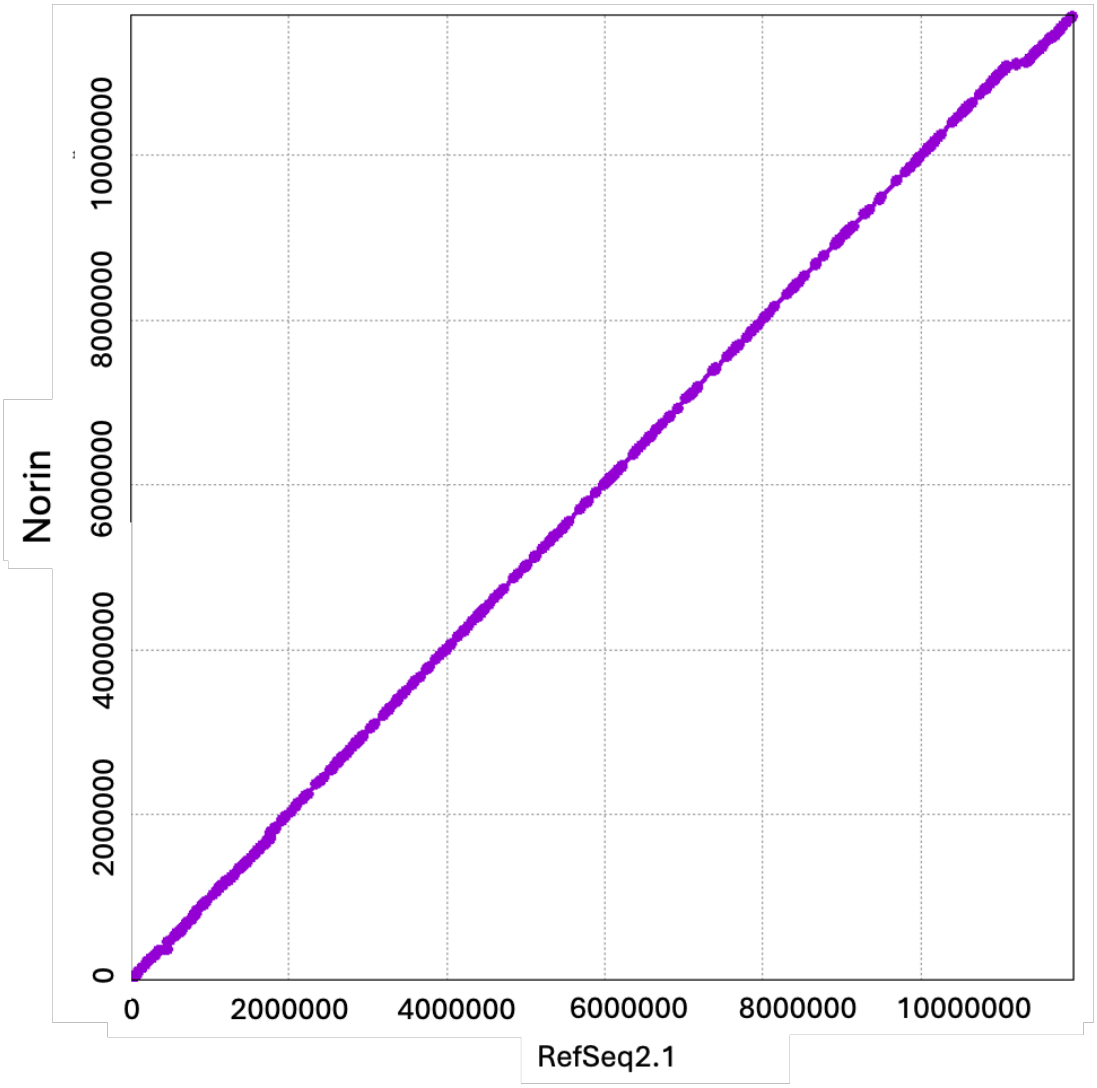


Figure 21: NUCmer alignment of the 1B QTL region between Norin 61 (Y axis) and IWGSC v 2.1 (X axis). Axes starts at 0 base pairs and denotes the absolute length of the QTL in Norin and RefSeq2.1

The alignment of the sequence of the 1B QTL region between IWGSC RefSeq v2.1 and Mace shows (Figure 22) a region with similar structure with no major inversions or insertions. Two small insertions are present at about 3,100,000bp (Figure 22, labelled A) and 10,100,000bp (Figure 22, labelled D) and a larger insertion at 7,000,000bp (Figure 22, labelled B) in Chinese Spring. The region upstream from this insertion has the lowest similarity to Mace (Figure 22, labelled C) and may contain genes that are absent from Chinese Spring.

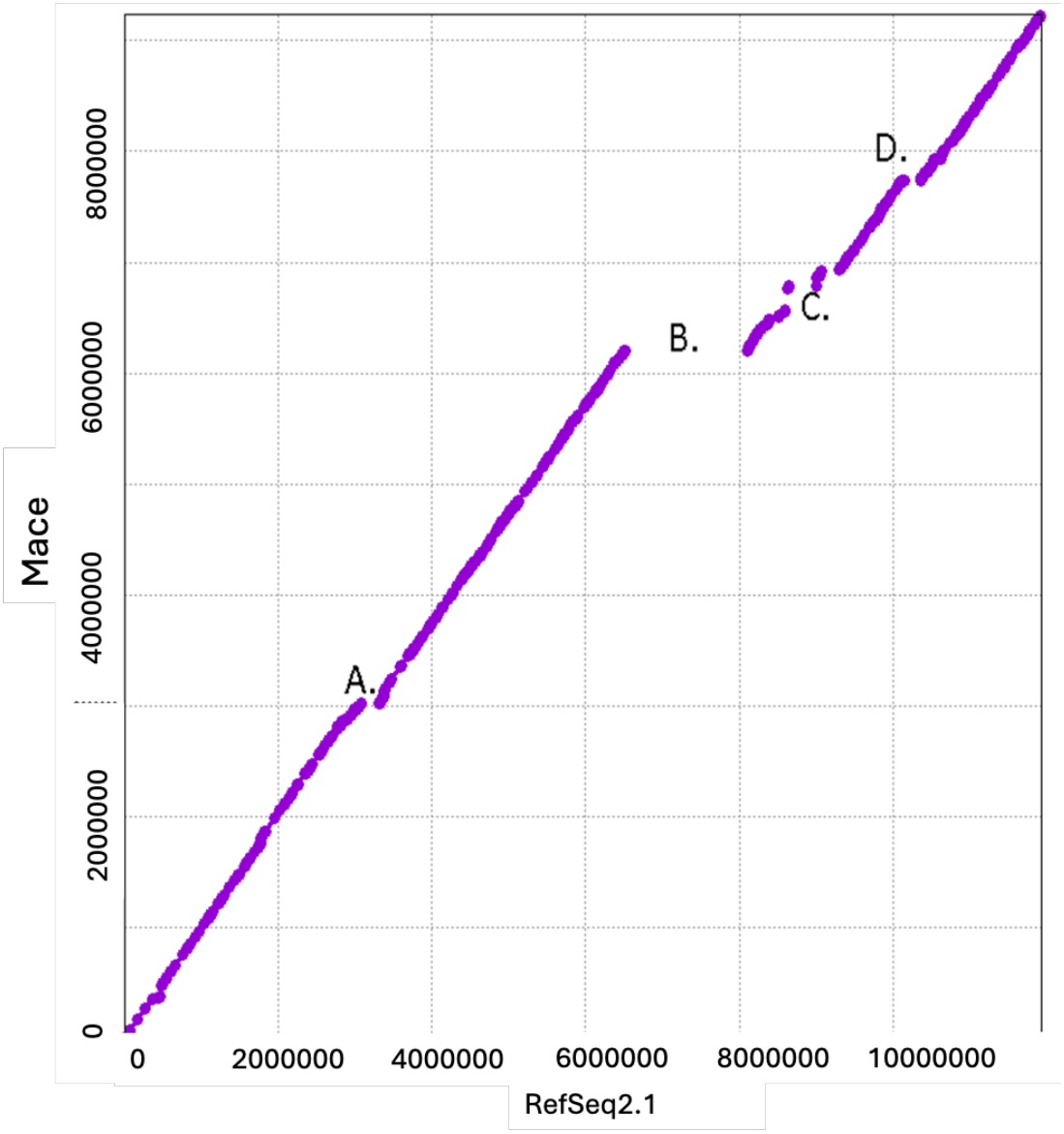


Figure 22: NUCmer alignment of the 1B QTL region between Mace (Y axis) and IWGSC v 2.1 (X axis). The axis starts at 0bp and denotes the absolute length of the QTL and not the genomic positioning on chromosome 1B.

The 1B QTL region of Mattis is similar to that of Mace and displays the same deletions, showing the same characteristic insertions as Chinese Spring (Figure 22, labelled A, B and D). Mattis also displays a similar region of divergence at point labelled C on Fig 23. This region, like that in Mace, displays insertions of DNA in both Mattis and Chinese Spring, identified by the upwards shift in plot points denoting insertions in Mattis, and a shift to the right denoting the insertion in Chinese Spring.

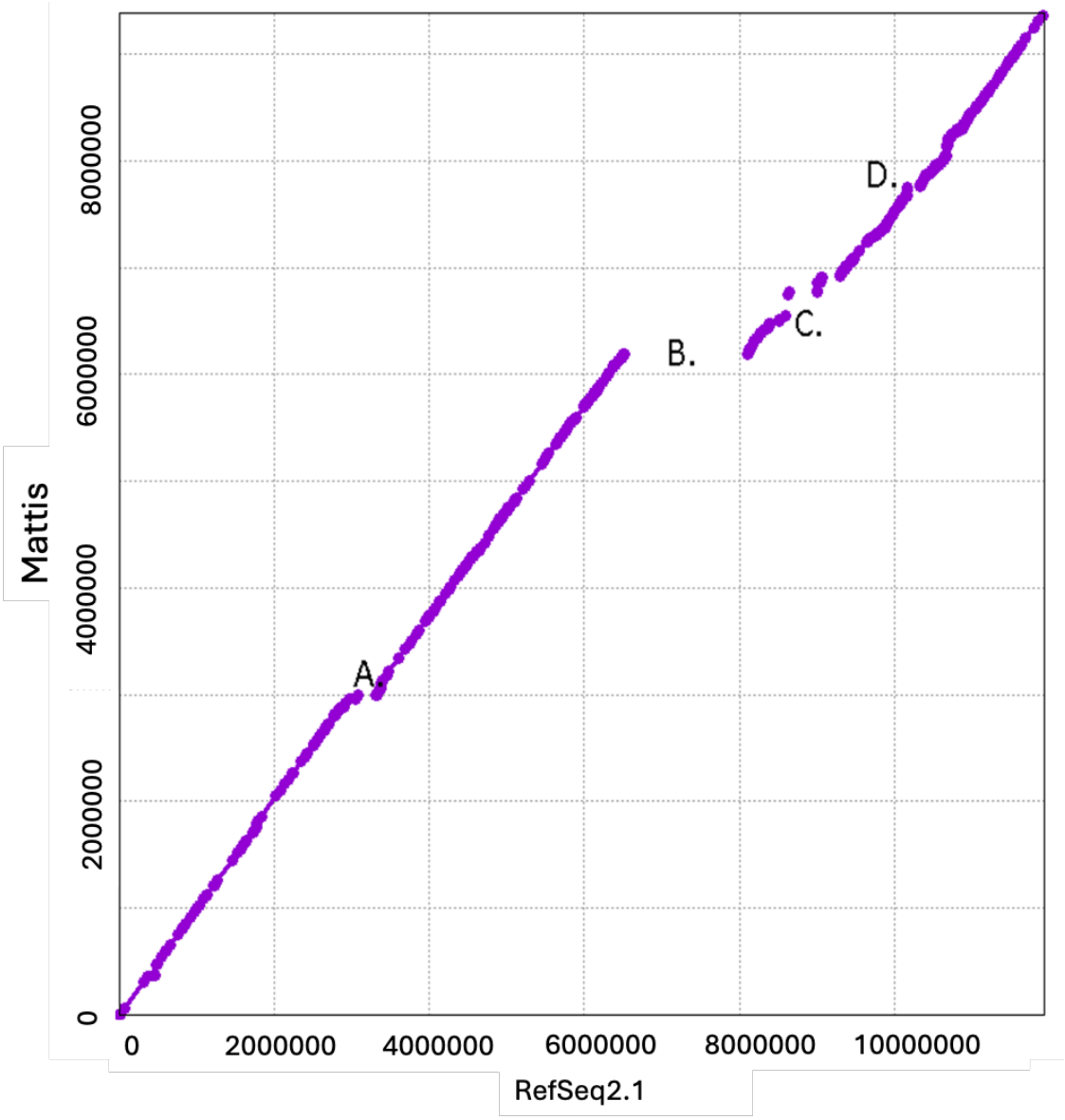


Figure 23: NUCmer alignment of the 1B QTL region between Mattis (Y axis) and IWGSC v 2.1 (X axis). Axis starts at 0bp and denotes the absolute length of the QTL and not genomic positioning on chromosome 1B.

The dot plot alignments of Chinese Spring against Jagger showed very similar insertions in Chinese Spring as shown in Figure 24, labelled A and C. However, Jagger contains a unique insertion that is not detected in the other Pangenome alignments, which is displayed in Figure 24, labelled B. Moreover, unlike the previously described pangenome cultivars, the region at point D, -analogous to point C in Figures 21-23, shows only a single insertion in Chinese Spring and a much smaller insertion in Jagger, immediately before that in Chinese Spring. This indicates a higher degree of similarity between Jagger and Chinese spring at this region. However, it is

worth noting that, the upstream regions of this QTL showed more insertions in Jagger that are not present in Chinese Spring. Therefore, it may be that this region contains genes that are absent from the Chinese Spring reference genome sequence. Despite this, any genes that are present in these regions are not unique to Jagger, as no orphan gene models (i.e. genes that do not have homologous annotations in the other pangenome cultivars) were detected, as can be seen from the Circos plot (Figure 28).

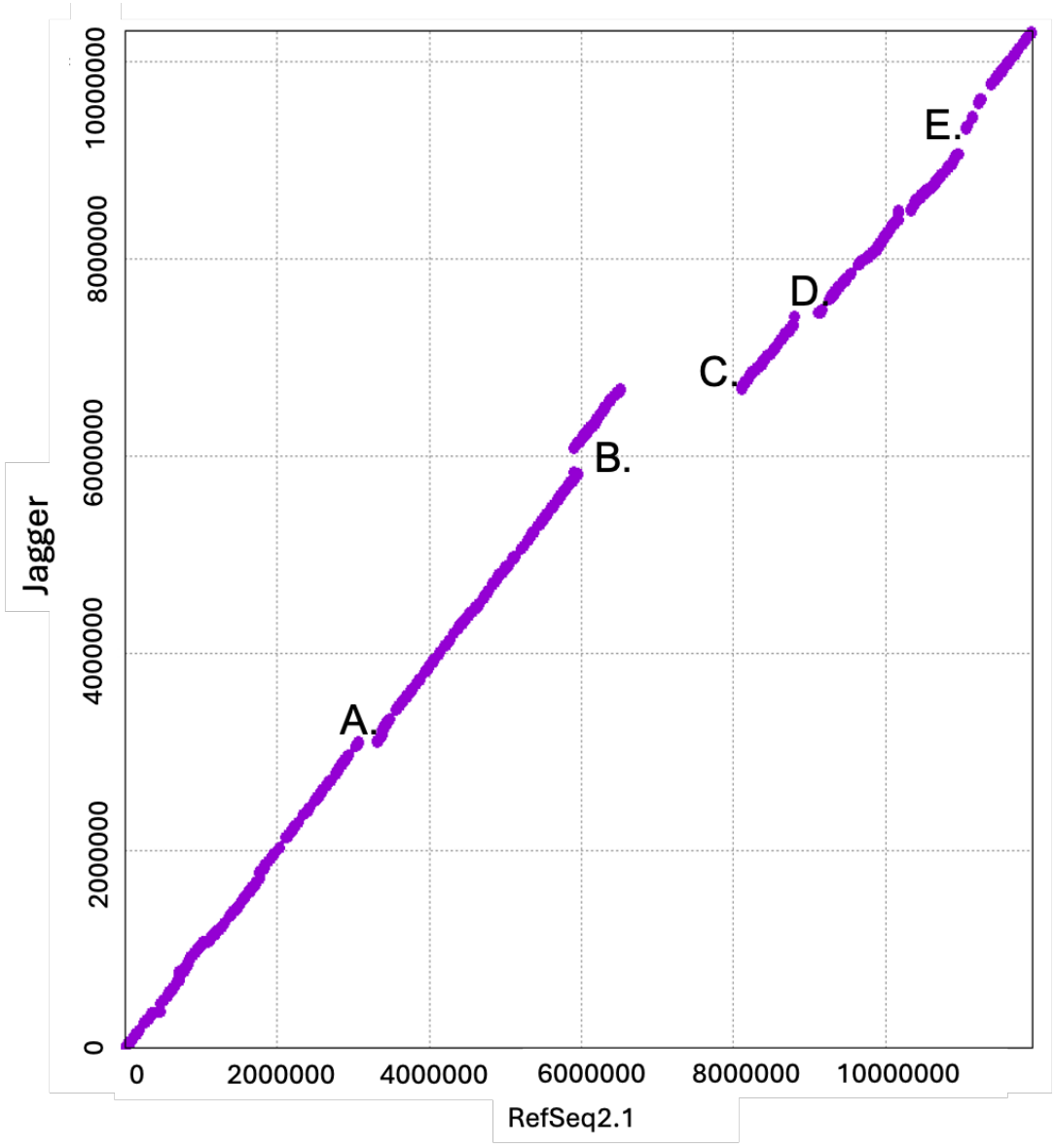


Figure 24: NUCmer alignment of the 1B QTL region between Jagger (Y axis) and IWGSC v 2.1 (X axis). The axis starts at 0bp and denotes the absolute length of the QTL and not the genomic positioning on chromosome 1B.

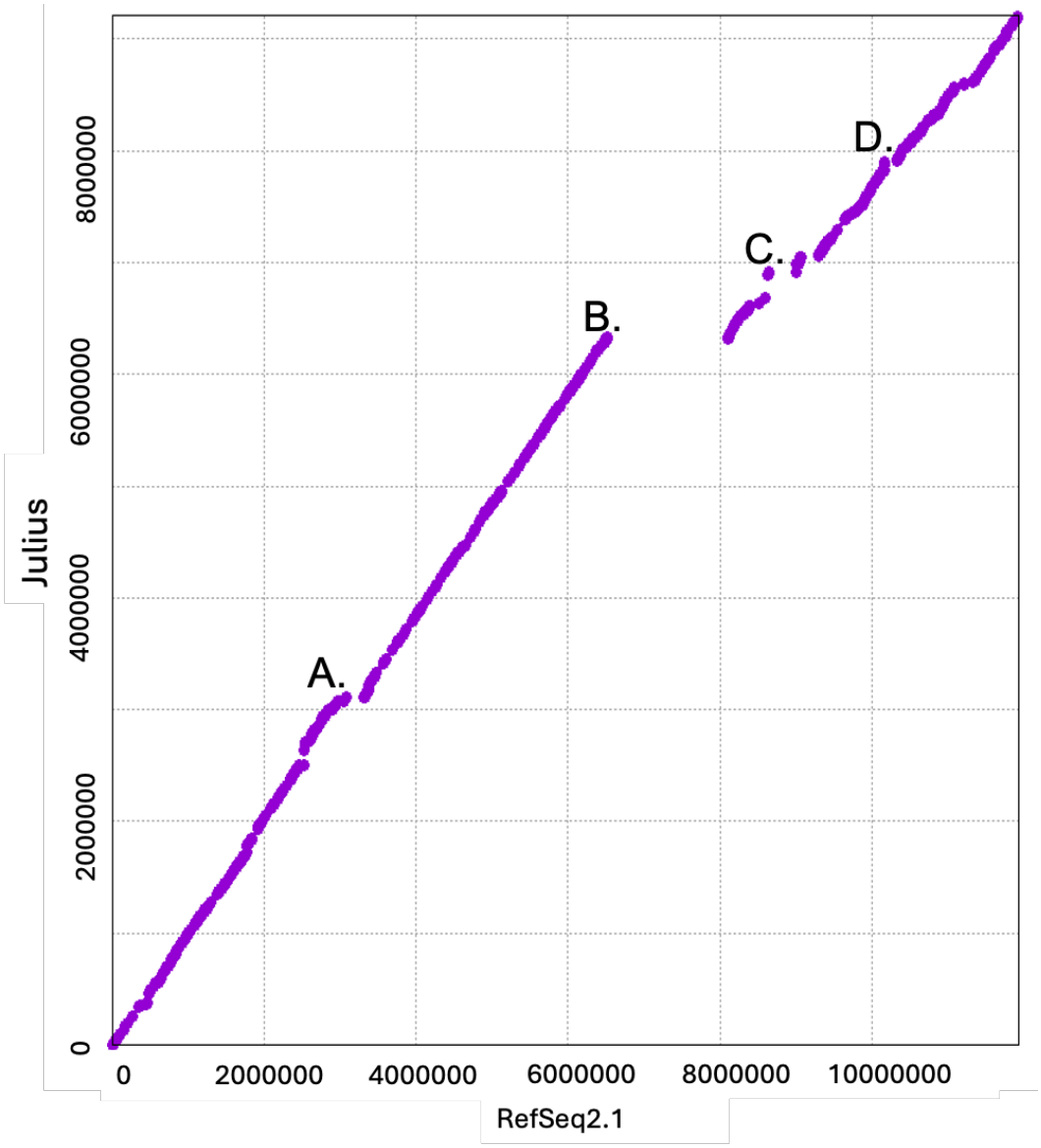


Figure 25: NUCmer alignment of the 1B QTL region between Julius (Y axis) and IWGSC v 2.1 (X axis). The axis starts at 0bp and denotes the absolute length of the QTL and not genomic positioning on chromosome 1B.

Similarly, the alignments of both Julius (Figure 25) and Landmark (Figure 26) show similar insertions and deletions, at positions A-D, to the other pangenome cultivars, suggesting that they are more representative of modern cultivars of wheat (that make up the pangenome) than Chinese Spring.

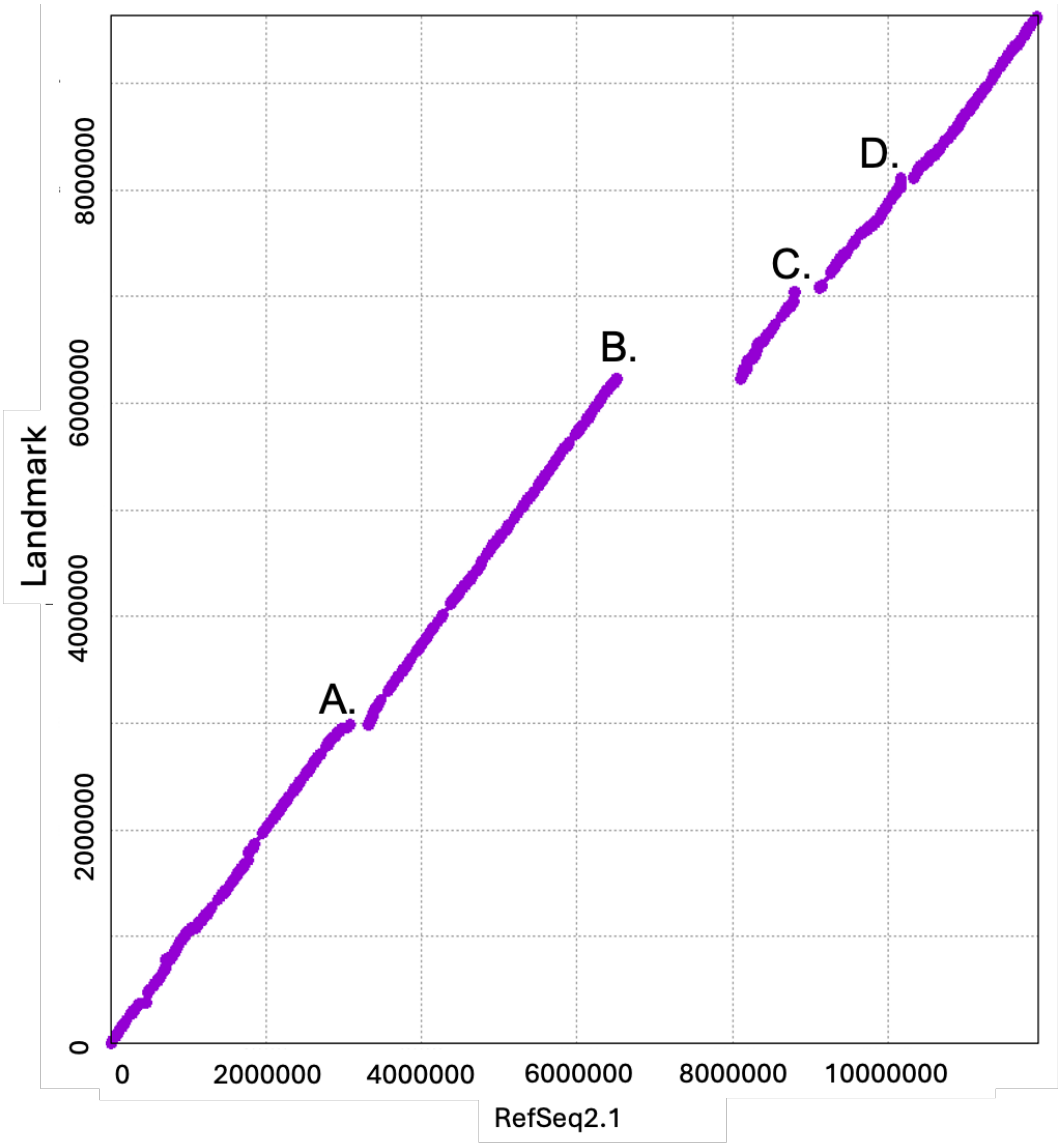


Figure 26: NUCmer alignment of the 1B QTL region between Landmark (Y axis) and IWGSC v 2.1 (X axis). (The axis starts at 0bp and denotes the absolute length of the QTL and not genomic positioning on chromosome 1B.

Comparison of the nucleotide sequences in Refseq V2.1 against Lancer (Figure 27) shows no alignment at ~7,000,000bp in Lancer, (Figure 27- B.) and some deletions in both Lancer and Chinese Spring at ~8,300,000bp (Figure 27- C), which coincides with a block of genes at the 65.2Gbp region of the 1B QTL in Lancer. The insertions in Refseq2.1 at position A and D (Figure 27) fall at the same positions as the with the alignments with the other pangenome cultivars.

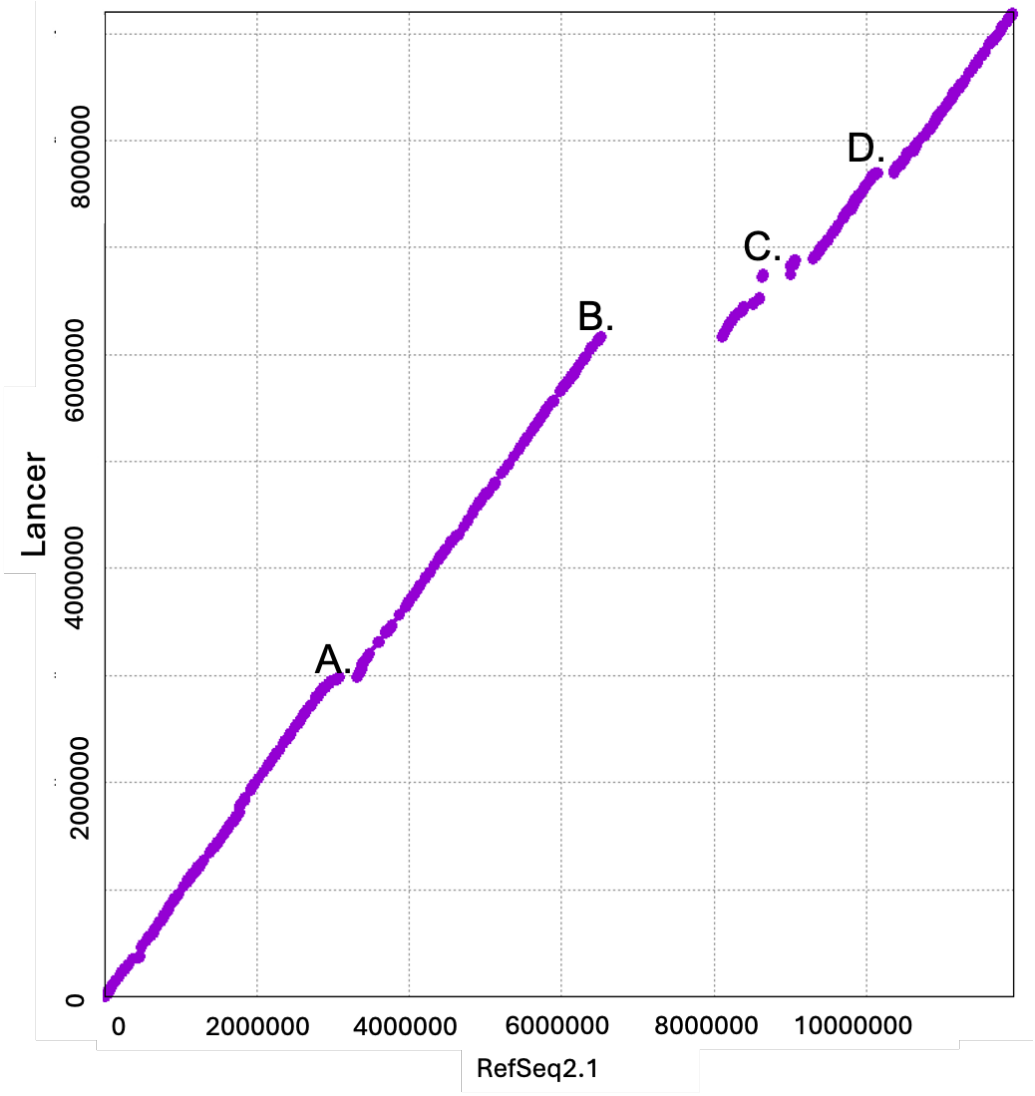


Figure 27: NUCmer alignment of the 1B QTL region between Lancer (X axis) and IWGSC v 2.1 (Y axis). Axis starts at 0bp and denotes the absolute length of the QTL and not genomic positioning on chromosome 1B.

This block includes 3 genes that were not detected in IWGSC 2.1 (see Figure 28b) in the region between 65.1Mbp-65.22Mbp. Alignment of endosperm-specific RNA-seq reads (Pellny *et al., unpublished)*, using HISAT2, from RNA-seq of Yumai-34 (with the high AX allele at 1B QTL) and Valoris (with the low allele at 1B QTL) to Lancer indicated that one of these genes is expressed in the developing endosperm in both high and low AX alleles. It is therefore necessary to analyse this gene for SNPs and differences in expression levels between allelic variants before it can be considered a candidate.

Yumai x Valoris recombinants have previously been shown to have a 1B QTL in the same region as that detected in the ParW crosses in a separate QTL analysis, where Yumai contains the increasing allele (Lovegrove *et al.*, 2020), suggesting that Paragon and Yumai-34 share the same high fibre allele at this locus. Similarly, the Lancer and Chinese Spring v2.1 sequences can only be partially aligned at positions 34,000,00bp-36,000,00bp (Figure 28a) with some gaps in Lancer. This region coincides with the physical location of the 648 - 649 Mbp region of Lancer, where there are ten genes of which only two are present in Chinese Spring. Therefore, the eight remaining genes may also be present in Paragon, that would not have been detected if only using f Refseq2.1 as the reference genome for putative candidate gene detection.

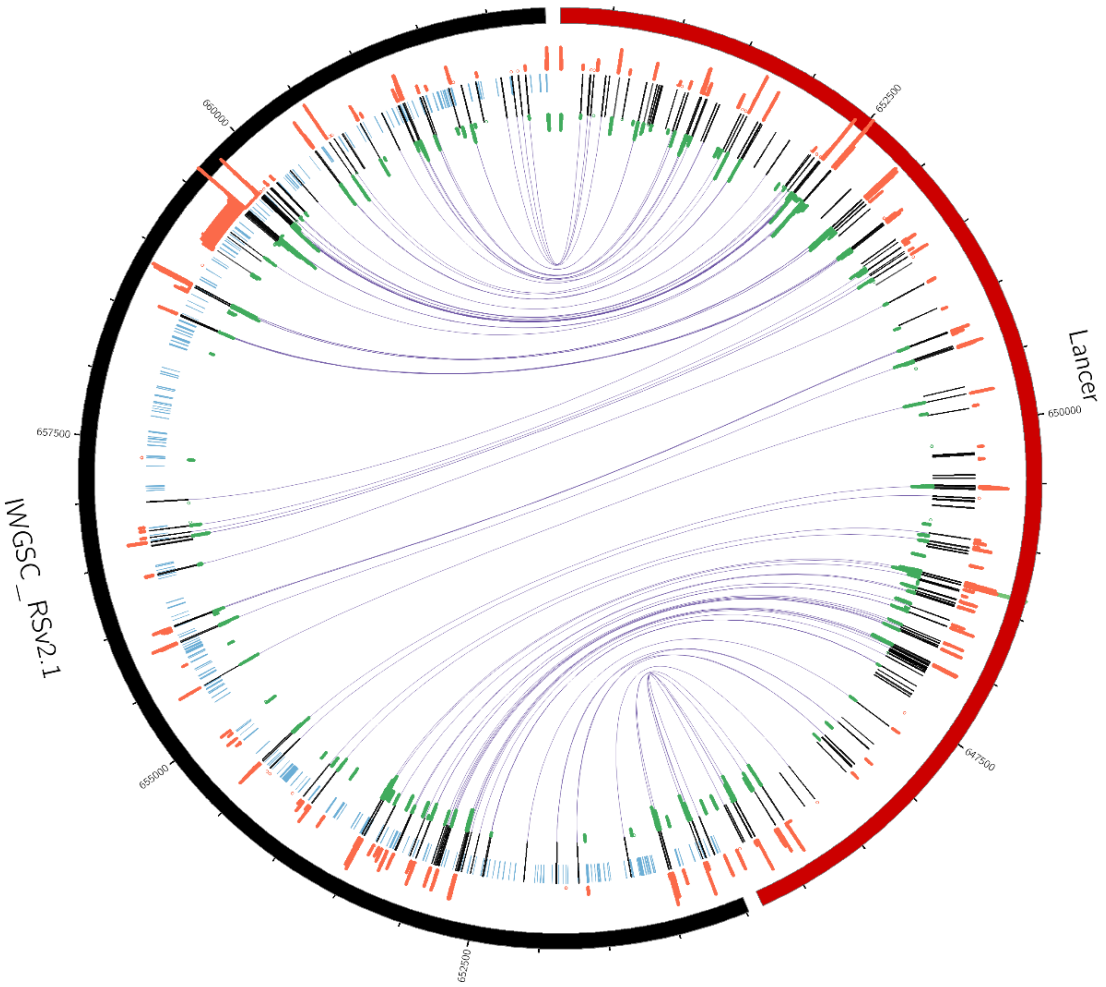


Figure 28a: Circos plot showing the genes (black tics) across the 1B QTL in both IWGSC and Lancer. Increments across QTL in 500kb intervals. Genes are joined together by links in brown indicating proteins belonging to the same HOG. Tics in blue represent newly annotated lncRNAs detected in the 2.1 IWGSC annotation. These are absent from the pangenome annotations. The inner histograms (orange and green) represent the log2 transformed RNA read coverage pileup at a locus across the 1B QTL in 1bp steps and mapped to IWGSC 2.1. Orange represents RNA derived from a high AX allele-containing YxV cross. Green represents RNA-seq from low AX allele-containing YxV cross.

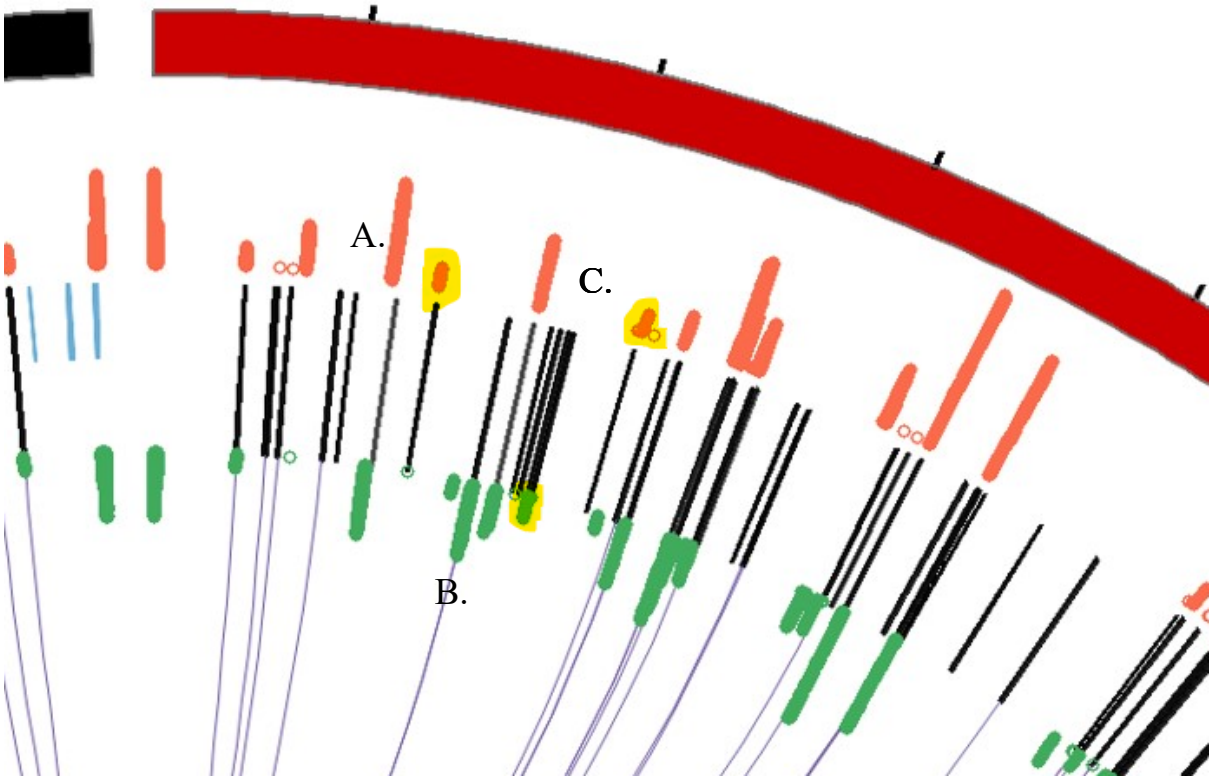


Figure 28b: Circos plot of the downstream 1B QTL end in Lancer with RNA-read coverage of Yumai-34 (orange) and Valoris (Green), from the YxV crosses used for RNA extraction, showing the only four genes in Lancer with no annotated orthologue in IWGSC v 2.1 and where there is zero expression in only one cultivar. Two genes, A. and C. expressed only in Yumai-34, and two closely positioned genes underlying B. expressed only in Valoris.

Of the 8 genes, 4 have complete differential expression between the high and low AX allele lines from the Yumai-34 x Valoris cross and can be identified from the Circos representation in Figure 28a. Moreover, many RNA transcripts are only present in one cultivar. Further study of these transcripts is therefore required to determine their identities and potential functions. A long block of reads mapping to the reference genome also occurs across 450Mbp at 65.9 Mbp in Chinese Spring. Reads were filtered to >Q30, which did not alter the mapping blocks, suggesting genuine read mappings. This could be due to chimeric read mapping, where transcription jumps from one locus to another (Frenkel-Morgenstern, 2020). The block at 65.9Mbp includes expressed genes for two β-purothionins and a probable RING E3 ubiquitin ligase. Reads across

the highly expressed β-purothionin genes using the interactive genome browser (IGV) show the potential for extensive splicing. These chimeric transcripts displayed in IGV coincide with the same block-patterning seen on the Circos plots, (Figures 28a and 28b). This chimeric transcription masks the differential transcription of two genes, with protein IDs XP_044360231.1 (a type-5 thionin like protein) and XP_044360241.1 (uncharacterised) being expressed in the low AX allele but not in the high allele. It is possible that trans-splice sites were introduced upstream of these genes in the high AX allele in Yumai-34, preventing transcription of the gene. Thionins are involved in plant defence and are cytotoxic to fungal cells (Höng *et al.*, 2021) but have no known relationship to cell wall polysaccharides.

Among the pangenome genotypes, Lancer, Mace and Julius were the most closely related to Paragon when considering SNPs in genes across the 1B QTL. As a chromosome level assembly was not available for Paragon, candidate gene identification was assisted by mapping RNA to these reference (pan)genomes (Lancer, Mace, Julius) alongside the IWGSC assemblies.
Mapping of the same RNA reads to Lancer did not show the same trans-splicing and while it is not known if Lancer has the high allele, it is the most similar to Yumai-34 in terms of the occurrence of SNPs within genes across the 1B QTL. Four additional candidate genes were identified from Lancer (Figure 28b), downstream of the RING-U gene. Two of these genes are expressed only in Yumai-34, and two expressed only in Valoris in the endosperm tissue and are shown in Table 10.

Table 10: Candidate genes derived from RNA-mapping of Lancer, based solely on zero expression in a single cultivar.

Start	End	ID	Protein domains
653906327	653907481	TraesLAC1B03G00389700	F-Box domain
654029389	654032521	TraesLAC1B03G00389770	F-Box domain
654038782	654042327	TraesLAC1B03G00389790	Armadillo-type fold
654310906	654311463	TraesLAC1B03G00389910	PROKAR_LIPOPROTEIN

4.3 Discussion

Pan genomes represent a more effective way to analyse QTL, particularly when the cultivar being assessed does not have a complete genome assembly. For example, in one study, utilising a novel pangenome construction pipeline (PSVCP), 12 representative rice genomes were analysed, which identified a new QTL (qPH8-1) associated with plant height on chromosome 8 (Wang *et al.*, 2024). This QTL was undetectable using traditional SNP-based GWAS approaches, anchored onto the Nipponbare reference genome. Another study, through a pangenome-wide association study (PanGWAS) combined with transcriptomic analysis, a novel QTL conferring resistance to both panicle and leaf blast diseases in rice was identified. The candidate genes associated with this QTL were absent in the Nipponbare reference genome, highlighting the advantage of pangenome approaches in QTL analysis (Wang *et al.*, 2023).

Four genes in Table 10 were identified as absent in the Chinese Spring reference, two belong to the F-box superfamily, and one to the Armadillo (ARM) repeat type superfamily. F-box proteins

are widely involved in regulating E3 protein ubiquitin ligase activity, being part of the SKF1-Cullin-Fbox (SCF) E3 ubiquitin ligase complex and provide components essential for substrate recognition. They have a wide variety of roles in plants. Specific roles in Arabidopsis include regulation of leaf senescence (Woo et al. 2001), branching of lateral shoots (Stirnberg et al. 2002) and circadian rhythms (Schultz 2001). In wheat, F-box proteins have been shown to be involved in regulating flower spike development, by interacting with S-phase kinase-associated proteins (SKP1) and cullin (a scaffold protein) to form an SCF E3 ligase complex that targets the proteins involved (Hong et al. 2012). Hong et al (2012) showed that the F-box protein TaCFBD was upregulated in early inflorescence development and active in recruiting the SCF ligase complex, targeting proteins for degradation by the 26s proteosome, although the substrates for the TaCFBD SCF ligase complex were not known. Moreover, analysis of published transcriptional analysis of wheat across various stages of embryogenesis and endosperm development (Xiang et al. 2019) showed 1206 *T.* F-box (TaFBX) genes, suggesting a wide role in regulating cellular functions during grain development. There are no published data on the role of TaFBX genes in regulating cell wall synthesis in wheat, but it is possible that SCF ligase complexes may degrade proteins involved in either the regulation or direct synthesis of xylan, or arabinosylation of the xylan backbone of arabinoxylan. The four genes in Lancer (Table 10) represent cloud genes, and are not part of the core machinery, however, it is impossible to assess their function on dietary fibre without determining the WE-AX content of the pangenome cultivars.

Chapter 5
Bulked Segregant Analysis of Paragon x Watkins RIL Populations to Reduce the Size of the 1B QTL Associated With WE-Pentosans.

5.1 Introduction.

Several studies have identified the 1B locus associated with the contents of WE-AX and TOT-AX in white flour. Notably, Lovegrove *et al.,* (2020), identified this locus in the Chinese bread wheat Yumai-34 by analysing the progeny of crosses with the cultivars Claire, Ukrainka and Valoris, while Ibba *et al.,* (2021) identified genetic markers within the QTL through a genome-wide association study (GWAS). Ibba *et al., (2021)* also suggested putative candidate genes from the GWAS study, based on gene annotations and proximity to associated markers, but did not confirm their role in AX synthesis or expression in endosperm tissues. In addition to the mapping of biparental populations, Lovegrove *et al.* (2020) also reported a GWAS study on the WHEALBI panel of bread wheat genotypes. This identified significant markers associated with relative viscosity and WE-AX which mapped close to those later reported by Ibba *et al.,* (2021). The most significant marker for this association was in the gene: TraesCS1B02G426200, located within the 1B QTL.

5.1.1 Brief overview of bulked segregant analyses

Bulked segregant analysis is an effective method for QTL detection (Michelmore et al., 1991). The original methodology was based on the formation of two contrasting pools, or 'bulks', from segregating populations that are phenotypically or genetically identical for the trait of interest. DNA fractions from these two contrasting bulks are then screened for non-segregating markers to highlight regions linked to the trait of interest. The methodology has since been modified,

and a tailed approach to genotyping described (Darvasi and Soller, 1995), whereby the phenotypic tails of a segregating population are pooled into contrasting bulks and genotyped. More recently, multiple modifications have been made for next generation sequencing (NGS), such as using Euclidean-distance statistics to improve QTL detection (de la Fuente Cantó and Vigouroux, 2022) and used to map QTLs in a range of species, including *Brassica napus* (Itoh *et al.*, 2019), melon (Chayut *et al.*, 2011), and wheat, where QTL for fusarium head blight were detected on chromosomes 3BS, 2D and 6B in a population of recombinant inbred lines (RILs) (Shen *et al.*, 2003)candidate gene was also identified in bread wheat using a modified bulked segregant analysis (BSA) approach (Li *et al.*, 2020), where the aim was to use RNA transcript reads derived from the two contrasting bulks to identify significant differential expression of putative candidate genes. This modified approach, termed BSR-seq, was also successfully used to link the basic 7S globulin gene on chromosome 3D to Hagberg Falling Number- an important wheat quality trait relating to α-amylase activity and starch quality. This shows that BSA is a highly versatile method that can be used to detect QTLs and indicate putative candidate genes.

5.1.2 The PxW RIL populations.

As discussed in Chapter 3, the 1B QTL allele from Paragon has an increasing effect on WE-AX in two of the PxW mapping populations (PxW471 and W694), and accounts for around. It therefore provides an opportunity to determine the presence of the increasing allele in the Watkins collection by using the PxW RIL collection at the John Innes Centre (JIC). The JIC has produced about 60 f4-f7 RIL populations by crossing Paragon with different Watkins accessions. Of these ~60 populations, 24 were in current stock and had been genotyped using either a KASP array or the Axiom 35K breeders' array. All of these populations were grown at the JIC between the years 2014-2022. 24, where available, were collected from each population, and divided equally between allelic classes, such that 15 lines contained the Paragon 1B allele and 15 contained the

respective Watkins allele. Equal numbers of mature grain from the lines of each allelic class were pooled together into two bulks, one comprising the grain from the 15 lines from the Paragon allelic class, and the other sample comprising the 15 lines from the Watkins allelic class. Following the methodology described in section 2.12, the study aimed to identify regions of the QTL most closely associated with WE-pentosans, by applying genotypically selected bulked segregant analysis to 24 different RIL populations of Paragon x Watkins accessions. The QTL locus on 1B has been detected in multiple studies of AX content in bread wheat. However, the mechanistic mode of action of the QTL on AX synthesis has not been identified. The QTL locus of 1B, defined from the QTL analysis in section 3, contains 110-115 protein-coding genes and spans 8-11Mbp on the long arm of chromosome 1B. This makes identifying candidate genes difficult without reducing the QTL to a tighter interval and therefore reducing the number of putative causal genes, however, the gene tables and their annotations, for genes which are expressed in endosperm tissue can be found in supplementary table 8.

5.2 Results

5.1 Allele effects on phenotype in mapping populations

Lines with parental haplotypes, Watkins and Paragon, across the QTL locus were selected and grouped into the two allelic classes. The WE-AX content of lines within these allelic classes were determined and are shown as boxplots for 2019 (Figure 29 A), 2020 (Figure 29 B) and 2021 (Figure 29 C). A significantly lower content of WE-AX was observed within the Watkins allelic class, compared to the Paragon class in populations ParW471 (P= 6.6e-06) and ParW694 (P=0.028), but not in ParW145 (P=0.55) in 2019. This agrees with the results of the QTL analysis of these three populations grown in 2019, as the 1B QTL was detected in ParW471 and ParW694, but not in ParW145 in 2019, (Section 3, Table 8). The QTL on 1B was not detected in 2020 with no significant effects of allelic class on the WE-AX contents in any of the populations (Figure 29

B). This may be due effects of environmental factors, such as rainfall or temperature at harvest, as we know that the AX trait has a 30% environmental component, which could obscure the allelic differences in different years. The QTL was subsequently detected in 2021 for the ParW471 population, and can be seen in Figure 29 C, with a significant difference between allelic classes (P= 2e-09) suggesting that environmental effects may have masked the effect of the QTL in 2020 from the 'rQTL' detection pipeline, though the effect is still evident in Figure 29 B.

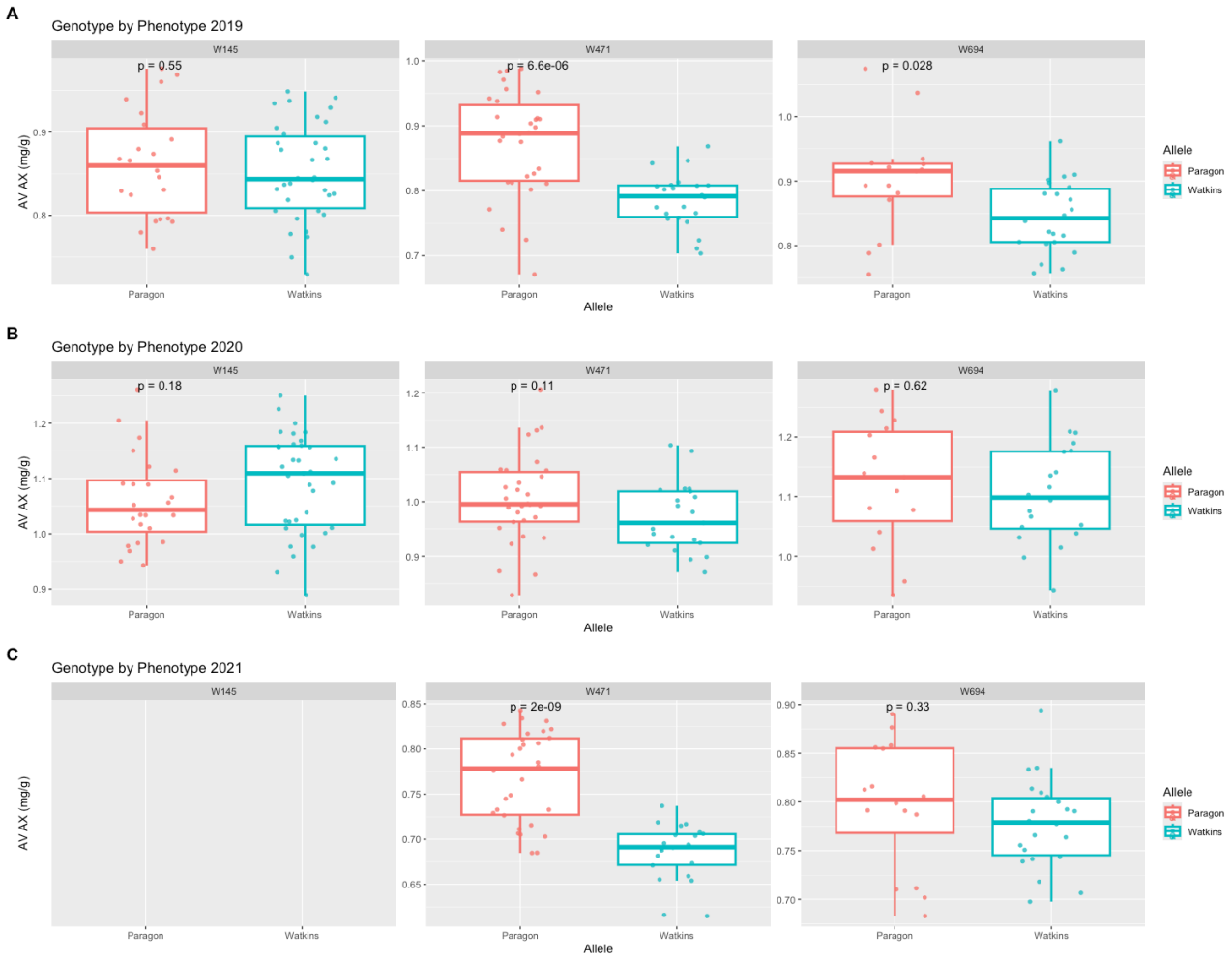


Figure 29 A and B: Genotype by phenotype boxplots plot showing distribution of WE-AX content between parental allelic classes across the 1B QTL in three experimental Paragon x Watkins mapping populations in the years 2019(A) and 2020 (B) and 2021 (C) and statistical T test for the difference in the mean WE-AX (% per mg) between allele classes.

To test whether a bulked segregant approach would work for determining the alleles present in the different Watkins accessions for which Paragon x Watkins recombinant inbred lines were available, ParW145 and ParW471 were selected for a bulked segregant analysis, as these two

populations consistently showed the presence of (ParW471) or failed to show the presence of (ParW145) the 1B QTL. The mean WE-AX content of the bulked flour of the 1B Paragon allelic class, 'Bulk A' (9.1mg/g) was significantly higher in ParW471 than the Watkins allelic class, 'Bulk B' (7.9mg/g) (P=0.033), but no differences were observed between the two classes in ParW145 (P=0.9), as shown in Figure 30

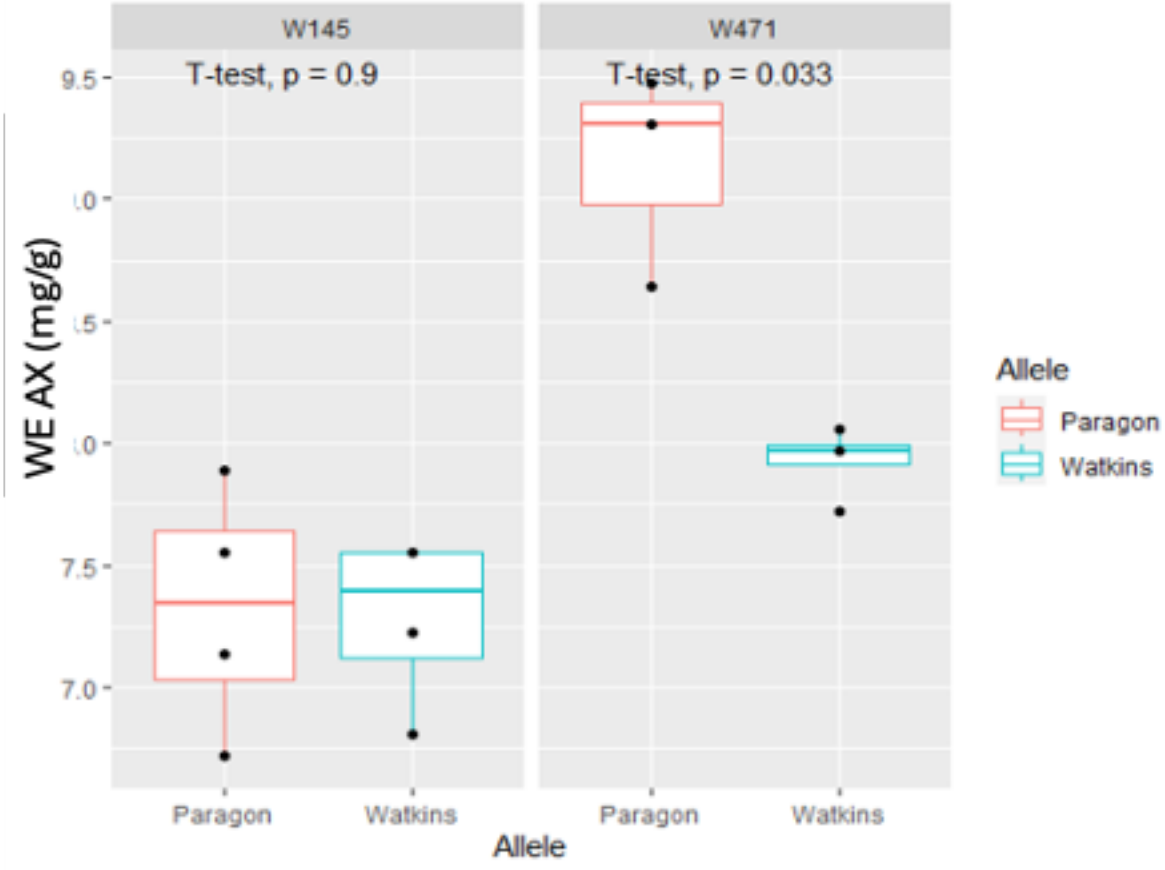


Figure 30: Box plots showing the distribution of replicate WE-pentosans from the two bulk samples corresponding to the respective allelic class (Paragon and Watkins). T test P values displayed show significance of the mean difference between the two allelic classes according to the respective population (W145 and W471) which the two allelic classes were derived.

5.2.1 Bulked segregant analysis on the 24 Watkins lines

Following the methodology from section 2.12, the 24 Paragon x Watkins populations were selected based on availability at the John Innes Centre, pooled into allelic classes across the 1B QTL, phenotyped for WE-pentosans and a t-test of significance was performed to determine significant differences in the WE-AX content for each bulk within a population (Table 11). The WE-AX content is categorised as 'high' or. 'low', only relative to the population being measured, and not an indicator of nutritionally high WE-AX content. From this, 15 Watkins accessions containing the high allele and 9 Watkins accessions with the low allele were identified (Table 12). Three populations, WATDE0013, WATDE0039 and WATDE0042, defined in table 12 as 'high-Watkins', deviated from the predicted trends of the Watkins bulk as it was expected that the Watkins bulk, WE-pentosan measurements would be equal to, or below, that of the Paragon bulk. These three populations, however, were shown to have significantly higher WE-contents than the Paragon bulk (P=0.003, 0.04 and 0.01 respectively), suggesting that the Watkins allele in these accessions may be more advantageous in increasing WE-AX content than the Paragon allele.

Table 11: Table showing the mean AX derived from the WE-pentosans assay for each of the bulks, A= Paragon, B =Watkins, for each population. Last three digits of the Watkins accession given in 'Sample'. T test P values show the significance of the difference between the mean AX content between each bulk within a population. An asterisk on the population name highlights a significant difference observed between bulk groups A and B.

Sample	Group	Mean AX (mg/g)	SD	Variance	T_test Pvalue
685 *	A	6.366	0.074	1.155	0.000
685	B	5.064	0.059	1.160	
45	A	6.762	0.205	3.000	0.250
45	B	7.193	0.516	7.170	
700 *	A	7.782	0.039	0.542	0.000
700	B	5.716	0.105	1.845	
740 *	A	8.778	0.211	3.128	0.001
740	B	6.649	0.116	2.062	
694 *	A	8.194	0.183	2.238	0.012
694	B	7.350	0.277	3.764	
103 *	A	5.749	0.018	0.310	0.003
103	B	5.071	0.058	0.977	
471 *	A	7.713	0.141	1.827	0.000
471	B	6.571	0.080	1.210	
475 *	A	7.085	0.175	2.471	0.005
475	B	6.232	0.188	2.997	
79	A	5.001	0.460	12.051	0.111
79	B	5.955	0.369	6.196	
110 *	A	7.355	0.419	5.695	0.016
110	B	5.934	0.674	12.174	
456	A	7.621	0.258	3.406	0.167
456	B	7.247	0.283	3.910	
274	A	7.975	0.231	2.893	0.784
474	A	7.871	0.574	7.290	
474	B	5.973	0.248	4.106	0.556
273	A	8.173	0.479	6.113	
273	B	7.377	0.786	10.654	0.423
398	B	7.849	0.472	6.016	
145	A	6.023	1.180	19.552	0.956
145	B	6.084	0.226	3.572	
397	A	8.169	0.220	2.695	0.840
397	B	8.138	0.118	1.450	
299 *	A	7.416	0.419	5.648	0.042
299	B	8.136	0.075	0.921	
238	A	6.722	0.306	4.068	0.577
238	B	6.727	0.427	6.352	
300	A	6.053	0.291	4.812	0.189
300	B	5.683	0.281	4.953	
560 *	A	5.874	0.209	3.091	0.047
560	B	3.874	0.582	14.284	
811	A	6.932	0.274	3.959	0.962
811	B	6.919	0.308	4.445	
308	A	4.677	0.083	1.777	0.131
308	B	3.658	0.411	9.745	
749 *	A	7.459	0.596	7.961	0.046
749	B	6.216	0.286	4.598	
750	A	6.396	0.128	2.000	0.210
750	B	6.001	0.393	6.489	
49	A	8.122	0.113	1.394	0.773
49	B	8.041	0.510	6.339	
216	A	7.676	0.778	10.140	0.746
216	B	7.839	0.234	2.988	
223	A	7.618	0.562	7.364	0.342
223	B	8.169	0.455	7.935	
44	A	8.803	0.377	4.288	0.495
44	B	8.579	0.124	1.444	
704 *	A	10.237	0.240	2.315	0.002
704	B	8.728	0.259	2.793	
254	A	7.907	0.393	5.317	0.286
254	B	8.088	0.133	1.641	

Table 12: The Watkins accessions selected for the bulked segregant analysis. The WATDE codes are used by the GRU seed store at the JIC, and the allele determined by phenotyping of the bulked segregants. Allele R is the new code associated on with each Watkins accession in Figure 30. The number of lines in each bulked sample =15.

Watkins accession	WATDE Code	allele	Allele R
1190045-1	WATDE0009	high	H1
1190079-1	WATDE0010	high	H2
1190145-1	WATDE0019	high	H3
1190273-1	WATDE0036	high	H4
1190397-1	WATDE0052	high	H5
1190398-1	WATDE0053	high	H6
1190474-1	WATDE0063	high	H7
1190546-1	WATDE0069	high	H8
1190238-1	WATDE0031	high	H9
1190300-1	WATDE0040	high	H10
1190750-1	WATDE0108	high	H11
1190811-1	WATDE0115	high	H12
1190103-1	WATDE0013	high-watkins	HW1
1190299-1	WATDE0039	high-watkins	HW2
1190308-1	WATDE0042	high-watkins	HW3
1190471-1	WATDE0062	low	L1
1190475-1	WATDE0064	low	L2
1190685-1	WATDE0091	low	L3
1190694-2	WATDE0093	low	L4
1190700-1	WATDE0095	low	L5
1190740-2	WATDE0103	low	L6
1190110-1	WATDE0014	low	L7
1190560-1	WATDE0071	low	L8
1190749-1	WATDE0107	low	L9

5.2.2 Reducing the QTL area using the haplotypes across the 1B QTL in the three Watkins parents.

In order to reduce the QTL region surrounding the high AX allele, the haplotypes of each Watkins accession from the three original mapping populations were examined on a gene-per-gene basis across the QTL. Analysing the haplotypes of the genes within the region of the 1B QTL, with knowledge of the alleles of the Watkins accessions (W145, increasing allele; W471 and W694,

decreasing allele) and their WE-AX content, allowed the QTL to be reduced in size, as shown schematically in Figure 31-A (the new (reduced) region of the QTL is shown in orange). This was achieved by comparing genes which shared haplotypes between increasing and decreasing allelic classes. The start of the new 1B QTL boundary originates at TraesCS1B02G425900, where both Low allele Watkins accessions share the same haplotype (8) but differs from W145, which has a gene haplotype (10). The end boundary was assigned to TraesCS1B02G430600 where the same haplotypes were observed and is indicated in figure 31-B and shown more clearly in figure 32, drawing the boundary from the green haplotype 8, and purple haplotype 10 at either side.

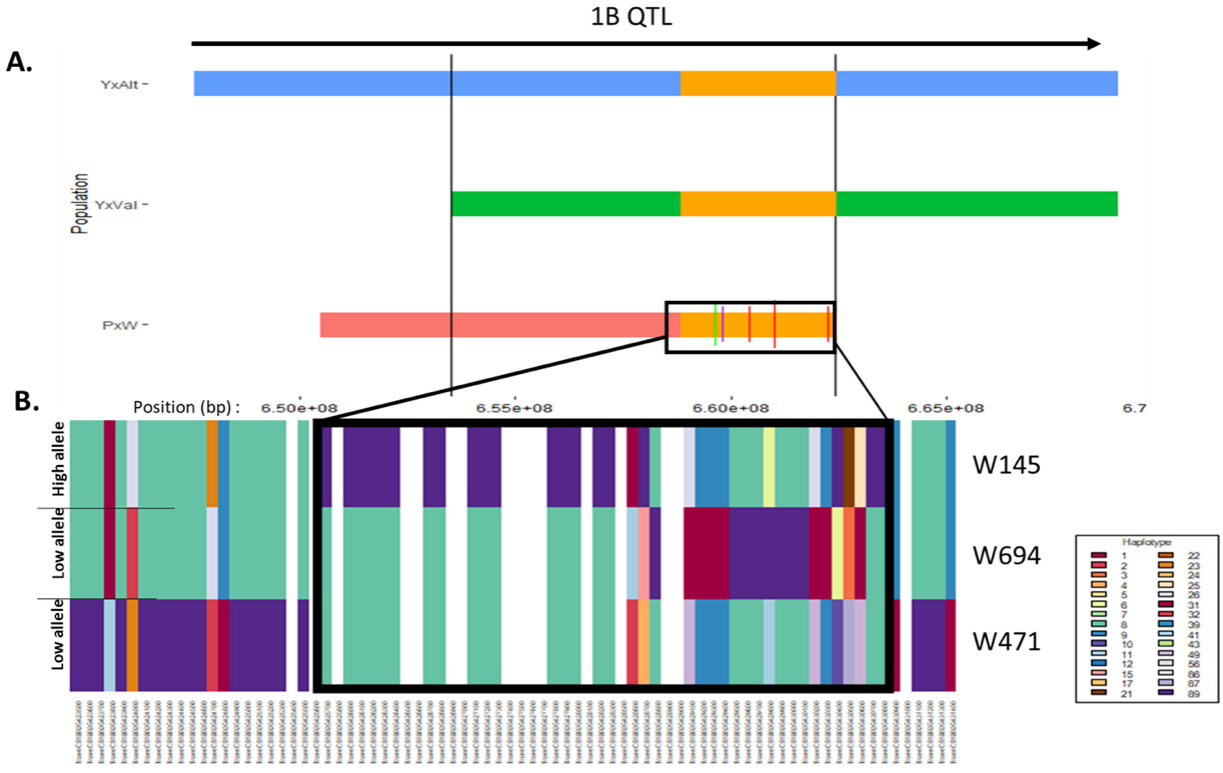


Figure 31: (A) to scale schematic representation of the QTL boundaries in the ParW471, Yumai34 x Valoris and Yumai34 x Altigo populations. The orange overlay across this schematic is the proposed reduced QTL region, whilst the horizontal-coloured bars indicate significant markers associated with AX content (Green from the Whealbi panel, Lovegrove et al., (2020), red markers from Ibba et al., (2021) and purple being a purothionin gene. (B) The haplotype calls from AGIS, China, for each gene across the 1B QTL in each of the three selected mapping ParW populations (W145; W694; W471).

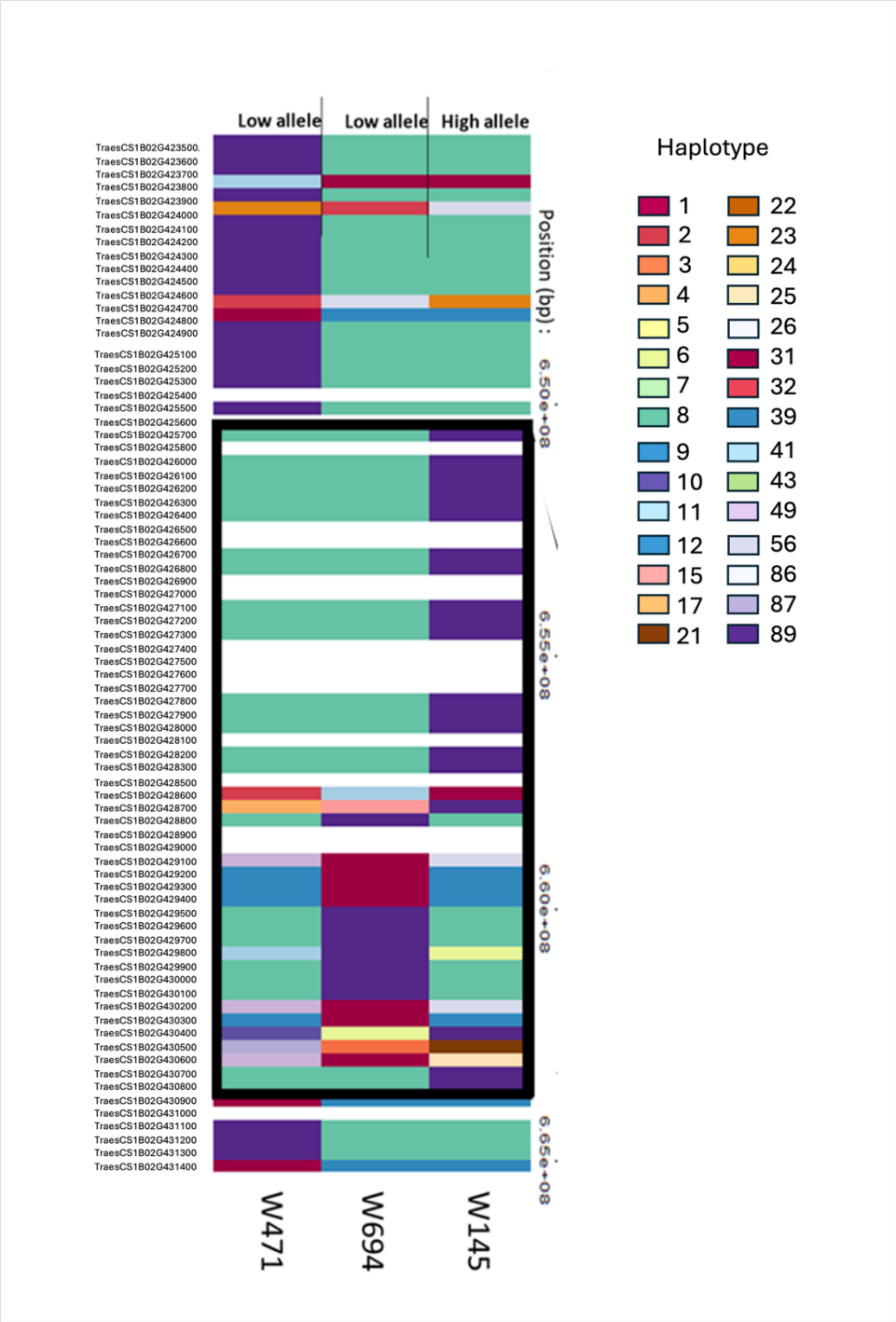


Figure 32: Zoomed view of the schematic haplotypes of each Watkins parental accession from Figure 30, B. Coloured haplotypes depict the haplotype of each gene, determined from the Watseq project, data derived from Shifeng *et al., (unpublished)*

To narrow this reduced region further, the haplotypes of the 24 others Watkins accessions, identified by bulked segregant analysis were examined across the QTL, with phenotypic data shown (Table 10) (Full phenotyping results in Appendix, supplementary table 5). This panel of

24 includes the original W145, W471 and W694 accessions, but derived from field material from the John Innes Centre, and not the original populations grown for this study, at Rothamsted Research. Figure 32A shows the gene haplotype for each of the Watkins accessions, which have been named according to the information in Table 12. Only 24 populations were available for the haplotype analysis, due to missing samples from the gene haplotype data provided by AGIS, China (Shifeng *et al., 2024)*. To show the regions of interest, genes are highlighted depending on whether the two allele classes (High and Low) differed in haplotypes but shared a common haplotype within allele classes (Figure 33B). This identified the regions around TraesCS1B02G426100, which includes 7 potential genes, and TraesCS1B02G430300, including 4 potential gene candidates.

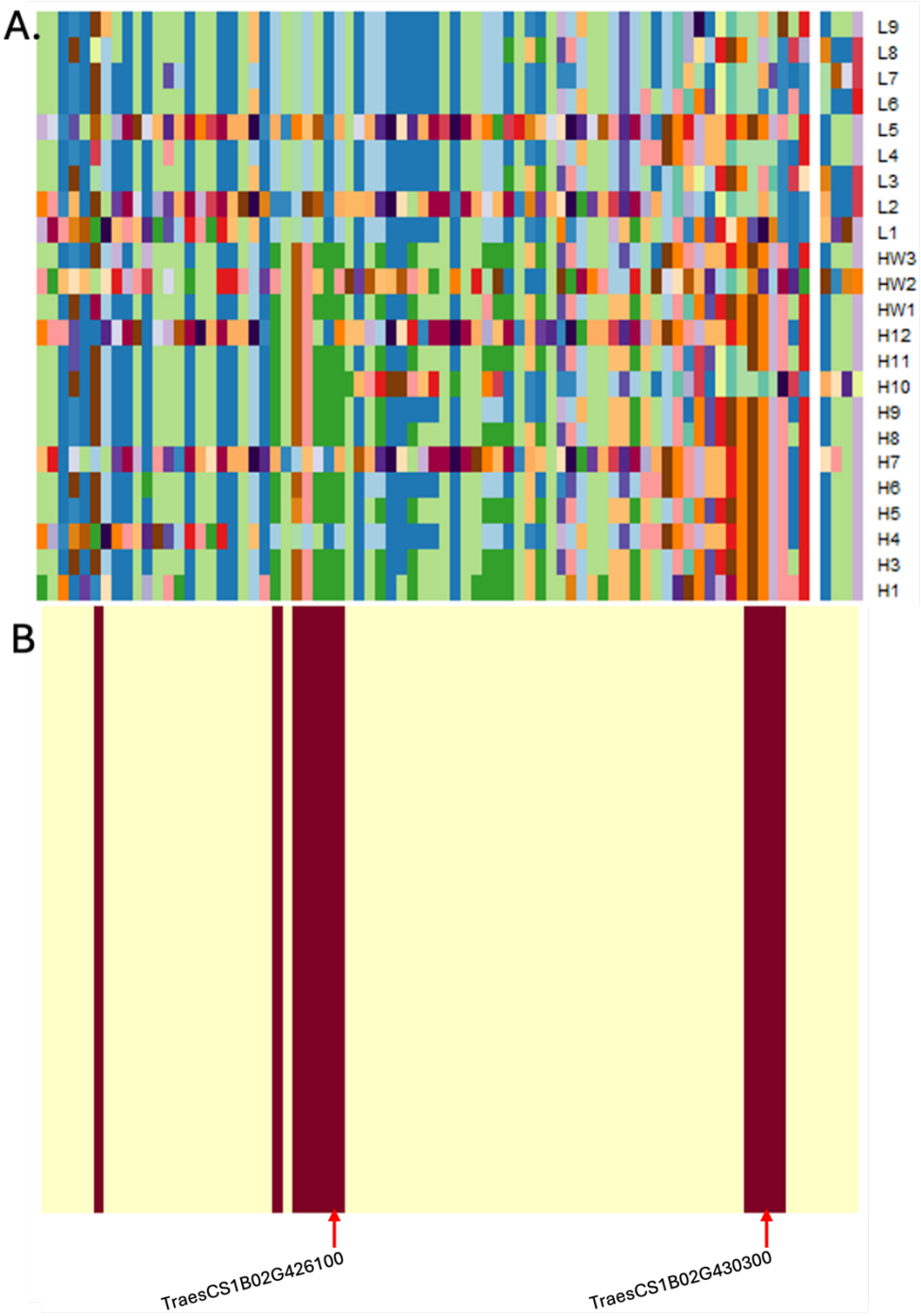


Figure 33. A, heatmap depicting the haplotypes within each gene across the 1B QTL in each Watkins parent, assigned to an allelic class, H, HW or L (see table 11); B, shows the region in the heatmap where the consensus haplotype was the same within but different between allelic classes.

The difference in haplotypes around TraesCS1B02G426100 (encoding an α-purothionin) and the adjacent gene TraesCS1B02G426200 (encoding an E3 RING-Ubiquitin ligase (Ring-U)), indicated by the leftmost arrow on Figure 33-B is consistent with the report of Lovegrove *et al.,* (2020), in which the peak GWAS marker was within the Ring-U gene. The second region, around

TraesCS1B02G423030, indicated by the rightmost arrow on Figure 33-B, is in agreement with the marker determined by GWAS at CIMMYT on a collection of elite bread wheats (Ibba *et al.,* 2021). This marker was the third most significant marker associated with WE-AX on chromosome 1B and lies close to the region around TraesCS1B02G423030.

Secondly, a GWAS study performed at the JIC on a panel of elite bread wheats, consisting of 320 commercial breeding and pre-breeding lines identified several markers on 1B that were significantly associated with the content of WE-AX determined at Rothamsted Research. The most significant marker, identified using the new 45K TanG SNP genotyping array, was close to TraesCS1B02G423030. The second most significant markers were KASP markers designed for the purothionin and RING-U genes, that were used to genotype the elite bread wheat panel and subsequently used in the GWAS pipeline using a general linear model (Wingen et al., unpublished). However, subsequent re-analysis of the GWAS data at the John Innes Centre showed that more appropriate models that account for the relatedness of the elite genotypes did not identify the marker close to TraesCS1B02G423030 with the most significant marker being the KASP marker designed for a SNP in the RING-U gene (Alabdullah et al., unpublished).

5.3 Yumai-34 x Valoris RNA-seq data to evaluate the two small regions within the 1B QTL.

Variant calling files (VCF) were compiled from RNA-seq data generated by Novogene using RNA fractions from endosperm tissues, at 21dpa, prepared from bulk segregants at Rothamsted (Oszvald et al., unpublished). Four sets of DH lines from the Yumai-34 x Valoris cross were selected based on results from Lovegrove *et al.,* (2020), each comprising five lines with the increasing alleles at the 1B QTL from Yumai-34 (sets A and D) and the decreasing allele from

Valoris (sets B and C), which were determined from the genotypes in the Axiom linkage maps, with three replicate RNA samples from each set being analysed.

Variants were called from the RNA-seq reads corresponding to SNPs occurring between Yumai-34 or Valoris against Paragon (section 2.11). The VCFs generated for each replicate per group were plotted as a heatmap (Figure 34) showing the presence of the SNPs across the two regions identified in figure 33. The SNPs are displayed as coloured bars, while the white spaces show where the samples have the same SNP as the reference sequence (Paragon). Groups B and C have SNPs in the purothionin and Ring-U genes, indicated by the green to yellow regions in figure 34. Furthermore, the regions upstream and downstream of the purothionin and Ring-U genes in Valoris are the most dissimilar to Paragon, whereas the regions in Yumai-34 (A and D) shows a high degree of similarity to Paragon. Moreover, the region surrounding the marker identified by GWAS using the general linear model (Wingen et al., unpublished) showed that there are no major similarities between Yumai-34 and Paragon around this locus, indicating that the candidate gene is closer to the purothionin and RING-U genes. This is also consistent with the failure to identify the marker close to TraesCS1B02G423030 using the updated mixed linear model for GWAS (Alabdullah et al., *unpublished)*

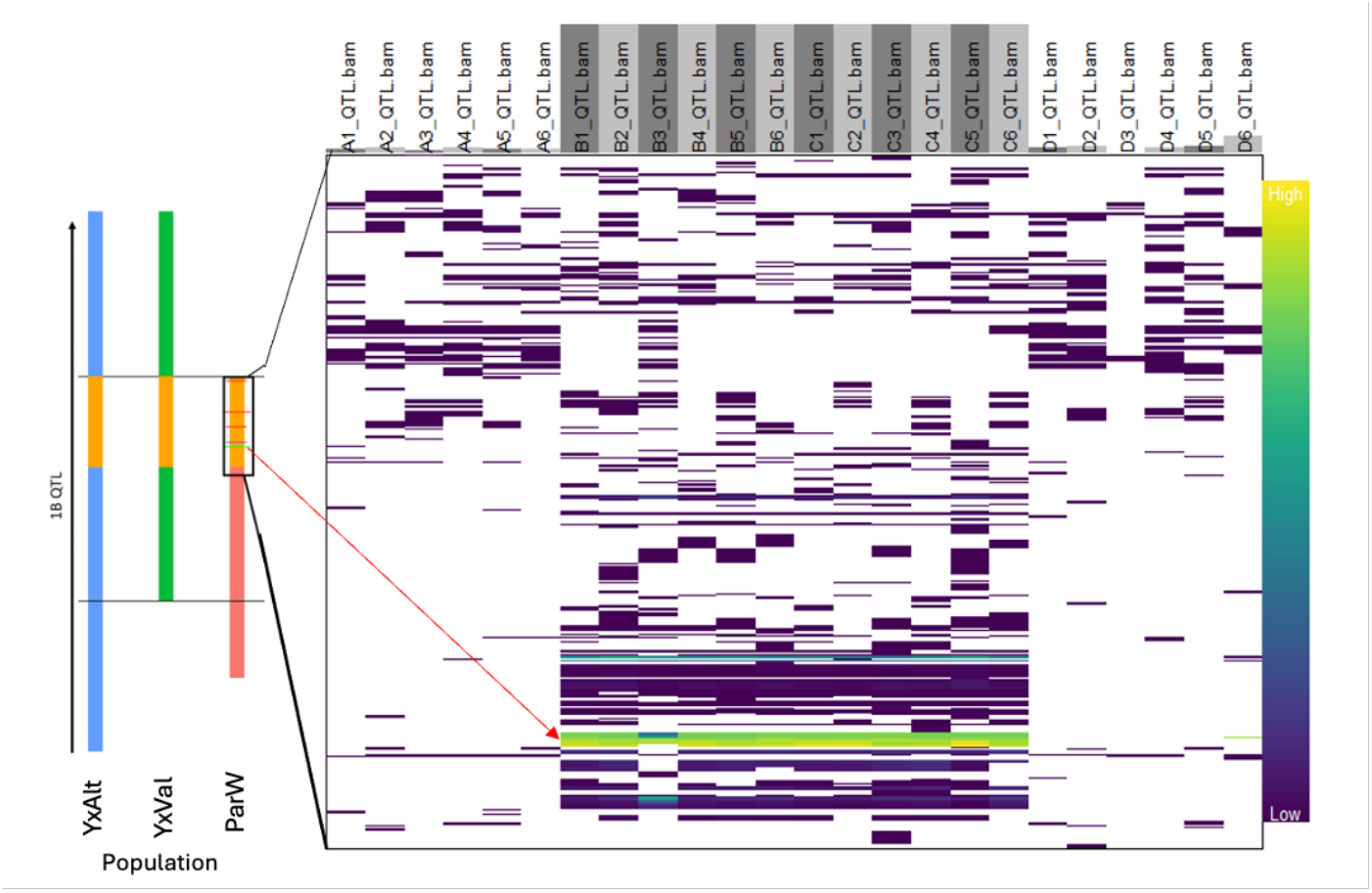


Figure 34: heatmap of SNPs from RNA-seq of Yumai-34 (group A and D) and Valoris (groups B and C, marked as grey on the sample name line). Blank space indicates where the SNP position is the same as the reference (Paragon) and the colour scale indicates read depth. SNPs are placed in positional order, where the reduced QTL boundary is at the bottom of the heatmap and proceeds upwards to the end of the reduced QTL boundary (orange) described in Figure 30, and shown here as a cutout, to the left of the heatmap. Colour scales on heatmap show the read depth of mapped RNA reads to Paragon..The Red arrow points to the location of the Purothionin putative candidate gene marker on the ParW QTL and its position in the heatmap.

5.4 Discussion

This study aimed to refine the 1B QTL associated with water-extractable arabinoxylans (WE-AX) in white flour by applying a genotypically informed bulked segregant analysis (BSA) approach to a diverse set of Paragon x Watkins recombinant inbred lines (RILs). Previous studies had established the relevance of the 1B locus to WE-AX and total AX content (Lovegrove, Wingen, Plummer, Wood, Passmore, Kosik, Freeman, Rowan A.C. Mitchell, *et al.*, 2020b), but the wide

QTL interval spanning up to 11 Mbp and encompassing up to 115 genes had hindered the identification of candidate causal genes. By leveraging both phenotypic and haplotype data, this work successfully reduced the candidate region and identified novel alleles within the Watkins collection that could enhance WE-AX content.

The consistent detection of the QTL in the ParW471 and ParW694 populations, but not in ParW145, provided a strong experimental framework to test the resolution capacity of BSA. In the selected RILs, the bulked grain analysis showed that allelic effects were largely concordant with previous QTL analyses, although year-to-year environmental variation could obscure these effects.

Crucially, applying BSA across 24 PxW populations enabled the phenotypic classification of Watkins alleles as either high or low for WE-AX. While the majority followed the expected trend of Paragon contributing the high, WE-AX allele, three Watkins accessions (WATDE0013, WATDE0039, WATDE0042) outperformed Paragon. This suggests that certain Watkins lines contain putative, novel, and potentially superior alleles, demonstrating the untapped diversity within this collection. Crucially, these three lines may be further scrutinised when a candidate gene is identified underlying the 1B locus, to ascertain the differences between these three Watkins lines and the Paragon/ Yumai-34 allele

Haplotype analysis further refined the QTL region by leveraging recombination events and allele-specific patterns across both the original mapping populations and the wider Watkins panel. This narrowed the interval to two sub-regions around TraesCS1B02G426100 and TraesCS1B02G430300, collectively containing 11 candidate genes. These findings significantly enhance the resolution of the 1B QTL and provide tangible targets for future functional studies or marker-assisted selection efforts.

In summary, this demonstrates that the integration of bulked segregant analysis, haplotype dissection, and historical QTL data can effectively dissect complex loci in wheat. The

identification of novel high AX alleles in Watkins accessions not only validates the approach but also highlights new genetic resources for improving dietary fibre content in modern wheat varieties.

Chapter 6
Identification of Putative Candidate genes

6.1 Introduction

It is important to identify the candidate gene for the high AX 1B QTL in order to develop perfect markers for selection by breeders, as previous markers have been based on linkage determined by biparental crosses and GWAS (Lovegrove *et al.,*2020; Ibba *et al.,*2020*)*. The following study has therefore focused on a set of putative candidate genes identified within the reduced 1B QTL to identify putative functions of these genes in controlling AX accumulation and therefore their role in defining the higher A phenotype observed.

6.2 Results

6.2.1 Differentially expressed putative candidate genes between Yumai-34 and Valoris

The differential RNA expression analysis performed on the four bulked segregant groups derived from the Yumai-34 and Valoris population, described in section 2.12, were mapped against the Chinese Spring RefSeq v2.1 available from NCBI. The most recent annotation of Chinese spring was used, as it provides the most reliable data on gene annotation. Of the genes present in the reduced 1B QTL boundary, only three were significantly differentially expressed between Yumai-34 and Valoris. The first, is a type-IV thionin (NCBI accession LOC123081759), which was upregulated in Yumai-34 with a positive Log2 fold change of 15.12927. The second is an uncharacterised gene (NCBI accession LOC123081771), containing DUF4220 and DUF594 with a Log2 Fold change of 7.614335. The final gene is phytoene-synthase related (NCBI accession LOC123146215), with a Log2 fold change of 6.541014.

6.2.2 Assessing the α-Purothionin gene as a Putative Candidate Gene

The reduced QTL region from Chinese Spring 2.1 contains 28 genes that are expressed in endosperm tissues, only 10 of which have more than 10 normalised RNA read counts. The most highly expressed gene is XP_044421416 (TraesCS1B02G426100), encoding α-1 purothionin. Analysis of endosperm specific RNA-seq data from the Yumai-34 x Valoris bulk segregants (unpublished work of M. Oszvald), showed that XP_044421416 is the most abundantly expressed gene in the region, with a normalised read count of up to 166,000 reads, compared to other genes in the QTL, ranging from ~40-3000 normalised read counts. The RNA-seq reads were mapped to Chinese spring v1.1 (Refseq1.1) by Novogene and remapped in this study to Chinese Spring v2.1 (Refseq2.1). Both alignments were then analysed for differential gene expression using the DESEQ2 pipeline. XP_044421416 showed significantly lower gene expression in the Yumai-34 (high 1B allele) than in Valoris (low 1B allele) ($P<0.05$), when mapped to v1.1 of the Chinese Spring reference, but not when mapped to v.2.1 (Table 11). A much higher log2 fold change (FC) value of 2.0916 was given by the DEG analysis using Refseq1.1 than Refseq2.1 at 0.6511. Overall, ~164,000 more reads mapped to the Refseq2.1 assembly than v1.1, indicating that the mapping was more accurate.

Table 13: table showing RNA reads mapping to both v1.1 and v2.1 of the Chinese spring reference genome in the low 1B allele groups (B, C) and the high allele (A, D) groups for the α-1 purothionin gene.

Purothionin V2.1 x V1.1				
Refseq Version	Read count (High group)	Read count (Low group)	log2FC	FC
V2.1	199247.2	312891.9	0.651105	1.57037
V1.1	34434.05	146761.4	2.091564	4.2621
Difference	164813.2	166130.5		
%decrease	82.71792	53.09518		

Furthermore, it was shown that the sequence of the high allele from Yumai-34 was more similar to the D homeologue of the α–1purothionin gene than the Valoris allele. To determine whether the differential gene expression was genuine or due to artefacts of read alignment, an early access assembled sequence of chromosome 1B from Paragon was obtained from the Earlham Institute (Hall *et al*., unpublished). The 1B chromosome from the Chinese spring v2.1 assembly was replaced with the Paragon 1B chromosome to create a new assembly (RefSeq_Par1B) and was used to map the reads from the RNA-seq groups A and D (Yumai-34 1B allele). The DEG analysis was repeated using feature counts from groups A and D aligned to the RefSeq_Par1B assembly and the feature counts from groups B and C aligned to the unmodified Refseq2.1 assembly. Groups A and D were combined into a single group of samples containing the 1B Yumai34 allele and groups B and C into a group containing the Valoris allele. There was no significant differential expression of the α–1 purothionin gene using this method and the mean normalised read count between these two groups was approximately equal (Table 13), suggesting that the differences in aligned read counts in the RefSeq assemblies was probably due to artefacts arising from misalignment of the 1B □–1purothionin reads to the homeologues on 1A and 1D.

Table 14: The mean normalised read count of the α–1purothionin gene from the Yumai34 x Valoris RNA-seq datasets. Groups A and D were mapped to the Chinese spring Refseq 2.1 genome, with the Paragon 1B chromosome assembly in place of the Refseq2.1 1B chromosome assembly. Groups C and B were mapped to Chinese Spring Refseq2.1.

	Group	
	A & D	C & B
Mean Read Count	126383.1	130532.1163

The wild-type form of TraesCS1B02G426100 transcribes a protein on the missense strand, with a terminal 3' tyrosine residue; however, the Paragon and Yumai-34 alleles contain a stop-gain nonsense mutation. The alignment shows that the low-AX allele produces more transcripts aligning further into the 3' region, while the high AX allele may result in truncation of most transcripts, (Figure 35), which may result in the loss of functional domains or key regulatory 3' elements. The nonsense mutation that is gained contains an 'ochre' (UAA) stop codon which would result in termination of translation of the protein in Yumai-34, truncating by up to 46AA (Figure 36), and thus probably also in Paragon, but this mutation is not observed in Valoris. As a result, there is a change in the terminal protein of the Yumai-34 isoform, which modifies an 'AAT' codon to 'GAA', encoding a glutamine as opposed to a tyrosine residue in the protein.

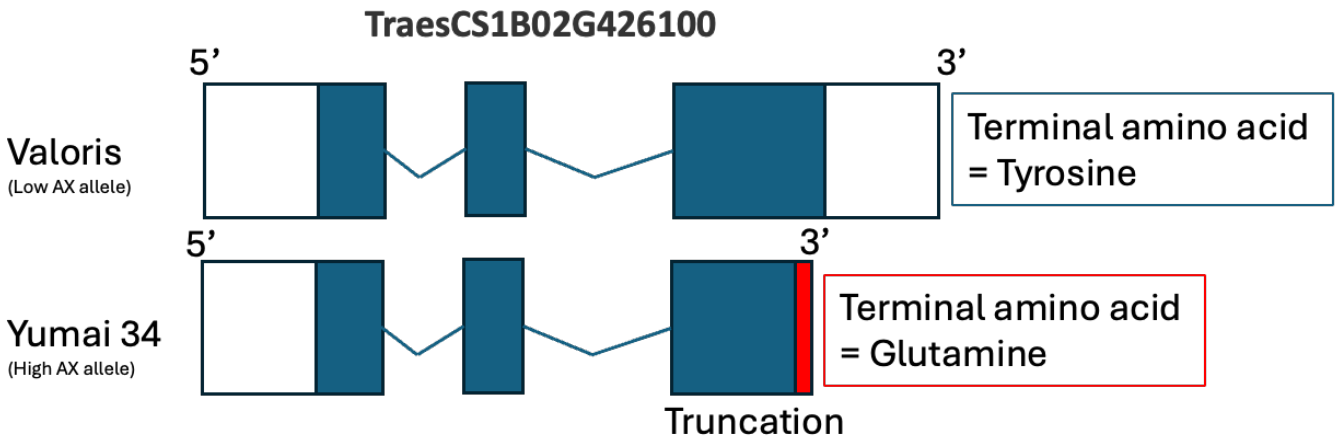


Figure 35: Schematic representation of the transcripts of the purothionin candidate gene. Valoris represents the wild-type isoform, where the protein encodes a tyrosine at the terminal end, whereas the Yumai-34 isoform contains a premature stop codon, which causes the C terminal amino acid to encode a glutamine residue.

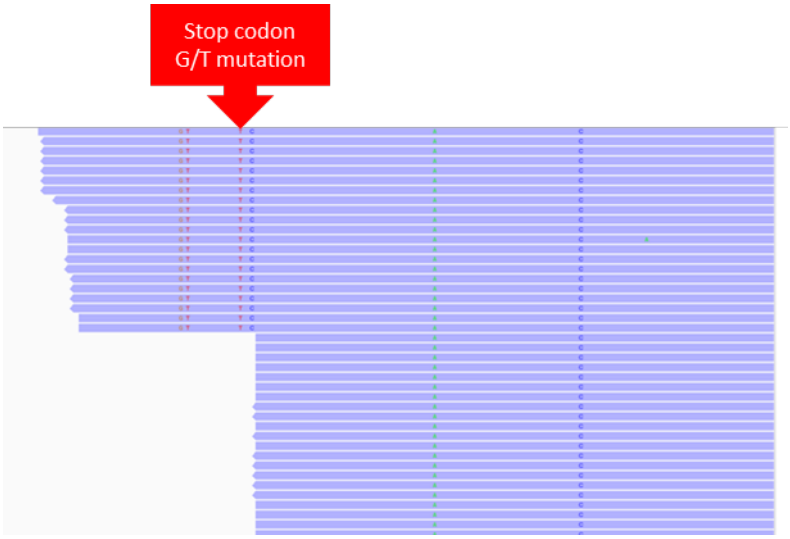


Figure 36: IGV alignments showing the extended RNA reads due to loss of stop codon (Red pointer), through adjacent SNP A/C mutation in Yumai-34. Base pair colours as follows: A- green, C- blue, T -red, G- orange. Truncated RNA reads seen below the extended reads, showing the impact of the stop codon on the transcription of TraesCS1B02G426100.

Purothionins are cell wall proteins identified in wheat endosperm during development of the grain (Cherkaoui et al., 2019). Purothionins are toxic to mammalian cells in *in vitro* assays (Hughes *et al.*, 2000). Their physiological function is unknown, and they have no known role in arabinoxylan synthesis. Their high contents of intra-chain disulphide bonds means that they could have similar functions to thioredoxins, which are present in all organisms and act as regulators of many biological processes including plant immunity through thiol/disulphide exchanges(Mata-Pérez and Spoel, 2019). The deposition of cell wall polysaccharides is an important response to invasion or damage of plants by pests and pathogens and this may be associated with the thioredoxin pathway, as an increase in reactive oxygen and nitric oxide(ROS/RNS) is known to influence cell wall rigidity and metabolism (Delledonne *et al.*, 2001; Torres, Jones and Dangl, 2006; O'Brien *et al.*, 2012). In vitro enzyme assays have shown that purothionins have weak thioredoxin activity in an FBPase assay and act synergistically with thioredoxin-f (Wada and Buchanan, 1981). It is therefore plausible that reduced expression of α-1 purothionin in the high AX allele could lead to increased deposition of AX into endosperm

cell walls. Furthermore, it is possible that the loss of the tyrosine residue in the a-1purothionin protein in the high AX allele in Yumai-34 and Paragon may contribute to an increased deposition of AX through a disruption of cross-linking between the C-terminal tyrosine residue and the ferulic acid residues on the AX chains. Evidence suggests tyrosine-containing peptides oxidatively couple to FA residues via peroxidases (Oudgenoeg *et al.*, 2001) This disruption of cross-linking may therefore allow more sites on the AX chains in the cell wall to be available for an increased deposition of AX, due to a decrease in the number of purothionin peptides bound to the FA residues in the cell wall.

A novel KASP marker (Puro1B) was developed targeting the G->T mutation at the C terminal end and validated at the John Innes Centre on two genetically distinct samples of Yumai-34 (called Yum-1 and Yum-2). Yum-1 was the parental line used to develop the populations of Yumai-34 x Altigo and Yumai-34 x Valoris by INRAE (France) while Yum-2 was used to develop the population of Yumai-34 x Claire at the John Innes Centre. These three populations were analysed by Lovegrove et al., (2020) study which showed that they have the same high AX allele at the 1B QTL. However, they differed when tested with the Puro1B Kasp marker, with Yum-2 of the same genotype as Valoris. This difference between Yum-1 and Yum-2 indicates that the SNP is not causal, and together with the lack of differential expression between Yumai-34 and Valoris, suggests that the gene is not a valid candidate gene for high AX content.

6.2.2.1 Putative candidate genes identified in RNA-seq SNP and Gene Haplotype analysis.

The region in the reduced 1B QTL interval which showed the most consistent differences between the pooled RNA samples with the high and low AX alleles from the Yumai-34 X Valoris data is displayed in Figure 37. This region contains 7 protein-coding genes, 6 of which are expressed in the endosperm and overlap with the genes identified in the gene haplotype analysis shown in Figure 33.

The RNA-seq data show that the E3-RING-ubiquitin ligase, RAS-related protein and the first F-box gene (Loc-123146003 in Figure 36) have the most consistent patterns of variants between groups. The uncharacterised gene model (highlighted in red in Figure 37) and both the second and third F-box genes (XM-044565588 and Loc 12346028 in Figure 37) showed no SNP variants that were common to all of the samples derived from Valoris with the low AX allele. The inconsistency in variant calling in these three genes suggests that the variants are non-causal, as they are not present in all lines with the low AX allele. Furthermore, none of the 7 genes are differentially expressed between theYumai-34 and Valoris, which suggests that any associations with the AX phenotype resulted from a mutation that affects protein structure rather than differences in expression.

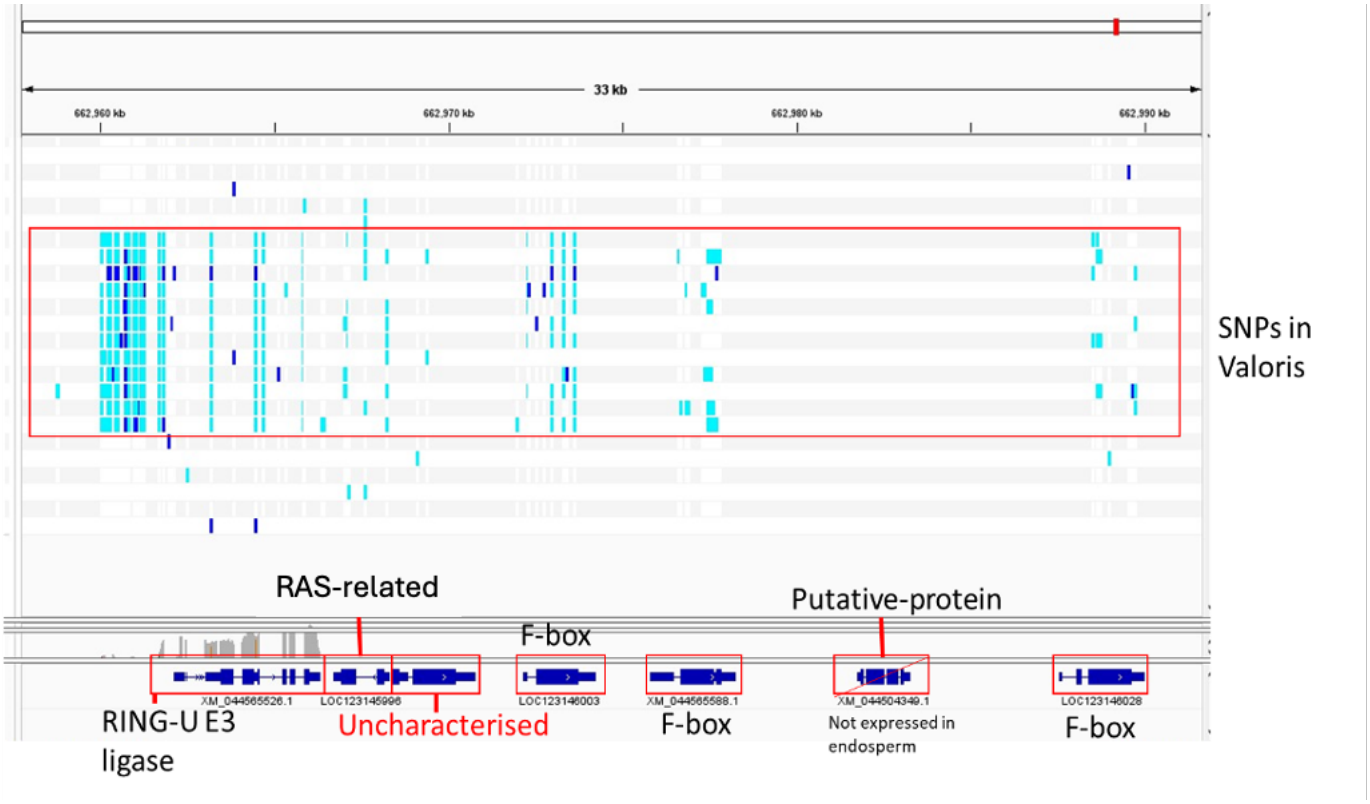


Figure 37: A schematic view of the genes overlapping the first significant region associated with AX content in the bulked segregant Paragon x Watkins Haplotype analysis which also displays significant difference in SNPs between the Yumai-34- and the Valoris-experimental groups (Outlined in the central red box). Light blue and dark blue blocks indicate the presence of a SNP.

In order to analyse each gene further, the variants in each of the Watkins parental accessions used in the bulked segregant analysis were plotted as heatmaps and grouped together depending on the presence of the high or low AX allele. Accessions that contained either the high AX Paragon allele or the low AX allele across the 1B QTL are outlined in the yellow boxes in Figures 38-42. The variant calls were made against the draft Paragon 1B reference sequence, therefore identifying variants common between the high AX alleles in Paragon and the Watkins accessions. Due to the draft nature of the Paragon 1B chromosome assembly, the gene annotations do not include features such as exons or introns and the current SNPs cannot be mapped accurately to these locations. Moreover, it was not possible to determine if a SNP was synonymous or nonsynonymous in these Watkins accessions, though later analysis can determine this in the future.

6.2.2.2 E3-Ring Finger Ubiquitin ligase TraesCS1B02G426300

The first of these genes is displayed in Figure 38 and shows the variants in the E3-Ring Finger Ubiquitin ligase (RING-U). Most of the variants were similar to Paragon, indicated by a blue colour. Six loci showed the alternate allele in the low allelic Watkins accessions, at 662963188bp, 662964578bp,662964706bp, 662965840bp, 66296632bp and 662966422bp. Although this alternative allele is highly represented in the low AX allelic class, it is also present in ~45% of the high AX allelic accessions, which showed a very similar haplotype to the low allele lines. However, two deletions, at 662962743bp and 662964396bp, were present in the low AX allelic class but completely absent from the high AX allelic group, except for WATDE0009, which displays the same deletion at 662964396bp as in the low AX allele group. However, WATDE0009 also contains three other deletion mutations while the remainder of the gene is identical to Paragon. It is possible therefore that additional mutations restore the high AX phenotype.

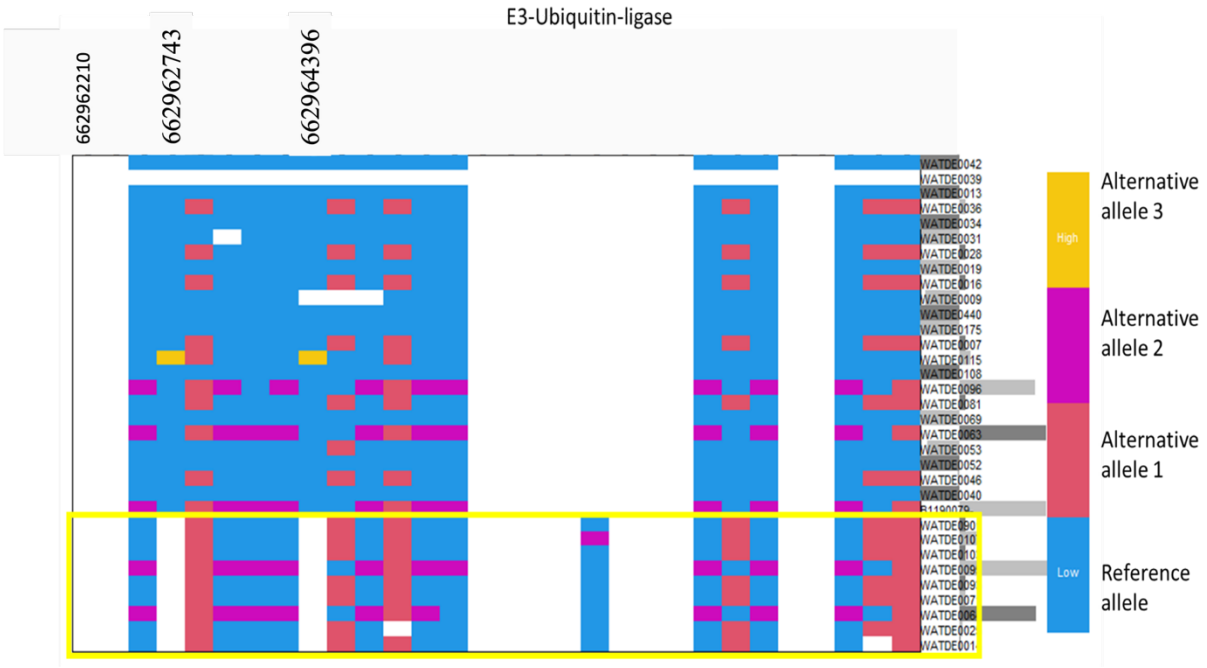


Figure 38: Heatmap of the SNPs in the E3-ubiquitin ligase gene model present in the Watkins accessions from the bulked segregant analysis. Columns indicate the physical position in the chromosome of each SNP and rows relate to the Watkins Accession. Blue colouration indicates no change from Paragon reference sequence and Red, purple and yellow indicate three separate alternative mutations. The histogram across the WATDE row labels indicate higher percentage of allele changes from reference.

6.2.2.3 RAS-Related protein TraesCS1B02G426400

The gene encoding a Rat Sarcoma Virus (RAS)-related protein displayed some differences between the high and low AX allelic groups as shown in Figure 39 where deletions are indicated by the triangles below the variant(s). The first deletion was present only in the low AX allele accessions, whereas the second region of deletions occurred with similar frequencies in the high and low AX groups. Because the high AX allele group contains accessions with a similar set of deletions to those in the low AX lines, it is unlikely to be the causal reason for conferring a low AX phenotype, suggesting that the first deletion is more likely to be the cause by affecting protein function.

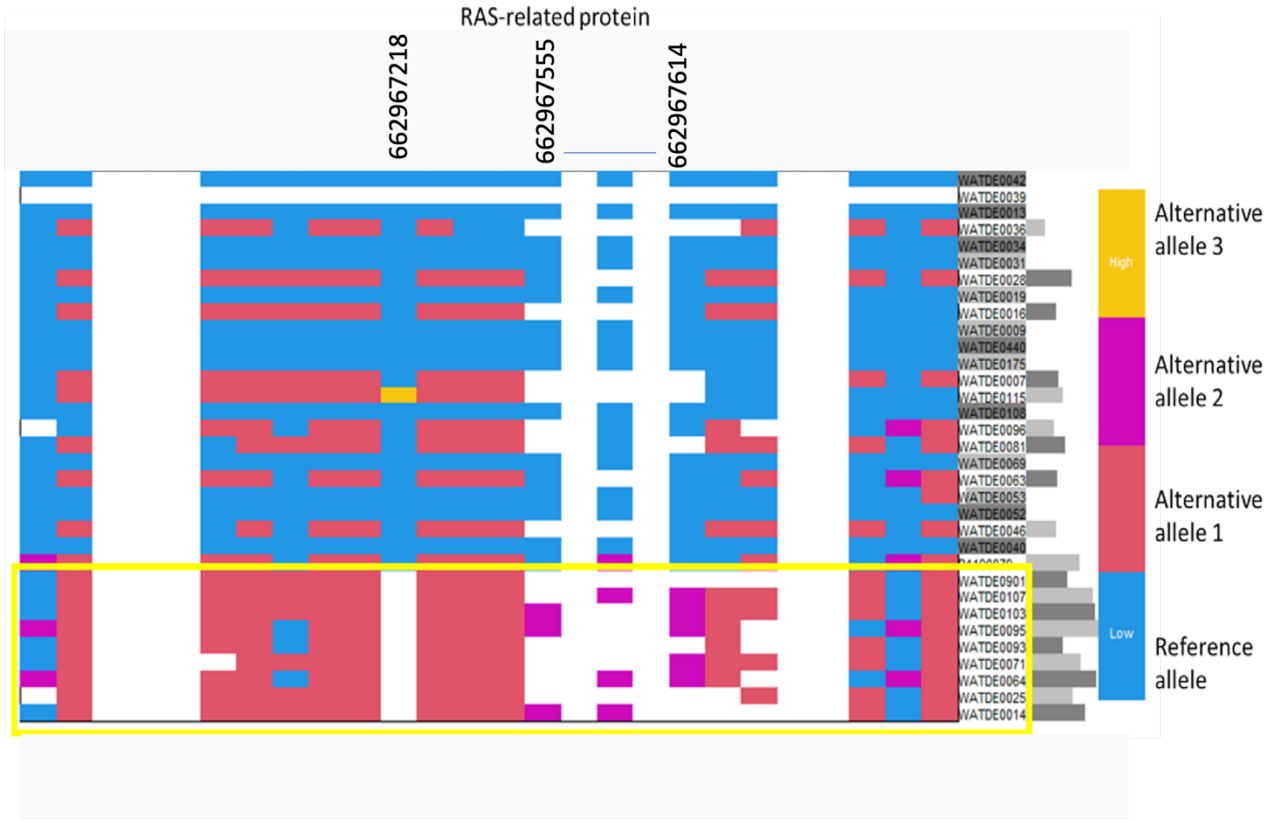


Figure 39: Heatmap of the SNPs in the RAS-related gene model (TraesCS1B02G426400) present in the Watkins accessions from the bulked segregant analysis. Columns indicate the physical position in the chromosome of each SNP and rows relate to the Watkins Accession. Blue colouration indicates no change from Paragon reference sequence and Red, purple and yellow indicate three separate alternative mutations. The histogram across the WATDE row labels indicate higher percentage of allele changes from reference.

6.2.2.4 Uncharacterised Locus TraesCS1B02G426500

The uncharacterised locus, given the gene ID TraesCS1B02G426500 in the Chinese Spring v1.1 reference sequence, displays only one deletion that is common in all the low AX allelic accessions at 662969713bp and not present in any of the high AX allelic accessions (Figure 40). In addition,

the alternative allele at 662970412bp is much more frequent in the low AX allelic group than in the high AX allele group (being present in only 3 of the 24 high AX lines), while the reference Paragon allele is not present in any of the low AX allele accessions. By contrast, the final variant in this gene, is only observed in two of the high allele accessions, but in about half of the low allele lines. The uncharacterised protein encoded by this gene contains an F-box binding domain, suggesting a possible role in the regulation of ubiquitination.

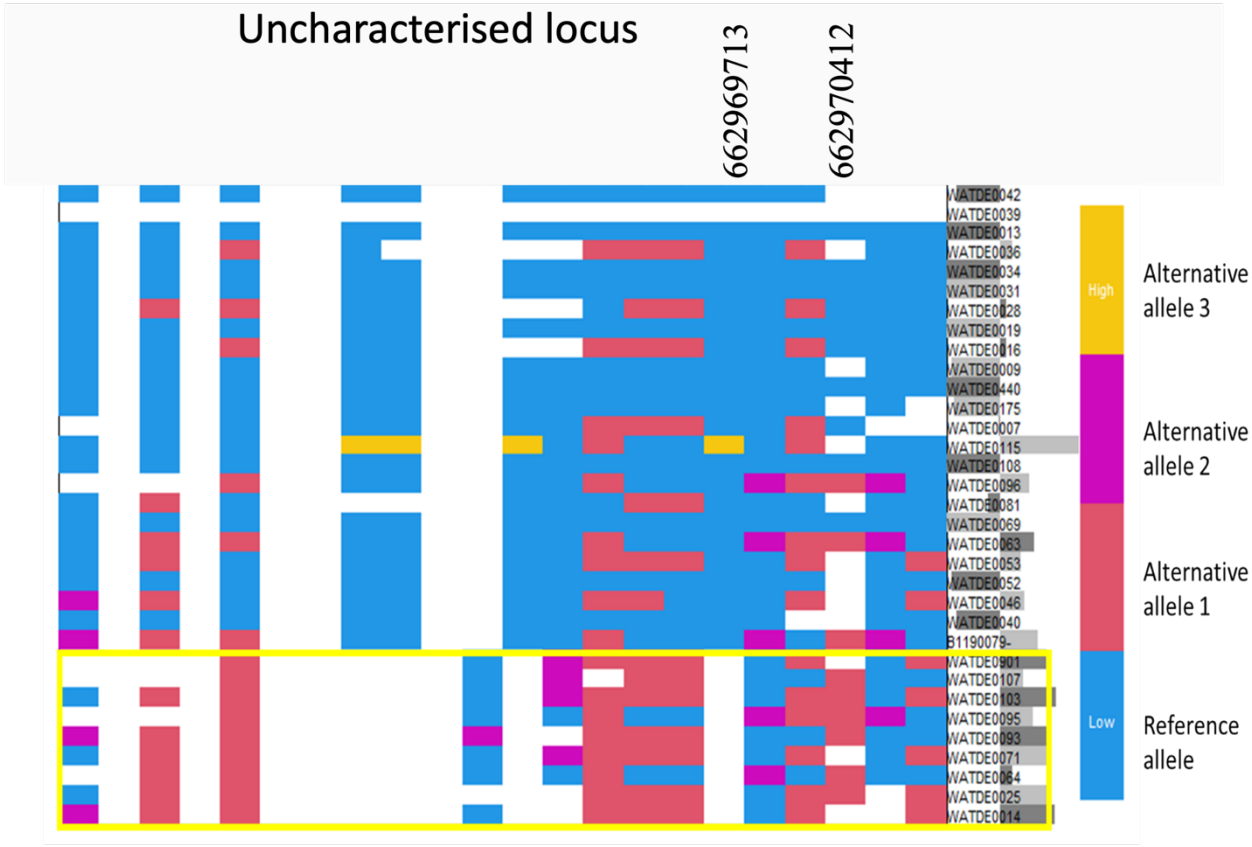


Figure 40: Heatmap of the SNPs in the uncharacterised gene model (TraesCS1B02G426500) present in the Watkins accessions from the bulked segregant analysis. Columns indicate the physical position in the chromosome of each SNP and rows relate to the Watkins Accession. Blue colouration indicates no change from Paragon reference sequence and Red, purple and yellow indicate three separate alternative mutations. The histogram across the WATDE row labels indicate higher percentage of allele changes from reference.

6.2.3 F-box binding protein TraesCS1B02G426500

Variants in the encoding an F-box binding protein are shown in Figure 41. Four deletions are unique to the low AX allele in the Watkins lines, indicated by the numbered inserts (1-4) within the yellow box. The fifth deletion is present in one high AX Watkins line (WATDE0115) while Watkins line WATDE0039, which displays missing data in all the genes present cannot be regarded, due to errors during alignment. No other SNPs are unique to either allele. However, one SNP at 662972303bp that is fixed to this allele class, and the reference allele does not occur, despite the alternative SNP occurring in the high allele class. The available annotated protein domains in the gene TraesCS1B02G426500, from ensemble plants, contains the F-box-like domain superfamily (accession SSF81383) and a domain of unknown function (DUF) 295 (accession PF03478).

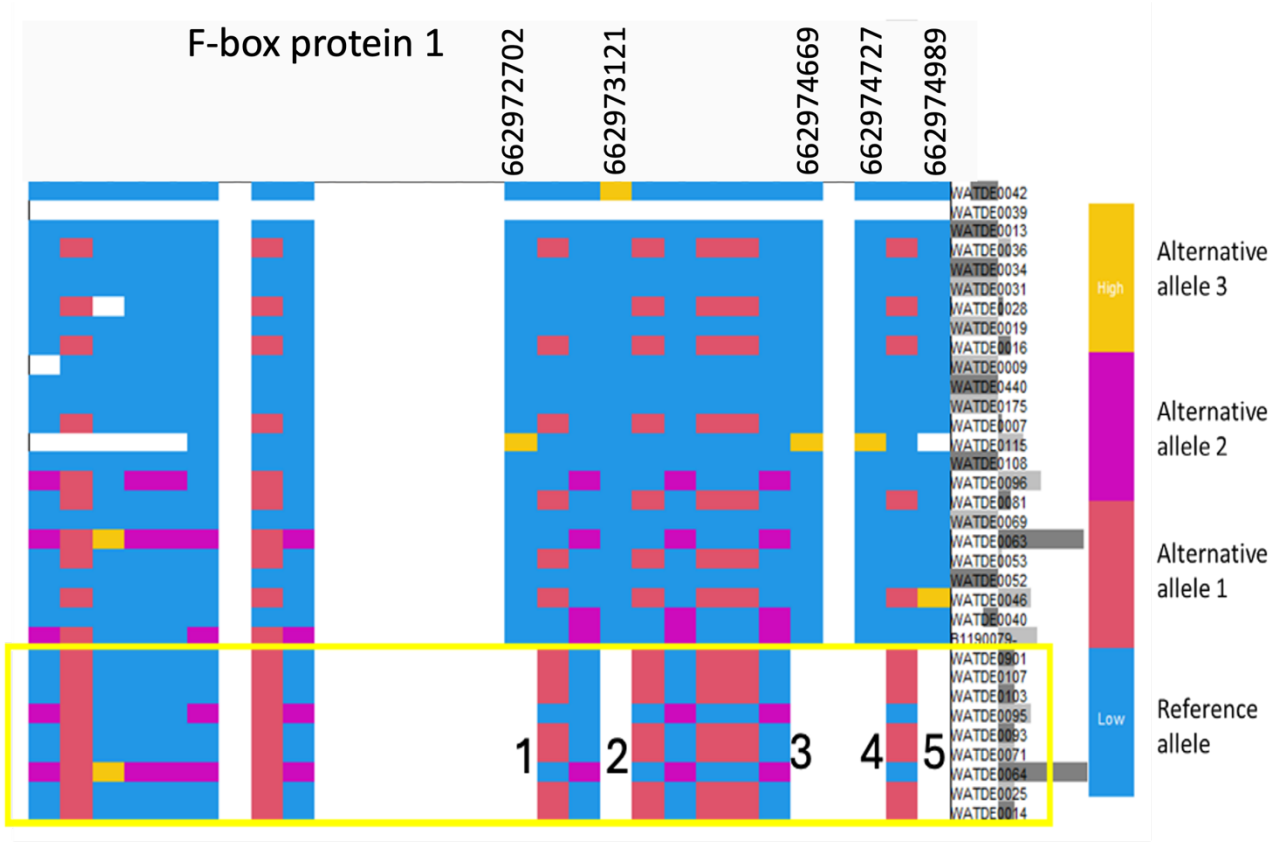


Figure 41: Heatmap of the SNPs in the F-box gene model (TraesCS1B02G426500) present in the Watkins accessions from the bulked segregant analysis. Columns indicate the physical position in the chromosome of each SNP and rows relate to the Watkins Accession.

Blue colouration indicates no change from Paragon reference sequence and Red, purple and yellow indicate three separate alternative mutations. The histogram across the WATDE row labels indicate higher percentage of allele changes from reference.

6.2.3.1 F-box binding protein (second) TraesCS1B02G426600

The second F-box binding protein displays a different set of deletions, numbered 1-3 in Figure 42. These are again unique to the low AX allele class and are not observed in the high AX allele lines. Similarly, none of the SNPs are unique to either low or high AX allele classes. The gene TraesCS1B02G426600 contains the same protein-coding domains as TraesCS1B02G426500, comprising the SSF81383 F-box-like and the DUF295 domains.

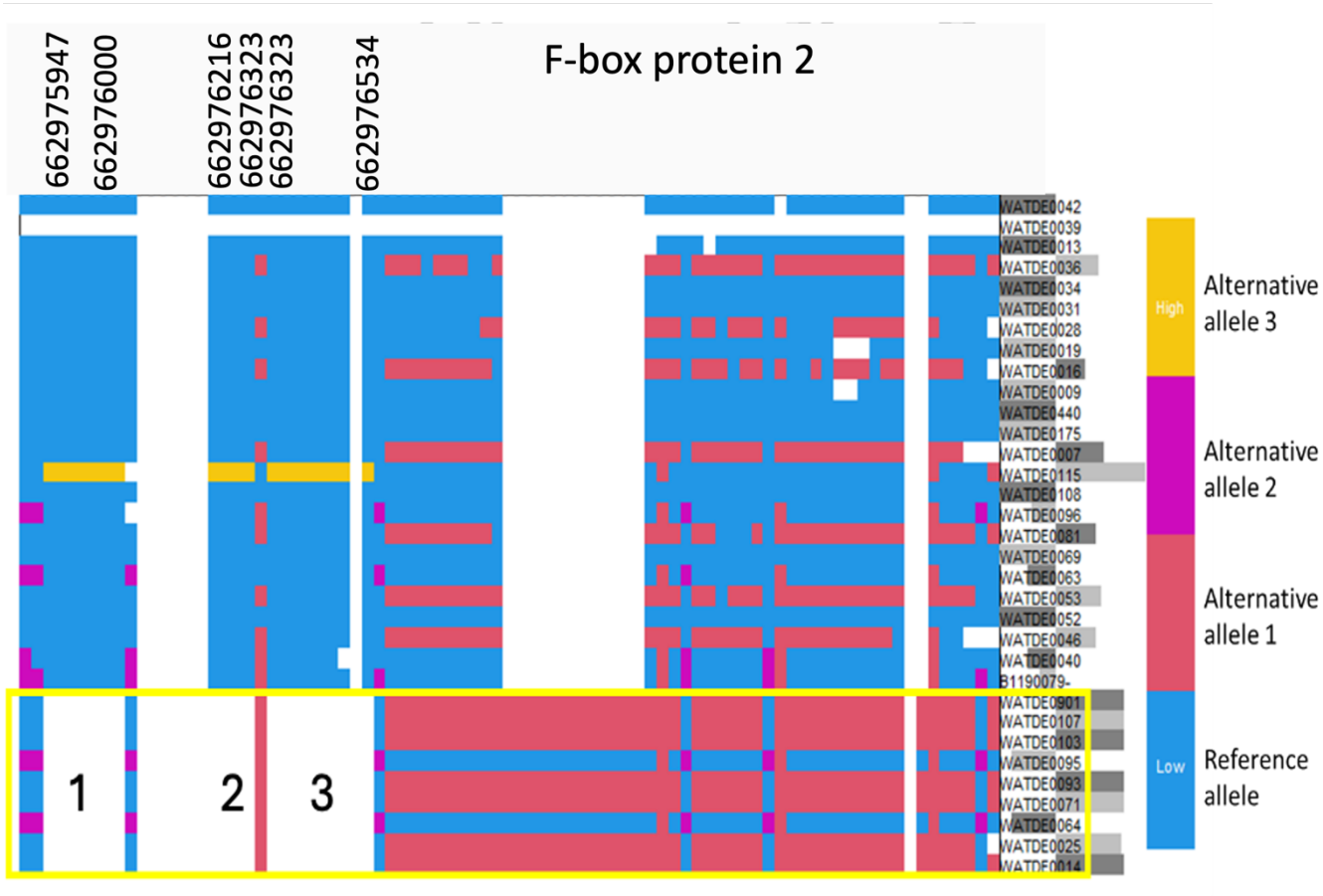


Figure 42: Heatmap of the SNPs in the second F-box gene model (TraesCS1B02G426500) present in the Watkins accessions from the bulked segregant analysis. Columns indicate the physical position in the chromosome of each SNP and rows relate to the Watkins Accession. Blue colouration indicates no change from Paragon reference sequence and Red, purple and yellow indicate three separate alternative mutations. The histogram across the WATDE row labels indicate higher percentage of allele changes from reference.

6.3 Discussion

Initial results suggested that the $\alpha-1$purothionin gene was a likely candidate gene for the 1B QTL that controlled the amount of AX. However, it was not differentially expressed in the high AX (Yumai-34) and low AX (Valoris) alleles. In addition, a KASP marker based on the most significant SNP (which resulted in a change from glutamine to tyrosine in the encoded protein) showed differences between two forms of Yumai-34 which shared the high AX allele. Hence it is unlikely that mutations in the $\alpha-1$ purothionin gene are responsible for the high AX trait on chromosome 1BL.

Of three genes that were upregulated in Yumai34 compared to Valoris. One encoded a type IV thionin, which may have a similar function to purothionins, with a role in cellular defence (Höng *et al.*, 2021). The second encoded phytoene synthase which is one of the first steps in carotenoid synthesis in plants (Zhou *et al.*, 2022), and does not have a known interaction with AX synthesis pathway. The type 5 thionin and the phytoene synthase genes are therefore unlikely candidates for the 1B AX QTL. The third gene is uncharacterised gene and contains DUF4220 and DUF594 domains. This is a more likely candidate as mutants of the rice gene OsSAC1, which contains both DUF4220 and DUF594, had elevated levels of sucrose and glucose in young developing rice leaves when compared to the wild type (Zhu *et al.*, 2018). However, DUF4220 and DUF594 have not been linked to any component of the AX synthesis pathway or shown to affect sugar metabolism in wheat leaves or grain.

However, the remaining candidate genes have many deletions that are unique to the low AX allelic class in the bulked segregant lines from the Watkins crosses. It is not known whether these deletions have effects on protein structure, or which may be responsible for the allelic variation in AX content. The presence of SNP mutations is also not indicative, as all SNPs were usually represented in both the high and low AX allelic classes. However, it is possible that

deletion and/or SNPs could have impacts on the structures and functions of the encoded proteins.

6.3.1.1.1 Biological function of E3-Ring finger Ubiquitin ligase

Proteins of the E3-RING finger ubiquitin ligase family play a vital role in protein degradation, through ubiquitination. This process begins with ubiquitin (Ub) activation, which is dependent on ATP, and mediated by ubiquitin-activating enzyme (E1). This forms a thioester link between the E1 active site cysteine and C-terminus of Ub. Transthiolation then occurs between Ub and ubiquitin-conjugating enzyme (E2), giving rise to the Ub-E2 thioester. The Ub-E2 thioester is then bound to a substrate catalysed by E3- protein ligases, which confer specificity to a substrate (Metzger et al., 2014). 2014). E3 proteins function via two main mechanisms, either acting as intermediates to catalyse ubiquitination, or directly transferring ubiquitin to the substrate. E3 proteins of the Really Interesting New Gene (RING) family, act via the second mechanism, by creating a platform of two Zn^{2+} ions to bind E2 enzymes to the substrate. E3-RING proteins have been implicated in many biological functions and most notably, may play a role in seed development in wheat (Parveen et al., 2021). Parveen et al. (2021) identified a set of 698 RING-E3 proteins expressed at all stages of seed development in wheat, and from this identified a set of 10 RING proteins involved in amylose biosynthesis. Given the diverse roles of E3-Ring proteins, and that they are implicated in a carbohydrate synthetic pathway, it is possible that the RING-U gene is involved in the biosynthesis of xylan, by targeting a substrate directly involved in the biosynthetic pathway.

6.3.1.1.2 Biological function of RAS-related proteins

RAS proteins are small GTP-binding proteins and members of the GTPase family. They are considered to be ‘binary switches’, which are either in an ON or OFF state and affect downstream cell signalling to regulate various biological processes, such as cell proliferation and

growth. The RAS subfamily is absent from plants, indicating that this gene does not encode a true RAS protein, but may share similarities to animal and fungal RAS proteins. In fact, the gene contains RAS and RHO protein domains, suggesting that it is more likely part of the RHO subfamily of the RAS superfamily. RHO proteins are the only small GTPases present in plants (Nagawa, Xu and Yang, 2010), and therefore the only mediators of GTPase-mediated cell signalling. RHO proteins have been associated with actin formation in pollen, with overexpression of RIC1 proteins, which are part of the RHO family, resulting in reduced actin formation. To date, there is no direct evidence for the involvement of RAS-like or RHO proteins in cell wall development in plants.

6.3.1.1.3 Biological function of F-box related proteins

Of the five genes for which SNP analysis was carried out, only the F-box genes may have a direct link to cell wall metabolism, but to secondary cell wall thickening as opposed to primary wall synthesis. As discussed in section 4.3, F-box proteins are involved in the E3 protein ubiquitin degradation pathway, as a constituent of the SKF1-Cullin-Fbox (SCF) E3 ubiquitin ligase complex. A subfamily of the F-box genes, secondary cell wall affecting (SAF1) F-boxes, have been shown to affect secondary cell wall thickening in anthers of Arabidopsis, with mutants in which SAF1 was over-expressed displaying decreased cell wall thickness (Kim *et al.*, 2012).

However, the F-box proteins encoded by the genes identified here contain only a conserved F-box-like domain and the DUF295 domain. The expression of F-box proteins containing a DUF295 domain has previously been shown to be upregulated in pollen tubes of Arabidopsis, with the loss of two key arabinogalactan peptides, AGP6 and AGP11 (Costa *et al.*, 2013). The authors proposed that the loss of these two AGPs resulted in cellular recycling of AGPs to compensate for the loss, though no mechanism was discussed- though AGPs are primarily signalling molecules and not cell wall components. It is possible that diverged F-box proteins containing

DUF295 domains could be responsible for regulating the AX synthesis pathway, with effects on the amount of AX deposited in the cell walls of the developing starchy endosperm.

Chapter 7
Final Discussion and conclusion

7.1 Final discussion

This project was to initially identify new QTL as targets for breeding and provide novel, accurate markers for QTLs associated with a high arabinoxylan content, by exploiting the genetic diversity of the Watkins landrace collection, and to identify novel mechanisms that are involved in the arabinoxylan (AX) synthesis pathways in hexaploid wheat. The core machinery driving AX synthesis has been well characterised, though regulatory mechanisms affecting this core machinery remain unknown.

To address this, a QTL analysis using three populations of recombinant inbred lines, where one parent was always Paragon, and the second- a high AX Watkins accession. This QTL analysis was repeated over three years of field trials, to detect robust QTL that are associated with the phenotyping data on water-extractable and total AX produced from KU Leuven on our field material. From these QTL analyses, only one robust QTL was detected on chromosome 1BL. This QTL, for which Paragon carried the high allele, was a major QTL that had been identified recently by Lovegrove *et al., (*2020), using both a standard QTL analysis on crosses of the cultivars Yumai-34 with Valoris, Altigo, and Claire, and a GWAS of the Whealbi panel. In addition, it was detected by Ibba *et al.,* (2021) during a GWAS study of a bread wheat panel from CYMMT. The QTL that was detected in the Paragon x W471 population in this study, in both 2019 and 2021 field trials overlapped with the physical location of the QTL and associated markers by both Lovegrove *et al.,* and Ibba *et al.* However, neither of these studies produced markers with a high predictive power for the AX trait. Results from Leverington-waite *et al., (unpublished)* showed that the KASP markers produced from Ibba *et al., (2021)* did not all successfully show the high allele across Paragon, from this study and Yumai, or the low allele in Valoris and Claire, which

Lovegrove *et al., (2020)* showed in their QTL analysis. Moreover, neither study have experimentally shown putative candidate genes underlying the 1B locus and were limited to available gene annotations. The effect of the 1B QTL, depending on the year and cultivar, may be up to 32% of the AX content in wheat, and as shown by Lovegrove *et al., (2020),* affects both total and water-extractable AX in white flour. Between 60-70% of bread purchased in the UK is white bread, which lacks the enrichment of AX from the bran, so the 1B QTL is an important target for breeders, in order to achieve new commercial varieties that have increased AX in the endosperm- the component of the grain which makes up white flour.

To address the lack of informative markers for the 1B QTL, this project aimed at characterising the 1B QTL in detail. At the start of this project, sequence data for neither Yumai, Valoris, Paragon or the any of the Watkins accessions was available, so the Wheat 10+ pangenome was utilised to understand the nature of the 1B QTL. The Chinese Spring reference genome (V1.1), which had been used in the previous literature was aligned to each of the pangenome cultivars to ascertain the structural rearrangements that may be occurring in the QTL. As a sub-telomeric loci, these often recombine frequently and can lead to large structural rearrangements, that may have been responsible for affecting the AX trait. However, no major structural changes were detected, though large insertions of non-coding DNA were seen in Chinese spring and Norin. This led to a presence/absence variation (PAV) analysis of protein coding genes to determine if any significant genes were absent from Chinese spring. The Orthologous Matrix (OMA) software was used to identify homologues between cultivars and highlight PAV across the pangenome. Results from AGIS, China, during the WATSEQ project collaboration, which sequenced the Watkins collection in its entirety, determined that of the pangenome, Lancer was most closely related to Paragon during their variant calling pipeline, which led to focussing on the PAV between Chinese spring and Lancer. This identified F-box genes and Armadillo-type fold

proteins that were present in Lancer, and absent in Chinese Spring. From this, it became important to reduce the QTL boundaries and therefore the number of genes to examine. Additional collaboration with the Watseq project allowed access to the haplotype calls for each gene in the QTL, for all of the Watkins collection. By looking at the shared haplotypes of W471 and W694 (Low allele), with W145 (high allele), we could draw a preliminary QTL reduction to further scrutinise the gene in this region. Further to this, the gene-by-gene haplotype analysis was expanded to 34 unique Paragon x Watkins recombinant inbred populations, by bulking related 1B haplotypes within the population and performing a bulked segregant analysis via phenotyping the WE-AX content of each of the bulked samples from each population. This highlighted two specific regions of genes within the first QTL reduction, focusing scrutiny on these two new loci. One of the loci incorporated an alpha-purothionin gene and a RING-ubiquitin E3 ligase (RING-U), which were highly expressed, based on initial RNA-seq from Pellny *et al., unpublished.* The second downstream region contained genes that were not expressed in the Yumai34 x valoris endosperm tissue.

With a new RNA-seq dataseq produced from Oszwald *et al.,* unpublished, from the Yumai34 x Valoris crosses, we could look at this region in detail. Towards the end of the project, early access to the Paragon 1B chromosome assembly became available to the project from Hall *et al., unpublished,* at the Earlham Institute. This allowed the mapping of the Yumai34 x Valoris RNA reads to a reference that shared the high allele with Yumai. To give an insight into which SNPs occurred between Yumai34 and Valoris and were also shared with Paragon. From these results, it was clear that the genes surrounding and including the purothionin and RING-U genes were almost identical between Yumai34 and Paragon, but a significant number of SNPs were seen in Valors. Contrastingly, the second region, highlighted by the bulked segregant haplotype analysis showed no discernible differences between Yumai, Valoris and Paragon, suggesting that the locus containing the purothionin, RING-U and the neighbouring genes may be responsible. This

gave us a final candidate gene list, including the purothionin, RING-U, RAS-related protein, and a series of three F-box genes. To date, there is no evidence for these genes affecting dietary fibre, or cell wall polysaccharides in wheat, however, there is some evidence that another sub-family of F-box genes, SAF1, affect cell wall composition in the pollen tubules (Kim *et al.*, 2012), and DUF295 F-box proteins may be implicated in the recycling of AGP proteins of Arabidopsis (Costa *et al.*, 2013).

This study has identified a small set of putative candidate genes underlying the agronomically important QTL associated with increased water extractable AX in wholemeal flour in Paragon on chromosome 1BL. This same locus has also been associated with increased total and water extractable AX in Yumai34 Lovegrove *et al., (2020*). The putative candidates described in this study are not directly correlated with the core machinery of AX synthesis, but instead link to either vesicle transport, through the RAS-related protein, or the E3 ubiquitin ligase degradation pathways, through the RING-U gene, or the 3 described F-box genes. These gene candidates have not previously been linked to AX synthesis in the literature, and as such, may represent a novel pathway that regulates the core AX synthesis machinery. Furthermore, A KASP marker designed on the RING-U putative candidate is currently the most accurate marker for predicting AX content in wheat and may provide breeders with a better marker for predicting high AX in their breeding programmes.

7.2 Limitations of the Study and Future Work

Although this study has aided in the identification of novel putative candidate genes and the development of more accurate markers, there may have been some limitations in this study. Firstly, the QTL mapping populations were composed of a total number of lines equating to less than 100, which would impact resolution and QTL detection power, which may explain the absence of novel QTL other than 1B. With the more recent data from AGIS, and the Watseq project, which performed short-read WGS sequencing of the Watkins collection, it would be

possible to run a genome-wide association study, with phenotypic data for WE-AX and TOT-AX. Future work on this may look to have the Watkins collection phenotype for these traits, so that a GWAS may be performed, increasing the possibility to detect new marker associations with increased AX.

Secondly, inferences had to be made from the gene expression RNA-seq data of Yumai34 x Valoris and the Paragon parent used in our mapping populations, based on the shared increasing allele for WE-AX at the 1B QTL. Further work, looking to analyse the differences in gene expression in Paragon x Watkins 471 could look to perform endosperm specific RNA-seq at multiple timepoints in the developing grain to ascertain differential gene expression in these two cultivars and make direct comparisons between the Yumai34 x Valoris's dataset.

Following on from this, the Yumai34 and Valoris cultivars do not have whole genome assemblies available, or WGS sequencing reads available, so all sequence inferences were made from the available RNA-seq data. As such, it is impossible to know if genes present in the RNA-seq data, co-localise within the boundaries of the QTL, or if a gene translocation has occurred. Moreover, we do not know the full gene content of Paragon Valoris or Yumai34, as gene annotations for these cultivars were also unavailable. Therefore, any presence/absence variation that may occur in these cultivars could not be determined, and thus reliant on the pangenome and Chinese spring gene annotations. The Paragon genome assembly has recently been released as early access, along with the draft annotation, so further work, could expand on the analysis in this study, to determine gene content differences between Paragon and Chinese spring. Yumai34 and Valoris could be nominated for whole genome sequencing and assembly. This would allow for a direct comparison between these cultivars. As Watkins 471 has been sequenced, but remains unassembled and unannotated, this could also provide a case of further research, to assemble and annotate the W471 genome and compare with Paragon.

In order to advance the confirmation of the candidate, genes identified in this study, Cadenza TILLING lines, containing knockout stop codon mutations within these genes may be selected, grown and backcrossed with Cadenza wildtype. This would allow for the biochemical confirmation of the effect of each candidate gene knockout on the WE-AX phenotype in a Cadenza background, when comparing the mutant lines, with the wild type of control. This could be achieved by backcrossing TILLING knockout mutant lines with the Cadenza wild-type in a randomised glasshouse experiment, for 2 generations, and screening the progeny by designing KASP markers to isolate the TILLING mutant SNP and analysing the progeny for homozygous mutants and homozygous wild-type lines. The homozygous mutants and wild-types would then be assayed for WE-AX via colorimetric determination of pentosans, and then statistically scrutinised to determine if there is a significant difference between the mutants and the wild-types. Furthermore, stacked TILLING lines can be produced, to knockout each homeologue of the candidate genes, to ensure gene redundancy does not mask the effect of the gene on the WE-AX phenotype. If TILLING confirmed one of the 6 candidate genes presented in this thesis, it could highlight a regulatory pathway that affects AX synthesis in wheat, as none of the candidates are genes involved directly in known synthetic pathways of AX. This could provide a novel pathway for AX synthesis, advancing our understanding of the wider regulatory network of cell wall components in bread wheat, and potentially other closely related crop species, such as barley or rice.

Once the candidate gene is confirmed, it may prove beneficial to use gene modification technologies, such as CRISPR/Cas9, to edit an existing commercial wheat cultivar to contain the high 1B AX allele, along with genes already identified on chromosome 6B that increase soluble AX, and perform in vitro digestibility assays on the high AX commercial cultivar, compared to the un-edited cultivar, using model digestive systems, such as the one hosted at the Quadram Institute in Norwich. It is also important to assess the mechanical properties of the grain during

industrial processes, such as milling and baking, to ensure other quality parameters do not fall as a consequence of higher dietary fibre, or effect overall yield.

Appendices

Supplementary Table 1: Independent t-test statistics highlighting the effect on freezing overnight on the relative viscosity of a sample.

Sample	T-statistic	P-value	Adjusted P-value	Significant
145_079	20.3153	0.0003	0.0005	TRUE
145_046	21.4118	0.0001	0.0003	TRUE
Hereward	110.2004	0.0001	0.0003	TRUE
Yumai-34	-11.9077	0.0058	0.0058	TRUE
694_094	22.2042	0.0003	0.0005	TRUE
694_055	10.0152	0.004	0.0046	TRUE
471_015	18.7928	0.0001	0.0003	TRUE
471_064	8.3118	0.004	0.0046	TRUE

Supplementary Table 2: Water extractable AX phenotyping data from KU Leuven from the ParW145 RIL population.

id	WE-Pent.2019_r1	WE-Pent.2019_r2	WE-Pent.2019_mean	WE-Pent.2020
ParW145_001	0.860660963	0.813037683	0.836849323	1.156840953
ParW145_002	0.887957275	0.847691748	0.867824512	1.023096622
ParW145_003		0.835664122	0.835664122	
ParW145_004	0.847749701	0.734524153	0.791136927	1.037622934
ParW145_005				
ParW145_006	0.818639269	0.857904902	0.838272086	1.024525546
ParW145_007	0.841130831	0.866176045	0.853653438	1.033488978
ParW145_008		0.891186954	0.891186954	0.968570831
ParW145_009	0.884939619		0.884939619	1.086751867
ParW145_010				
ParW145_011	0.905270618	0.842360354	0.873815486	0.98475895
ParW145_012	0.984842001	0.912726216	0.948784109	1.168291888
ParW145_013	0.913526811		0.913526811	1.025235738
ParW145_014	0.908055892	0.852767747	0.88041182	0.976522119
ParW145_015	0.820897636	0.7696363	0.795266968	1.010002092
ParW145_016	0.915182899	0.894778626	0.904980763	1.088510574
ParW145_017				
ParW145_018	0.830912512	0.80610687	0.818509691	0.929901653
ParW145_019	0.78322293		0.78322293	0.956458743
ParW145_020	0.965788181	0.902796959	0.93429257	1.159597524
ParW145_021	0.918728363	0.87608699	0.897407677	1.10611239
ParW145_022				
ParW145_023	0.883247048	0.775413422	0.829330235	0.982805063
ParW145_024	0.942477632	0.765426842	0.853952237	1.145814669
ParW145_025				0.949923109

ParW145_026	0.952012121	0.906400939	0.92920653	1.13527893
ParW145_027	0.892295039	0.881343511	0.886819275	1.183618545
ParW145_028	0.954875803	0.890310981	0.922593392	1.033488978
ParW145_029	0.995014493	0.942342701	0.968678597	1.09068325
ParW145_030				
ParW145_031	0.8489214	0.802895289	0.825908345	0.99772494
ParW145_032		0.777681279	0.777681279	1.225755227
ParW145_033				
ParW145_034	0.862321483	0.800732824	0.831527154	1.112199205
ParW145_035	0.964354156	0.956216897	0.960285527	1.261111887
ParW145_036	0.8067913	0.794866336	0.800828818	0.976047093
ParW145_037				
ParW145_038	0.966756609	0.908138953	0.937447781	1.156840953
ParW145_039		0.830680042	0.830680042	1.121523531
ParW145_040	0.903138991	0.933217534	0.918178263	1.184406663
ParW145_041		0.941262513	0.941262513	1.121523531
ParW145_042	0.976075273		0.976075273	1.150407113
ParW145_043	0.83317681	0.85178626	0.842481535	1.038308401
ParW145_044				1.056308054
ParW145_045	0.82698161	0.822229008	0.824605309	1.0889341
ParW145_046	0.869367293	0.925172556	0.897269925	1.132527851
ParW145_047				
ParW145_048	0.746434947	0.812542853	0.7794889	0.942835661
ParW145_049	0.885065306	0.832977099	0.859021203	1.108892822
ParW145_050		0.879680116	0.879680116	1.173794047
ParW145_051	0.800774023	0.724257676	0.76251585	0.995420429
ParW145_052	0.768883273	0.823876761	0.796380017	1.114332878
ParW145_053	0.85213263	0.905526718	0.878829674	1.158175483
ParW145_054	0.851265875	0.880417323	0.865841599	1.065874111
ParW145_055	0.835018585	0.725282078	0.780150332	1.077506255
ParW145_056	0.78369783	0.804497874	0.794097852	1.039671159
ParW145_057	0.831771893	0.828632811	0.830202352	1.199861173
ParW145_058	0.848974746	0.904595359	0.876785053	1.085170108
ParW145_059				1.250041409
ParW145_060				
ParW145_061	0.870491934	0.83131447	0.850903202	1.046492597
ParW145_062	0.873068585	0.860061008	0.866564797	1.133514018
ParW145_063	0.889256593	0.935083969	0.912170281	1.091707277
ParW145_064	0.863897459	0.871539364	0.867718412	1.016982499
ParW145_065		0.939387488	0.939387488	0.97766384
ParW145_066				
ParW145_067	0.770020159	0.814739876	0.792380018	1.089519487
ParW145_068	0.884259474	0.889441974	0.886850724	1.161916719
ParW145_069	0.853136773	0.885861136	0.869498955	1.192676376
ParW145_070				
ParW145_071	0.838194743	0.772721015	0.805457879	1.010742691

ParW145_072	0.807174752	0.841038168	0.82410646	1.105017053
ParW145_073	0.74959632		0.74959632	1.109437121
ParW145_074				1.13803001
ParW145_075	0.921008307	0.917127577	0.919067942	1.1199633
ParW145_076	0.840969481	0.848108431	0.844538956	0.958981021
ParW145_077	0.788513961	0.759723737	0.774118849	1.010002092
ParW145_078		0.759723737	0.759723737	1.027349439
ParW145_079	0.829385125	0.861570469	0.845477797	1.021768975
ParW145_080		0.873282443	0.873282443	1.192676376
ParW145_081	0.914241632	0.903767743	0.909004688	1.205107328
ParW145_082	0.907242451	0.727958395	0.817600423	1.111592715
ParW145_083				
ParW145_084	0.885099202	0.905526718	0.89531296	1.071387253
ParW145_085	0.798951903	0.892976486	0.845964195	1.0341671
ParW145_086	0.753749969	0.838351145	0.796050557	1.001232613
ParW145_087	0.767803632	0.817906172	0.792854902	1.052091256
ParW145_088	0.712009599	0.74639525	0.729202425	0.888662584
ParW145_089	0.851055176	0.885861136	0.868458156	1.047244377
ParW145_090	0.782277406	0.828632811	0.805455109	1.181251308
ParW145_091	0.727863922	0.76695464	0.747409281	1.16596924
ParW145_092				
ParW145_093	0.750461938	0.805040591	0.777751265	0.965808747
ParW145_094	0.741922208	0.81310713	0.777514669	1.103912953

Supplementary Table 3: Raw mean WE and TOT AX phenotyping data from KU Leuven for the ParW471 RIL population

id	WE-Pentosans_2019	WE-Pentosans_2020	WE-Pentosans_2021	TOT-Pentosants_2019	TOT-Pentosans_2020	TOT-Pentosans_2021	RelVisc	Spec_Visc29r1
ParW471_001		NA	NA	NA	NA	NA	NA	NA
ParW471_002	0.846259589	1.093148335	0.67338914	5.64	5.684595667	5.597510184	NA	0.221051
ParW471_003	0.816737182	0.994862194	0.72741647	5.3	NA	5.403141501	2.1398	0.6388
ParW471_004	0.710875335	1.023668256	0.616190494	5.26	NA	5.391142363	1.6023	0.21936
ParW471_005	0.808153525	0.981133645	0.694051083	NA	5.319173351	5.308970591	NA	NA
ParW471_006	0.875093863	0.980348633	0.748741601	5.54	5.008550402	5.209275753	1.8807	0.371137
ParW471_007	0.783977147	0.894353234	0.641309426	5.25	NA	5.135595203	2.0306	0.44036
ParW471_008		NA	NA	NA	NA	NA	NA	NA
ParW471_009	0.751765667	0.89876235	0.670968624	5.33	NA	5.465372249	1.4896	0.200161
ParW471_010	0.7720199	1.052522101	0.678413755	5.74	5.018615592	5.474175583	2.066	0.415485

ParW471_011	0.813012163	0.950085298	0.695423553	NA	NA	5.397975833	NA	NA
ParW471_012	0.902055792	NA	0.77751414	5.67	NA	5.479523267	NA	0.39368
ParW471_013	0.789877562	NA	0.691781252	5.33	NA	5.290908247	2.1642	0.687776
ParW471_014	0.80882375	1.023715849	0.703769019	5.05	NA	5.282243564	NA	0.26943
ParW471_015	0.919857314	1.001614362	0.831789633	5.58	4.910343532	5.242975009	NA	0.386402
ParW471_016	0.876732673	0.992116485	0.729063832	NA	4.913769863	5.265008881	NA	NA
ParW471_017	0.797635501	1.031239653	0.768035007	5.55	5.221317336	5.205949198	NA	0.180082
ParW471_018	0.897653022	1.123162387	0.793654723	5.82	5.451977941	5.237122275	NA	0.3898
ParW471_019	0.793625723	0.991353482	0.667044407	5.59	5.437439334	5.284110479	1.812	0.381492
ParW471_020	0.970958555	1.205948038	0.819701441	6.05	NA	5.785572977	NA	0.310058
ParW471_021	0.774441217	0.940966153	0.681763243	NA	NA	5.655828311	NA	0.619563
ParW471_022	0.842337586	0.910642048	0.71874133	5.72	5.744175452	5.642337169	NA	NA
ParW471_023	0.758042259	0.924370707	0.655442113	5.36	5.116471607	5.338627149	NA	0.409737
ParW471_024	0.793044632	1.103613622	0.714935757	5.47	5.218064696	5.381741103	1.5924	0.223882
ParW471_025	0.832620126	NA	0.679048717	5.27	NA	5.316457588	NA	0.56393
ParW471_026	0.801687056	1.01336318	0.705106278	5.67	5.194376395	5.585405736	NA	0.2147
ParW471_027	0.942034014	1.059267869	0.785166861	5.4	NA	5.388092272	NA	0.394215
ParW471_028	0.756829544	1.002358331	0.659200517	5.23	4.927749293	5.15284841	2.2467	0.48855
ParW471_029	0.823329097	0.946107686	0.675338684	5.97	NA	5.745538765	NA	0.422834
ParW471_030	0.984733096	NA	0.816900421	5.66	NA	5.301477008	3.1307	0.714826
ParW471_031	0.951805349	1.046377727	0.812110705	4.8	NA	5.13893898	NA	0.573886
ParW471_032	0.764611311	1.008590744	0.690745669	5.39	5.174033968	5.404459466	1.8587	0.323742
ParW471_033	0.987967691	0.971290257	0.81039005	5.82	5.009342109	5.623819397	2.4014	0.533545
ParW471_034	0.810922398	1.026424993	0.684816846	5.59	5.314355197	5.440166017	NA	0.26896
ParW471_035	0.826898794	1.020981078	0.706511341	5.29	5.347616648	5.165448144	NA	0.27586
ParW471_036	0.774784992	NA	0.726454543	5.72	NA	5.292494312	1.833	0.232124
ParW471_037	0.82204485	0.989370775	0.6850546	5.39	NA	5.1777066	1.9656	0.346504
ParW471_038	0.771313499	0.922573175	0.702874892	5.7	NA	5.528635928	1.8752	0.341033
ParW471_039	0.801902163	0.929939206	0.689681091	4.85	4.93324606	5.407625212	NA	0.30386

ParW471_040	0.903666108	1.029185391	0.728625148	5.49	NA	5.473285146	NA	0.194565
ParW471_041	0.868476966	0.986625065	0.775748938	5.39	5.441391577	5.460362127	NA	0.385326
ParW471_042	0.80681236	1.021616817	0.716780033	5.64	5.448379276	5.55433461	NA	0.252757
ParW471_043	0.883342731	0.933578024	0.80629942	5.41	NA	5.624686655	NA	0.250988
ParW471_044	0.903658627	1.057382576	0.732666688	5.83	5.364746294	5.594554972	NA	0.283509
ParW471_045	0.91023863	1.005845034	0.804468782	6.07	5.757864493	5.862559117	NA	NA
ParW471_046	0.775115413	1.018865604	0.705942964	5.12	5.465957372	5.248894724	NA	0.171554
ParW471_047	0.982920792	1.072941722	0.842761578	NA	5.038131431	5.618775471	NA	NA
ParW471_048	0.844312553	0.970150806	0.730327921	5.69	5.420218847	5.351462888	NA	NA
ParW471_049	0.956584888	1.034654932	0.830985984	5.34	NA	5.374584814	NA	0.438143
ParW471_050	0.752060033	1.069602864	0.785864171	5.59	5.330822876	5.760421637	NA	0.100688
ParW471_051	0.937252408	1.025759194	0.754609774	5.3	NA	5.308053107	NA	NA
ParW471_052	0.813110581	0.996100888	0.744846408	5.72	5.340419317	5.476478664	1.9032	0.245177
ParW471_053	0.867073314	0.990587415	0.73381779	NA	5.231048116	5.226094485	NA	NA
ParW471_054	0.909473259	1.13584145	0.775976019	5.33	NA	5.875229661	1.8622	0.316956
ParW471_055	0.765427861	0.961090147	0.704496346	5.59	5.584782532	5.732575096	1.777	0.264234
ParW471_056	0.82971656	1.197694401	0.722224599	5.97	NA	5.376271264	1.8686	0.401112
ParW471_057	0.911546724	1.058013521	0.821927954	5.74	5.351592316	5.27775235	1.8012	0.217369
ParW471_058	0.90408653	1.041539262	0.777962043	5	4.910343532	4.922600425	NA	0.35427
ParW471_059	0.807651981	0.935452679	0.687601854	5.24	NA	5.23193415	NA	0.39188
ParW471_060	0.823506731	0.994104694	0.772499802	5.77	NA	5.779791382	1.7262	0.342698
ParW471_061	0.868391913	0.9924143	0.736916546	5.84	NA	5.50736465	1.7458	0.28481
ParW471_062	0.703260484	0.920728029	0.654115011	5.3	4.85589502	5.447058856	NA	NA
ParW471_063	0.887560424	1.130853982	0.827607712	5.51	NA	5.834101468	NA	0.395361
ParW471_064	0.852797037	0.927923058	0.771039458	5.1	NA	5.142434603	2.0568	0.486977
ParW471_065	0.826287582	0.994862194	0.70631928	5.6	NA	5.646958755	2.293	NA
ParW471_066	0.938007076	0.962733919	0.800323858	5.6	NA	5.116092056	NA	0.38525
ParW471_067	0.803378299	0.894228184	0.707578884	5.33	NA	5.488500162	NA	0.23118
ParW471_068	0.782670666	0.95283651	0.714999784	5.02	5.206206237	5.09882398	2.1057	0.46356

ParW471_069	0.723549936	0.870568687	0.614945441	5	5.015819706	5.093175191	NA	0.404216
ParW471_070	0.849288977	1.073289518	0.768842179	5.69	NA	5.386812391	1.7361	0.363436
ParW471_071	0.830879952	0.748747246	0.772261898	5.9	5.008550402	5.337690481	1.605	0.281622
ParW471_072	0.812460898	0.828462679	0.732756208	5.06	5.018615592	5.213943068	NA	0.31583
ParW471_073	0.671144269	1.021680629	0.711260144	NA	NA	5.691169615	NA	NA
ParW471_074	0.739818356	0.936203185	0.766137448	5.24	5.551752701	5.492251325	1.8756	0.239316
ParW471_075	0.88895312	0.951773001	0.780735183	5.11	NA	5.126037572	2.1996	0.597916
ParW471_076		NA	NA	NA	NA	NA	NA	NA
ParW471_077	0.799861483	0.87975527	0.692944614	5.4	NA	5.454243932	NA	0.305176
ParW471_078	0.739680924	NA	0.613992199	5.22	NA	5.300408625	NA	0.255818
ParW471_079	0.863977293	1.05399236	0.745838236	5.49	NA	5.27900627	NA	0.22289
ParW471_080	0.724257426	0.872561887	0.715508557	NA	5.297235333	5.637110355	NA	NA
ParW471_081	0.913456769	0.965474148	0.833883382	5.29	5.046167286	5.534138467	2.2663	0.498698
ParW471_082	0.833782946	0.866108601	0.726403716	5.16	NA	5.129226668	2	0.39667
ParW471_083		0.992275366	NA	5.466056338	NA	NA	1.965975	0.356011773
ParW471_084		NA	NA	NA	NA	NA	NA	NA
ParW471_085		NA	NA	NA	NA	NA	NA	NA
ParW471_086		NA	NA	NA	NA	NA	NA	NA
ParW471_088		NA	NA	NA	NA	NA	NA	NA
ParW471_089		NA	NA	NA	NA	NA	NA	NA
ParW471_090		NA	NA	NA	NA	NA	NA	NA
ParW471_091		NA	NA	NA	NA	NA	NA	NA
ParW471_092		NA	NA	NA	NA	NA	NA	NA
ParW471_093		NA	NA	NA	NA	NA	NA	NA
ParW471_094		NA	NA	NA	NA	NA	NA	NA

Supplementary Table 4: Raw mean WE and TOT AX phenotyping data from KU Leuven for the ParW694 RIL population

id	WE-Pentosans_2019	WE-Pentosans_2020	WE-Pentosans_2021	TOT-Pentosans_2020	TOT-Pentosans_2021
ParW694_001	0.860963425	1.216679467	0.834555218	5.483814133	5.469663786

ParW694_002	1.07	1.333076061	0.872916428	5.377085591	5.872859776
ParW694_003	0.763468944	1.177189098	0.718118752	5.445696797	5.660399461
ParW694_004	0.901845876	1.14105353	0.780506698	5.16584579	5.440962922
ParW694_005	0.770774619	NA	0.777555221	4.82759091	5.807127727
ParW694_006	0.897356819	1.189781289	0.763677667	5.473698379	5.377315696
ParW694_007	0.845601566	1.076967413	0.741121373	4.995912227	5.228974927
ParW694_008	0.838320973	1.151288883	0.823502377	5.151282084	5.429680031
ParW694_009	NA	NA	NA	NA	NA
ParW694_010	0.776737287	0.979737252	0.753395103	4.645486848	4.940044745
ParW694_011	0.97625452	1.110245382	0.808874436	5.384855664	5.660979297
ParW694_012	0.846974946	1.278735761	0.755581478	4.785740678	5.363252898
ParW694_013	0.837903978	1.01358614	0.750442411	5.120172177	5.587948917
ParW694_014	0.732668691	0.991020215	0.734348602	5.15600504	5.151887213
ParW694_015	NA	NA	NA	NA	NA
ParW694_016	0.91	NA	0.882962401	5.481988449	5.624857361
ParW694_017	0.805691074	1.052390991	0.750787673	4.894625352	5.377515643
ParW694_018	0.755431301	0.957838419	0.70175992	4.845679098	5.152929151
ParW694_019	NA	NA	NA	NA	NA
ParW694_020	NA	NA	NA	NA	NA
ParW694_021	0.921011586	1.193438206	0.816868047	5.62917313	5.792926605
ParW694_022	0.88	1.209012983	0.835052864	5.384807744	5.723453715
ParW694_023	0.907147789	1.048872611	0.743527356	4.743117055	5.227307677
ParW694_024	0.867257121	1.217820125	0.851179106	5.497989796	5.306519021
ParW694_025	0.961897289	NA	0.79242352	5.444611255	5.665724591
ParW694_026	0.906403025	1.174368357	0.844903331	5.113004698	5.865148052
ParW694_027	0.864038227	1.064290308	0.72452542	5.241333064	5.465351222
ParW694_028	0.940753446	1.168303392	0.730783023	4.85405159	5.450379002
ParW694_029	0.870116882	1.1781087	0.837792654	4.848799239	5.245641205
ParW694_030	0.836083643	1.086281573	0.769207375	5.092069561	5.75401628
ParW694_031	0.976406872	1.26151867	0.812264622	5.422184938	5.477866957
ParW694_032	0.91554484	1.109491322	0.805795004	4.837979396	4.80858128
ParW694_033	0.897160169	1.153330801	0.912754381	4.858689816	5.725919013
ParW694_034	NA	NA	NA	NA	NA
ParW694_035	0.821444455	0.943067624	0.790601574	5.32372132	5.165601644
ParW694_036	0.893212041	1.132608232	0.798684384	5.592550188	5.319769921
ParW694_037	0.801335039	0.934605402	0.682881081	5.158297063	5.276045223
ParW694_038	0.83	1.114659876	0.707397975	4.976574032	5.314698859
ParW694_039	NA	NA	NA	NA	NA
ParW694_040	0.892838896	NA	0.777453586	4.771821763	5.253824504
ParW694_041	0.921441722	1.012349368	0.816015188	5.067712334	5.109876994
ParW694_042	0.989133353	1.323867612	0.947299993	5.392307613	5.803154355
ParW694_043	0.870997116	1.04044417	0.87644992	4.947535493	5.356921287
ParW694_044	0.757271288	1.093824294	0.805224214	5.484142144	5.542354502
ParW694_045	0.808057428	1.052364716	0.787499813	5.774480711	5.238135717
ParW694_046	1.074525995	1.279760126	0.858005041	5.356494438	5.392185003
ParW694_047	0.926327638	1.165510634	0.856045199	5.094991453	5.875144406

ParW694_048	0.934594149	1.202985711	0.812738692	5.615876859	5.176028282
ParW694_049	0.927740434	1.243707834	0.711400134	5.577826569	5.776356536
ParW694_050	0.805431301	1.075642434	0.790504501	5.334660901	5.275060855
ParW694_051	0.821310194	1.030710597	0.722488182	4.873463819	4.991342629
ParW694_052	0.776819894	1.03689486	0.76043342	4.824293139	5.333325731
ParW694_053	NA	NA	NA	NA	NA
ParW694_054	0.880697293	1.207004083	0.800132646	5.289566798	5.139820821
ParW694_055	0.788166135	NA	0.71025969	5.327383055	5.615548419
ParW694_056	0.767194297	0.944921843	0.712161013	4.640982313	5.409221204
ParW694_057	0.917830861	1.139311383	0.787104596	4.73380196	5.66276843
ParW694_058	0.802820173	1.038371572	0.809639063	4.792318693	5.542428468
ParW694_059	1.03709724	1.228321604	0.854740698	5.512968121	5.728222588
ParW694_060	0.890522289	1.17529922	0.833481609	5.145436913	5.582233073
ParW694_061	0.712058808	1.133703616	0.73336563	4.574514648	5.103144501
ParW694_062	0.815464909	0.997945208	0.765764794	5.484142144	5.673531136
ParW694_063	0.871431165	1.014374067	0.813680548	5.102197947	5.509709152
ParW694_064	0.856040379	1.031458293	0.706568653	5.238211781	5.267619837
ParW694_065	0.838168399	1.115717637	0.894089227	5.068430374	5.204005188
ParW694_066	0.868405866	1.065729492	0.752875537	5.750782432	5.57610041
ParW694_067	NA	NA	NA	NA	NA
ParW694_068	0.818269886	1.135423331	0.741416469	5.726136221	5.233480462
ParW694_069	0.958540542	1.198192939	0.794412784	5.248345651	5.656876608
ParW694_070	0.83	0.985378734	0.6648141	4.917567564	5.029651732
ParW694_071	NA	NA	NA	NA	NA
ParW694_072	NA	NA	NA	NA	NA
ParW694_073	NA	NA	NA	NA	NA
ParW694_074	0.667200329	1.093196844	0.674173099	4.763492716	5.333449808
ParW694_075	NA	NA	NA	NA	NA
ParW694_076	0.881543875	1.080628764	0.79103596	4.917068007	5.475687571
ParW694_077	0.91	1.102860896	0.738912771	5.773451404	5.301198636
ParW694_078	NA	1.214246108	0.791410136	5.455316166	5.52008797
ParW694_079	0.826005849	NA	NA	NA	NA
ParW694_080	0.893200162	1.077625943	0.890228016	5.066165476	5.69393322
ParW694_081	0.732573136	1.001111447	0.719661336	5.201980175	4.843022838
ParW694_082	0.789129676	1.06649674	0.697594367	5.138230418	5.451025705
ParW694_083	NA	1.024869103	0.709102571	4.774193268	5.047125066
ParW694_084	NA	NA	NA	NA	NA
ParW694_085	NA	NA	NA	NA	NA
ParW694_086	NA	NA	NA	NA	NA
ParW694_087	0.942209386	NA	NA	NA	NA
ParW694_088	NA	1.135423331	0.806618533	5.530469607	5.051056449
ParW694_089	NA	NA	NA	NA	NA
ParW694_090	NA	NA	NA	NA	NA
ParW694_091	NA	NA	NA	NA	NA
ParW694_092	NA	NA	NA	NA	NA
ParW694_093	0.877015772	NA	NA	NA	NA

ParW694_094	NA	1.180918181	0.80976403	6.043846532	6.026843573

Supplementary Table 5: The Relative viscosity (RV) and specific viscosity phenotyping data for each of the lines assayed using relative viscometry and HPSEC-mals

Genotype	RV	SV	id	line
ParW145	1.7303	0.359091	6	ParW145_11
ParW145	1.731	0.320351	6	ParW145_51
ParW145	1.7594	0.293216	6	ParW145_48
ParW145	1.7712	0.269365	6	ParW145_31
ParW145	1.7931	0.410181	6	ParW145_73
ParW145	1.8994	0.410937	6	ParW145_88
ParW145	1.9134	0.447112	6	ParW145_77
ParW145	1.9844	0.427411	6	ParW145_86
ParW145	2.081	0.537947	6	ParW145_81
ParW145	2.0884	0.602806	6	ParW145_23
ParW145	2.1427	0.557195	6	ParW145_56
ParW145	2.1875	0.560496	6	ParW145_82
ParW145	2.2368	0.488033	6	ParW145_64
ParW145	2.2423	0.54989	6	ParW145_85
ParW145	2.2539	0.613364	6	ParW145_69
ParW145	2.2623	0.678871	6	ParW145_27
ParW145	2.2716	0.663414	6	ParW145_36
ParW145	2.3467	0.514771	6	ParW145_15
ParW145	2.3859	0.575018	6	ParW145_93
ParW145	2.3899	0.50083	6	ParW145_18
ParW145	2.4764	0.547503	6	ParW145_75
ParW145	2.5024	0.577964	6	ParW145_35
ParW145	2.5078	0.627774	6	ParW145_4
ParW145	2.5587	0.846879	6	ParW145_21
ParW145	2.6228	0.724148	6	ParW145_24
ParW145	2.718	0.815888	6	ParW145_62
ParW145	2.7566	0.871805	6	ParW145_14
ParW145	2.8202	0.674649	6	ParW145_20
ParW145	2.8428	0.778032	6	ParW145_38
ParW145	2.8496	0.8	6	ParW145_28
ParW471	1.4896	0.200161	8	ParW471_009
ParW471	1.5924	0.223882	8	ParW471_024
ParW471	1.6023	0.21936	8	ParW471_004
ParW471	1.605	0.281622	8	ParW471_071
ParW471	1.7262	0.342698	8	ParW471_060
ParW471	1.7361	0.363436	8	ParW471_070
ParW471	1.7458	0.28481	8	ParW471_061
ParW471	1.777	0.264234	8	ParW471_055
ParW471	1.8012	0.217369	8	ParW471_057
ParW471	1.812	0.381492	8	ParW471_019
ParW471	1.833	0.232124	8	ParW471_036
ParW471	1.8587	0.323742	8	ParW471_032
ParW471	1.8622	0.316956	8	ParW471_054
ParW471	1.8686	0.401112	8	ParW471_056

ParW471	1.8752	0.341033	8	ParW471_038
ParW471	1.8756	0.239316	8	ParW471_074
ParW471	1.8807	0.371137	8	ParW471_006
ParW471	1.9032	0.245177	8	ParW471_052
ParW471	1.9656	0.346504	8	ParW471_037
ParW471	2.0306	0.44036	8	ParW471_007
ParW471	2.0568	0.486977	8	ParW471_064
ParW471	2.066	0.415485	8	ParW471_010
ParW471	2.1057	0.46356	8	ParW471_068
ParW471	2.1398	0.6388	8	ParW471_003
ParW471	2.1642	0.687776	8	ParW471_013
ParW471	2.1996	0.597916	8	ParW471_075
ParW471	2.2467	0.48855	8	ParW471_028
ParW471	2.2663	0.498698	8	ParW471_081
ParW471	2.4014	0.533545	8	ParW471_033
ParW471	3.1307	0.714826	8	ParW471_030
ParW694	1.516	0.114283	4	ParW694_002
ParW694	1.6555	0.404535	4	ParW694_006
ParW694	1.6563	0.32201	4	ParW694_005
ParW694	1.6568	0.423881	4	ParW694_061
ParW694	1.7117	0.369142	4	ParW694_052
ParW694	1.7243	0.322675	4	ParW694_044
ParW694	1.7409	0.293629	4	ParW694_017
ParW694	1.761	0.371477	4	ParW694_037
ParW694	1.8256	0.46323	4	ParW694_038
ParW694	1.8802	0.314448	4	ParW694_016
ParW694	1.9227	0.4630394	4	ParW694_042
ParW694	1.9445	0.470435	4	ParW694_029
ParW694	1.988	0.43521	4	ParW694_033
ParW694	2.026	0.41092	4	ParW694_036
ParW694	2.053	0.524005	4	ParW694_041
ParW694	2.0596	0.524444	4	ParW694_004
ParW694	2.1094	0.382377	4	ParW694_022
ParW694	2.1187	0.511576	4	ParW694_081
ParW694	2.1437	0.6984	4	ParW694_046
ParW694	2.1723	0.455301	4	ParW694_027
ParW694	2.2003	0.502144	4	ParW694_047
ParW694	2.2245	0.490061	4	ParW694_023
ParW694	2.2294	0.697146	4	ParW694_050
ParW694	2.2524	0.7736454	4	ParW694_068
ParW694	2.2614	0.4963232	4	ParW694_078
ParW694	2.2919	0.614197	4	ParW694_028
ParW694	2.3176	0.618528	4	ParW694_021
ParW694	2.5408	0.595085	4	ParW694_011
ParW694	2.6041	0.691939	4	ParW694_040
ParW694	2.7527	0.824104	4	ParW694_048
ParW694	2.8133	0.82811	4	ParW694_059
ParW694	3.1316	0.73084	4	ParW694_060

Supplementary Table 6: Full QTL tables for both WE-AX and TOT-AX in the three ParW mapping populations

Population	chr	LOD	%var	mean	add eff	nearest marker	pos nearest marker	Clstart	Clend	start marker	end marker	Pvalue(F)	trait	env	population	increase All
ParW471	2B	3.65	26.333	0.379	-0.059	BS00022734	50.1	19.8	57.4	BS00063694	BS00022966	0.0	sv	ParW471	ParW471	A
ParW471	2B	2.721	16.178	5.466	0.105	BS00022734	50.1	19.8	57.4	BS00063694	BS00022966	0.002	tot	ParW471	ParW471	B
ParW471	1B	6.023	32.338	0.828	-0.04	BS00035268	96.0	83.0	109.9	BS00022539	BS00032697	0.0	we1	ParW471we	ParW471we	A
ParW471	1B	3.6	15.8	0.847	-0.042	BS00035268	96.0	83.0	109.9	BS00022539	BS00032697	0.0004	we2	ParW471we	ParW471we	A
ParW471	3B	3.3	14.0	0.847	-0.041	BS00060014	3.2	0.0	5.9	BS00022961	BS00089954	0.0009	we2	ParW471we	ParW471we	A
ParW471	2B	2.874	16.584	0.992	0.031	BS00022734	50.1	19.8	57.4	BS00063694	BS00022966	0.002	we3	ParW471we	ParW471we	B
ParW694	6A	3.688	39.321	2.108	0.242	BS00022951	7.2	0.0	18.5	BS00023192	BS00105466	0.0	RV	ParW694	ParW694	B
ParW694	2A	2.86	20.957	0.47	-0.039	BS00027847	43.8	15.8	73.3	BS00059111	BS00022995	0.002	SV	ParW694	ParW694	A
ParW694	3B	2.937	17.798	0.869	-0.031	BS00023120	7.8	0.0	16.1	BS00022315	BS00060073	0.002	we1	ParW694	ParW694	A
ParW145	2A	3.034	17.637	0.846	0.021	BS00053834	63.6	16.8	68.9	BS00078489	BS00107316	0.001	we2	ParW145we	ParW145we	B

Supplementary Table 7: WE-AX phenotyping data for the bulked segregant analysis, including absorbance values measure at A552 and A510. P-values from the T-test between the groups A and B within a sample (Watkins accession). Allele shows which group (A= Paragon, B=Watkins) contains the increasing 1B allele

Sample	Group	Rep	Flour quantity (mg)	Volume 1 (uL)	A552	A510	ΔA552-510	Sample AX quantity (mg/g)	Mean	SD	%Diff	Variance	Xylose (X)	Mean Xylose	T_test Pvalu	Significant	Allele
685	A	1	10	1000	0.152	0.072	0.0793						0.00340002				
685	A	2	10	1000	0.185	0.075	0.1101	6.42					0.00481325				
685	A	3	10	1000	0.183	0.075	0.1084	6.31	6.37	0.074	1.6338	1.1553	0.00473525	0.00431617			
685	B	1	10	1000	0.156	0.067	0.089	5.13					0.00384509				
685	B	2	10	1000	0.155	0.067	0.0878	5.05					0.00379003				
685	B	3	10	1000	0.155	0.068	0.0871	5.01	5.06	0.059	2.2956	1.161	0.00375792	0.00379768	0.000199	Sig	Hi:Par
45	A	1	10	1000	0.199	0.083	0.1157	6.76					0.0050702				
45	A	2	10	1000	0.189	0.077	0.1124	6.56					0.00491878				
45	A	3	10	1000	0.207	0.088	0.1191	6.97	6.76	0.205	6.0615	3.0309	0.00522621	0.00507173			
45	B	1	10	1000	0.196	0.078	0.118	6.90					0.00517574				
45	B	2	10	1000	0.217	0.084	0.1325	7.79					0.00584106				
45	B	3	10	1000	0.196	0.078	0.1178	6.89	7.19	0.516	12.504	7.1703	0.00516656	0.00539445	0.25043	NoSig	Hi:WAT
700	A	1	10	1000	0.208	0.086	0.1217	7.13					0.00534551				
700	A	2	10	1000	0.194	0.071	0.1228	7.19					0.00539598				
700	A	3	10	1000	0.207	0.084	0.1228	7.19	7.17	0.039	0.9383	0.5417	0.00539598	0.00537916			
700	B	1	10	1000	0.165	0.067	0.0983	5.70					0.00427182				
700	B	2	10	1000	0.161	0.064	0.0971	5.62					0.00421676				
700	B	3	10	1000	0.167	0.066	0.1005	5.83	5.72	0.105	3.639	1.8455	0.00437276	0.00428711	2.34E-05	Sig	Hi:Par
740	A	1	10	1000	0.2	0.08	0.12	7.02					0.0052675				
740	A	2	10	1000	0.192	0.078	0.1141	6.66					0.00499679				
740	A	3	10	1000	0.191	0.077	0.1139	6.65	6.78	0.212	5.5054	3.1277	0.00498761	0.00508397			
740	B	1	10	1000	0.194	0.094	0.0995	5.77					0.00432688				
740	B	2	10	1000	0.167	0.072	0.0957	5.54					0.00415252				
740	B	3	10	1000	0.171	0.074	0.0974	5.64	5.65	0.116	4.1155	2.0616	0.00423052	0.00423664	0.001269	Sig	Hi:Par
694	A	1	10	1000	0.218	0.083	0.1359	8.00					0.00599706				
694	A	2	10	1000	0.228	0.086	0.1418	8.36					0.00626778				
694	A	3	10	1000	0.221	0.081	0.1397	8.23	8.19	0.183	4.4052	2.2329	0.00617142	0.00614542			
694	B	1	10	1000	0.215	0.084	0.1305	7.67					0.00574929				
694	B	2	10	1000	0.203	0.081	0.1221	7.15					0.00536386				
694	B	3	10	1000	0.207	0.084	0.1234	7.23	7.35	0.277	6.9922	3.7637	0.00542351	0.00551222	0.011607	Sig	Hi:Par
103	A	1	10	1000	0.163	0.064	0.0988	5.73					0.00429476				
103	A	2	10	1000	0.181	0.082	0.0993	5.76					0.0043177				
103	A	3	10	1000	0.169	0.069	0.0993	5.76	5.75	0.018	0.5323	0.3073	0.0043177	0.00431005			
103	B	1	10	1000	0.171	0.069	0.1027	5.96					0.00447371				
103	B	2	10	1000	0.169	0.068	0.1019	5.92					0.004437				
103	B	3	10	1000	0.171	0.067	0.1038	6.03	5.97	0.058	1.9467	0.9774	0.00452418	0.0044783	0.003111	Sig	Hi:WAT
471	A	1	10	1000	0.215	0.082	0.1338	7.87					0.00590071				
471	A	2	10	1000	0.209	0.079	0.1293	7.59					0.00569423				
471	A	3	10	1000	0.213	0.082	0.1307	7.68	7.71	0.141	3.5695	1.8267	0.00575847	0.00578447			
471	B	1	10	1000	0.186	0.074	0.1119	6.53					0.00489584				
471	B	2	10	1000	0.191	0.077	0.1141	6.66					0.00499679				
471	B	3	10	1000	0.186	0.074	0.1118	6.52	6.57	0.08	2.1415	1.2104	0.00489125	0.00492796	0.000257	Sig	Hi:Par
475	A	1	10	1000	0.232	0.108	0.124	7.27					0.00545104				
475	A	2	10	1000	0.228	0.109	0.1183	6.92					0.0051895				
475	A	3	10	1000	0.219	0.098	0.1207	7.07	7.08	0.175	4.9223	2.4713	0.00529962	0.00531339			
475	B	1	10	1000	0.2	0.093	0.1069	6.22					0.00466642				
475	B	2	10	1000	0.195	0.091	0.1041	6.05					0.00453795				
475	B	3	10	1000	0.206	0.096	0.1102	6.42	6.23	0.187	5.9882	2.9975	0.00481784	0.00467407	0.004488	Sig	Hi:Par
79	A	1	10	1000	0.144	0.05	0.0944	5.30					0.00397133				
79	A	2	10	1000	0.135	0.058	0.0774	4.31					0.00323066				
79	A	3	10	1000	0.149	0.053	0.0962	5.40	5.00	0.603	21.839	12.051	0.00404976	0.00375058			
79	B	1	10	1000	0.187	0.076	0.111	6.26					0.00469458				
79	B	2	10	1000	0.176	0.069	0.1076	6.06					0.00454644				
79	B	3	10	1000	0.165	0.066	0.0987	5.54	5.96	0.369	11.998	6.1957	0.00415868	0.00446657	0.110652	NoSig	Hi:WAT
110	A	1	10	1000	0.191	0.062	0.1294	7.33					0.00549625				
110	A	2	10	1000	0.204	0.067	0.1373	7.79					0.00584045				
110	A	3	10	1000	0.186	0.063	0.1229	6.95	7.36	0.419	11.373	5.6954	0.00521305	0.00551659			
110	B	1	10	1000	0.15	0.052	0.0985	5.53					0.00414997				
110	B	2	10	1000	0.167	0.057	0.1101	6.21					0.00465537				
110	B	3	10	1000	0.157	0.07	0.0869	4.86	5.53	0.674	24.357	12.178	0.00364456	0.00414997	0.016435	Sig	Hi:Par
456	A	1	10	1000	0.212	0.076	0.136	7.71					0.00578381				
456	A	2	10	1000	0.202	0.072	0.1294	7.33					0.00549625				
456	A	3	10	1000	0.241	0.103	0.1379	7.82	7.62	0.259	6.4795	3.4008	0.00586659	0.00571555			
456	B	1	10	1000	0.194	0.068	0.1263	7.15					0.00536119				
456	B	2	10	1000	0.207	0.073	0.1335	7.57					0.00567489				
456	B	3	10	1000	0.19	0.066	0.1242	7.03	7.25	0.283	7.4549	3.9098	0.00526969	0.00543526	0.167127	NoSig	Hi:Par
474	A	1	10	1000	0.204	0.067	0.1374	7.79					0.00584481				
474	A	2	10	1000	0.22	0.075	0.145	8.23					0.00617593				
474	A	3	10	1000	0.207	0.068	0.1392	7.90	7.98	0.231	5.536	2.893	0.00592323	0.00598132			
474	B	1	10	1000	0.199	0.07	0.1293	7.32					0.0054919				
474	B	2	10	1000	0.211	0.073	0.1379	7.82					0.00586659				
474	B	3	10	1000	0.226	0.077	0.149	8.47	7.87	0.574	14.541	7.2898	0.00635021	0.0059029	0.784181	NoSig	Hi:Par
273	A	1	10	1000	0.167	0.056	0.1109	6.25					0.00469022				
273	A	2	10	1000	0.164	0.061	0.1031	5.80					0.00435038				
273	A	3	10	1000	0.194	0.09	0.1042	5.86	5.97	0.245	7.5863	4.1061	0.00439831	0.00447964			
273	B	1	10	1000	0.149	0.056	0.0931	5.22					0.00391469				
273	B	2	10	1000	0.171	0.064	0.1071	6.03					0.00452466				
273	B	3	10	1000	0.171	0.064	0.1077	6.07	5.77	0.48	14.691	8.3128	0.0045508	0.00433005	0.556425	NoSig	Hi:Par

398	A	1	10	1000	0.215	0.07	0.1457	8.28					0.00620643			
398	A	2	10	1000	0.18	0.059	0.1206	6.82					0.00511284			
398	A	3	10	1000	0.183	0.059	0.1244	7.04	7.38	0.786	19.766	10.654	0.00527841	0.00553256		
398	B	1	10	1000	0.206	0.069	0.1367	7.75					0.00581431			
398	B	2	10	1000	0.222	0.075	0.1472	8.36					0.00627178			
398	B	3	10	1000	0.198	0.067	0.1312	7.43	7.85	0.472	11.842	6.0164	0.00557468	0.00588692	0.422538	NoSig	Hi:WAT
145	A	1	10	1000	0.137	0.053	0.0841	4.70					0.00352257			
145	A	2	10	1000	0.192	0.069	0.123	6.96					0.00521741			
145	A	3	10	1000	0.175	0.062	0.1137	6.42	6.02	1.18	37.518	19.592	0.00481222	0.0045174		
145	B	1	10	1000	0.163	0.059	0.1042	5.86					0.00439831			
145	B	2	10	1000	0.173	0.066	0.1071	6.03					0.00452466			
145	B	3	10	1000	0.174	0.063	0.1116	6.29	6.06	0.217	7.0893	3.5721	0.00472072	0.0045479	0.956003	NoSig	Hi:WAT
397	A	1	10	1000	0.21	0.07	0.1398	7.93					0.00594937			
397	A	2	10	1000	0.216	0.071	0.1445	8.21					0.00615415			
397	A	3	10	1000	0.219	0.072	0.1473	8.37	8.17	0.22	5.3336	2.6952	0.00627614	0.00612655		
397	B	1	10	1000	0.218	0.074	0.1443	8.19					0.00614543			
397	B	2	10	1000	0.221	0.076	0.1447	8.22					0.00616286			
397	B	3	10	1000	0.211	0.07	0.141	8.00	8.14	0.118	2.6413	1.4496	0.00600166	0.00610332	0.840395	NoSig	Hi:Par
299	A	1	10	1000	0.151	0.039	0.1122	7.71					0.00578282			
299	A	2	10	1000	0.304	0.202	0.1016	6.94					0.00520234			
299	A	3	10	1000	0.302	0.191	0.1107	7.60	7.42	0.419	10.436	5.6477	0.00570067	0.00556194		
299	B	1	10	1000	0.161	0.042	0.1183	8.16					0.00611686			
299	B	2	10	1000	0.169	0.05	0.1189	8.20					0.00614972			
299	B	3	10	1000	0.18	0.063	0.1169	8.05	8.14	0.075	1.7948	0.921	0.00604019	0.00610226	0.041901	Sig	Hi:WAT
238	A	1	10	1000	0.123	0.028	0.0952	6.47					0.00485187			
238	A	2	10	1000	0.139	0.041	0.0978	6.66					0.00499425			
238	A	3	10	1000	0.147	0.05	0.0969	6.59	6.57	0.096	2.8878	1.4665	0.00494496	0.00493036		
238	B	1	10	1000	0.171	0.069	0.1026	7.01					0.00525711			
238	B	2	10	1000	0.146	0.044	0.1016	6.94					0.00520234			
238	B	3	10	1000	0.137	0.045	0.092	6.24	6.73	0.427	11.505	6.3524	0.00467663	0.00504536	0.577073	NoSig	Hi:WAT
300	A	1	10	1000	0.135	0.041	0.094	6.38					0.00478616			
300	A	2	10	1000	0.132	0.044	0.0881	5.95					0.00446306			
300	A	3	10	1000	0.127	0.041	0.0864	5.83	6.05	0.291	9.1677	4.8115	0.00436997	0.00453973		
300	B	1	10	1000	0.121	0.037	0.0842	5.67					0.00424949			
300	B	2	10	1000	0.128	0.04	0.0884	5.97					0.00447949			
300	B	3	10	1000	0.116	0.035	0.0807	5.41	5.68	0.281	9.8929	4.9533	0.00405783	0.00426227	0.188821	NoSig	Hi:Par
560	A	1	10	1000	0.113	0.032	0.0808	5.42					0.0040633			
560	A	2	10	1000	0.103	0.028	0.0742	4.94					0.00370188			
560	A	3	10	1000	0.109	0.036	0.0733	4.87	5.07	0.299	10.791	5.8923	0.00365259	0.00380593		
560	B	1	10	1000	0.082	0.028	0.0543	3.48					0.00261212			
560	B	2	10	1000	0.093	0.028	0.065	4.26					0.00319807			
560	B	3	10	1000	0.111	0.035	0.0754		3.87	0.552	20.17	14.262		0.0029051	0.046575	Sig	Hi:Par
811	A	1	10	1000	0.419	0.316	0.1035	7.08					0.00530639			
811	A	2	10	1000	0.204	0.1	0.1039	7.10					0.0053283			
811	A	3	10	1000	0.135	0.037	0.0972	6.62	6.93	0.274	7.0576	3.9587	0.00496139	0.00519869		
811	B	1	10	1000	0.161	0.058	0.1022	6.98					0.0052352			
811	B	2	10	1000	0.162	0.057	0.1051	7.19					0.00539401			
811	B	3	10	1000	0.148	0.051	0.0968	6.59	6.92	0.308	8.7584	4.4449	0.00493949	0.00518957	0.961668	NoSig	Hi:Par
308	A	1	10	1000	0.144	0.059	0.0847	4.74					0.00355187			
308	A	2	10	1000	0.143	0.045	0.0972									
308	A	3	10	1000	0.123	0.04	0.0828	4.62	4.68	0.083	2.5132	1.7771	0.00346372	0.0035078		
308	B	1	10	1000	0.128	0.045	0.0828									
308	B	2	10	1000	0.138	0.045	0.0933	5.27					0.00395091			
308	B	3	10	1000	0.155	0.049	0.1059	6.05	5.66	0.551	13.778	9.7425	0.00453554	0.00424323	0.130652	NoSig	Hi:WAT
749	A	1	10	1000	0.15	0.047	0.103	7.04					0.00527901			
749	A	2	10	1000	0.17	0.056	0.1145	7.88					0.00590877			
749	A	3	10	1000	0.136	0.044	0.0927		7.46	0.594	11.258	7.9606		0.00559389		
749	B	1	10	1000	0.126	0.035	0.0901	6.10					0.00457259			
749	B	2	10	1000	0.135	0.039	0.0962	6.54					0.00490663			
749	B	3	10	1000	0.125	0.036	0.0889	6.01	6.22	0.286	8.5748	4.5981	0.00450687	0.00466203	0.046353	Sig	Hi:Par
750	A	1	10	1000	0.149	0.055	0.0941	6.39					0.00479163			
750	A	2	10	1000	0.13	0.037	0.0925	6.27					0.00470401			
750	A	3	10	1000	0.135	0.039	0.096	6.53	6.40	0.128	3.9954	2.0002	0.00489568	0.00479711		
750	B	1	10	1000	0.122	0.036	0.0857	5.78					0.00433164			
750	B	2	10	1000	0.138	0.042	0.0955	6.49					0.0048683			
750	B	3	10	1000	0.126	0.04	0.0868	5.86	6.04	0.392	11.845	6.4892	0.00439187	0.0045306	0.209828	NoSig	Hi:Par
49	A	1	10	1000	990	0.062	0.1167	8.04					0.00602924			
49	A	2	10	1000	990	0.066	0.1196	8.25					0.00618805			
49	A	3	10	1000	990	0.064	0.1172	8.08	8.12	0.113	2.6071	1.3937	0.00605662	0.00609131		
49	B	1	10	1000	990	0.086	0.1223	8.45					0.00633591			
49	B	2	10	1000	990	0.071	0.1089	7.47					0.0056021			
49	B	3	10	1000	990	0.08	0.119	8.21	8.04	0.51	12.167	6.3393	0.00615519	0.00603107	0.772953	NoSig	Hi:Par
216	A	1	10	1000	990	0.062	0.116	7.99					0.00599091			
216	A	2	10	1000	990	0.061	0.0996	6.79					0.00509282			
216	A	3	10	1000	990	0.063	0.1196	8.25	7.68	0.778	19.023	10.14	0.00618805	0.00575726		
216	B	1	10	1000	990	0.06	0.1137	7.82					0.00586496			
216	B	2	10	1000	990	0.065	0.1109	7.62					0.00571163			
216	B	3	10	1000	990	0.063	0.1173	8.08	7.84	0.234	5.9609	2.9882	0.0060621	0.00587956	0.745763	NoSig	Hi:WAT
223	A	1	10	1000	990	0.077	0.1023	6.99					0.00524068			
223	A	2	10	1000	990	0.065	0.1142	7.86					0.00589234			
223	A	3	10	1000	990	0.065	0.1167	8.04	7.63	0.562	13.784	7.3654	0.00602924	0.00572075		
223	B	1	10	1000	990	0.069	0.1245	8.61					0.00645638			
223	B	2	10	1000	990	0.069	0.1224	8.46					0.00634138			
223	B	3	10	1000	990	0.066	0.1082	7.42	8.16	0.647	14.584	7.9335	0.00556377	0.00612051	0.3421	NoSig	Hi:WAT
44	A	1	10	1000	990	0.076	0.13	9.01					0.00675757			
44	A	2	10	1000	990	0.072	0.1303	9.03					0.006774			
44	A	3	10	1000	990	0.066	0.1212	8.37	8.80	0.377	7.5477	4.2876	0.00627567	0.00660241		
44	B	1	10	1000	990	0.08	0.1229	8.49					0.00636876			
44	B	2	10	1000	990	0.076	0.1253	8.67					0.00650019			
44	B	3	10	1000	990	0.078	0.1266		8.58	0.124	2.0426	1.4443		0.00643448	0.494675	NoSig	Hi:Par

704	A	1	10	1000	990	0.072	0.1432	9.97					0.00748042			
704	A	2	10	1000	990	0.076	0.1495	10.43					0.00782542			
704	A	3	10	1000	990	0.077	0.1477	10.30	10.24	0.237	4.4936	2.3145	0.00772685	0.00767756		
704	B	1	10	1000	990	0.085	0.1298	9.00					0.00674662			
704	B	2	10	1000	990	0.086	0.1228	8.48					0.00636329			
704	B	3	10	1000	990	0.08	0.126	8.72	8.73	0.256	5.8528	2.93	0.00653852	0.00654948	0.001716	Sig	Hi:Par
254	A	1	10	1000	990	0.073	0.1095	7.51					0.00563496			
254	A	2	10	1000	990	0.065	0.1171	8.07					0.00605115			
254	A	3	10	1000	990	0.067	0.1204		7.79	0.392	7.1228	5.0366		0.00584305		
254	B	1	10	1000	990	0.069	0.1194	8.24					0.0061771			
254	B	2	10	1000	990	0.069	0.1159	7.98					0.00598543			
254	B	3	10	1000	990	0.069	0.1168	8.05	8.09	0.133	3.1598	1.6408	0.00603472	0.00606575	0.285897	NoSig	Hi:WAT

Supplementary table 8: Table of endosperm expressed candidate genes underlying the 1B QTL

gene_ID	functional call
TraesCS1B02G415200	0, At ortho occurs plasmodesmata proteome. Interacts PAL.
TraesCS1B02G415400	0, nothing known
TraesCS1B02G415500	0, RING Ubiquitin ligase
TraesCS1B02G415800	1, At ortho desc Ubi-conjugating enzyme which may occur in Golig among other
TraesCS1B02G416800	0, PFAm domain hydrolase (not polysaccharide)
TraesCS1B02G417200	0, At ortho ribsomal protein
TraesCS1B02G417500	0, nothing known
TraesCS1B02G418600	0, At ortho nuclear transaminase
TraesCS1B02G419300	2, mitgoen activated protein kinase
TraesCS1B02G419800	4, At ortho Golgin candidate 5. Structural Golgi protein that interacts with Rab family of GTPases which may be involved in transport to/from Golgi e.g. CESA to PM.
TraesCS1B02G706700LC	
TraesCS1B02G423300	0, P450
TraesCS1B02G423500	1, similar to dehyrdration proteins of unknown function
TraesCS1B02G424400	0, disease induced
TraesCS1B02G424800	0, At ortho SUMO transferase involced cell cycle

TraesCS1B02G425700	1, Ubiqutin At ortho in golig among others..
TraesCS1B02G426000	0, thionin
TraesCS1B02G426100	1, thionin defense /stoage. There is a version present in CW in barley leaves. Endosperm version in protein bodies?
TraesCS1B02G426200	3, Os ortho Probable E3 ubiquitin-protein ligase ZFP1 implicated salinity stress
TraesCS1B02G426300	6, At ortho RAB GTPase homolog A1G KO affects cell wall composition. Could regulate composition by affecting trafficking of vesicles
TraesCS1B02G426400	0, nothing known
TraesCS1B02G427700	1, top At ortho ribsosomal protein. Some intreactome evidence for physical relation with Golgi proteins
TraesCS1B02G429900	0, nothing known
TraesCS1B02G431200	1, low expr At ortho in chloroplast proteome ; UDP-N-acetylglucosamine biosynthetic process; 16S rRNA processing protein RimM family
TraesCS1B02G431300	4, has DNA-binding domain
TraesCS1B02G729200LC	0
TraesCS1B02G431400	2, MAP kinase-stress signalling and development
TraesCS1B02G431800	0, At ortho RNA splicing, viral response
TraesCS1B02G432200	0, At ortho Eukaryotic translation initiation factor 3
TraesCS1B02G432300	0, has endsulphine domain poss cell cycle
TraesCS1B02G432400	0, At ortho ntegral membrane protein in the inner envelope of chloroplasts. Provide freezing tolerance

TraesCS1B02G434000	1, aspartyl proteinase that can occur in cell wall
TraesCS1B02G435700	5, similar to ERD6 which is proton-driven Glc transporter evidence that is in tonoplast so responsible for Glc export to cytosol under dry conds . Overexpression of sugarbeet ortho in Arabidopsis gives rise to lower Glc and slightly more starch doi: 10.1111/nph.12642. KOs accumulate Glc
TraesCS1B02G436500	0, At ortho mitochondrial protein ISU1
TraesCS1B02G443100	1, At ortho desc A member of RAN GTPase gene family. Encodes a small soluble GTP-binding protein. Likely to be involved in nuclear translocation of proteins. May also be involved in cell cycle progression. Found in CW, NaCl stress and Golgi (Hzlewood) proteomes

Glossary of terms

8

References

Abutair, A.S., Naser, I.A. and Hamed, A.T. (2016) 'Soluble fibers from psyllium improve glycemic response and body weight among diabetes type 2 patients (randomized control trial)', *Nutrition Journal*, 15(1), pp. 1–7. Available at: https://doi.org/10.1186/s12937-016-0207-4.

Alemui, S.K. *et al.* (2021) 'Genome-wide association mapping identifies yellow rust resistance loci in Ethiopian durum wheat germplasm', *PLoS ONE*, 16(5 May 2021), pp. 1–28. Available at: https://doi.org/10.1371/journal.pone.0243675.

Allen, A.M. *et al.* (2017) 'Characterization of a Wheat Breeders' Array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (*Triticum aestivum*)', *Plant Biotechnology Journal*, 15(3), pp. 390–401. Available at: https://doi.org/10.1111/pbi.12635.

Alonso, M.P. *et al.* (2021) 'QTL mapping of spike fertility index in bread wheat', *Crop Breeding and Applied Biotechnology*, 21(1), pp. 1–11. Available at: https://doi.org/10.1590/1984-70332021v21n1a3.
Ames, N. *et al.* (no date) 'A double-blind randomised controlled trial testing the effect of a barley product containing varying amounts and types of fibre on the postprandial glucose response of healthy volunteers'. Available at: https://doi.org/10.1017/S0007114515000367.
Anders, N. *et al.* (2012a) 'Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto xylan in grasses', *Proceedings of the National Academy of Sciences of the United States of America*, 109(3), pp. 989–993. Available at: https://doi.org/10.1073/pnas.1115858109.
Anders, N. *et al.* (2012b) 'Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto xylan in grasses', *Proceedings of the National Academy of Sciences of the United States of America*, 109(3), pp. 989–993. Available at: https://doi.org/10.1073/pnas.1115858109.
Anders, N. *et al.* (2012c) 'Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto xylan in grasses', *Proceedings of the National Academy of Sciences of the United States of America*, 109(3), pp. 989–993. Available at: https://doi.org/10.1073/pnas.1115858109.
Antoine, C. *et al.* (2004) 'Wheat bran tissue fractionation using biochemical markers', *Journal of Cereal Science*, 39(3), pp. 387–393. Available at: https://doi.org/10.1016/J.JCS.2004.02.001.
Balfourier, F. *et al.* (2019) 'Worldwide phylogeography and history of wheat genetic diversity', *Science Advances*, 5(5), p. eaav0536. Available at: https://doi.org/10.1126/sciadv.aav0536.
Barron, C., Surget, A. and Rouau, X. (2007) 'Relative amounts of tissues in mature wheat (Triticum aestivum L.) grain and their carbohydrate and phenolic acid composition', *Journal of Cereal Science*, 45(1), pp. 88–96. Available at: https://doi.org/10.1016/j.jcs.2006.07.004.
Bates, B. *et al.* (2014a) *National Diet and Nutrition Survey: Results from Years 1, 2, 3 and 4 (combined) of the Rolling Programme (2008/2009-2011/2012): A survey carried out on behalf of Public Health England and the Food Standards Agency*. Public Health England.
Bates, B. *et al.* (2014b) *National Diet and Nutrition Survey: Results from Years 1, 2, 3 and 4 (combined) of the Rolling Programme (2008/2009-2011/2012): A survey carried out on behalf of Public Health England and the Food Standards Agency*. Public Health England.
Baydoun, E.A., Waldron, K.W. and Brett, C.T. (1989) 'The interaction of xylosyltransferase and glucuronyltransferase involved in glucuronoxylan synthesis in pea (Pisum sativum) epicotyls.', *The Biochemical journal*, 257(3), pp. 853–858. Available at: https://doi.org/10.1042/bj2570853.
Boros, D. *et al.* (2002) 'Chromosome location of genes controlling the content of dietary fibre and arabinoxylans in rye', *Euphytica*, 128(1), pp. 1–8. Available at: https://doi.org/10.1023/A:1020639601959/METRICS.
Botticella, E., Savatin, D.V. and Sestili, F. (2021) 'The Triple Jags of Dietary Fibers in Cereals: How Biotechnology Is Longing for High Fiber Grains', *Frontiers in Plant Science*. Frontiers Media S.A. Available at: https://doi.org/10.3389/fpls.2021.745579.
De Brier, N. *et al.* (2015) 'The impact of pearling as a treatment prior to wheat roller milling on the texture and structure of bran-rich breakfast flakes', *LWT - Food Science and Technology*, 62(1), pp. 668–674. Available at: https://doi.org/10.1016/j.lwt.2014.08.015.
Brown, D.M. *et al.* (2005) 'Identification of novel genes in Arabidopsis involved in secondary cell wall formation using expression profiling and reverse genetics', *Plant Cell*, 17(8), pp. 2281–2295. Available at: https://doi.org/10.1105/tpc.105.031542.
Bulone, V., Schwerdt, J.G. and Fincher, G.B. (2019) 'Co-evolution of Enzymes Involved in Plant Cell Wall Metabolism in the Grasses', *Frontiers in Plant Science*. Frontiers Media S.A. Available at: https://doi.org/10.3389/fpls.2019.01009.
Cagirici, H.B. *et al.* (2021) 'Multiple variant calling pipelines in wheat whole exome sequencing', *International Journal of Molecular Sciences*, 22(19). Available at: https://doi.org/10.3390/IJMS221910400/S1.
Carpita, N.C. and Gibeaut, D.M. (1993) 'Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth', *The Plant journal : for cell and molecular biology*, 3(1), pp. 1–30. Available at: https://doi.org/10.1111/J.1365-313X.1993.TB00007.X.

De Carvalho, C.M. *et al.* (2017) 'Plasma glucose and insulin responses after consumption of breakfasts with different sources of soluble fiber in type 2 diabetes patients: A randomized crossover clinical trial', *American Journal of Clinical Nutrition*, 106(5), pp. 1238–1245. Available at: https://doi.org/10.3945/ajcn.117.157263.
Charmet, G. *et al.* (2009) 'Genetics of dietary fibre in bread wheat', *Euphytica*, 170(1–2), pp. 155–168. Available at: https://doi.org/10.1007/s10681-009-0019-0.
Chateigner-Boutin, A.L. *et al.* (2018) 'Ferulate and lignin cross-links increase in cell walls of wheat grain outer layers during late development', *Plant Science*, 276(September), pp. 199–207. Available at: https://doi.org/10.1016/j.plantsci.2018.08.022.
Chayut, N. *et al.* (2011) 'A bulk segregant transcriptome analysis reveals metabolic and cellular processes associated with Orange allelic variation and fruit β-carotene accumulation in melon fruit'. Available at: https://doi.org/10.1186/s12870-015-0661-8.
Cheng, S. *et al.* (2023) 'Harnessing Landrace Diversity Empowers Wheat Breeding for Climate Resilience', *bioRxiv*, p. 2023.10.04.560903. Available at: https://doi.org/10.1101/2023.10.04.560903.
Cherkaoui, M. *et al.* (2019) 'Cell Wall Proteome of Wheat Grain Endosperm and Outer Layers at Two Key Stages of Early Development', *International Journal of Molecular Sciences 2020, Vol. 21, Page 239*, 21(1), p. 239. Available at: https://doi.org/10.3390/IJMS21010239.
Chutkan, R. *et al.* (2012) 'Viscous versus nonviscous soluble fiber supplements: mechanisms and evidence for fiber-specific health benefits', *Journal of the American Academy of Nurse Practitioners*, 24(8), pp. 476–487. Available at: https://doi.org/10.1111/J.1745-7599.2012.00758.X.
Costa, M. *et al.* (2013) 'Expression-based and co-localization detection of arabinogalactan protein 6 and arabinogalactan protein 11 interactors in Arabidopsis pollen and pollen tubes', *BMC Plant Biology*, 13(1), p. 7. Available at: https://doi.org/10.1186/1471-2229-13-7.
Darvasi, A. and Soller, M. (1995) 'Advanced Intercross Lines, an Experimental Population for Fine Genetic Mapping', *LANDER and BOTSTEIN* [Preprint].
Delledonne, M. *et al.* (2001) 'Signal interactions between nitric oxide and reactive oxygen intermediates in the plant hypersensitive disease resistance response', *Proceedings of the National Academy of Sciences*, 98(23), pp. 13454–13459. Available at: https://doi.org/10.1073/PNAS.231178298.
Dost, K. and Tokul, O. (2006) 'Determination of phytic acid in wheat and wheat products by reverse phase high performance liquid chromatography', *Analytica Chimica Acta*, 558(1–2), pp. 22–27. Available at: https://doi.org/10.1016/j.aca.2005.11.035.
Douglas, S.G. (1981) 'A rapid method for the determination of pentosans in wheat flour', *Food Chemistry*, 7(2), pp. 139–145. Available at: https://doi.org/10.1016/0308-8146(81)90059-5.
Emily S, C. *et al.* (2013) 'Redefining agricultural yields: from tonnes to people nourished per hectare ', *Environmental Research Letters*, 8. Available at: https://doi.org/10.1088/1748-9326/8/3/034015.
Farre Martinez, A. *et al.* (2021) 'Resolving a QTL complex for height, heading, and grain yield on chromosome 3A in bread wheat', *Journal of Experimental Botany*, 72(8), pp. 2965–2978. Available at: https://doi.org/10.1093/jxb/erab058.
Finnie, S.M., Bettge, A.D. and Morris, C.F. (2006) 'Influence of Cultivar and Environment on Water-Soluble and Water-Insoluble Arabinoxylans in Soft Wheat', *Cereal Chemistry Journal*, 83(6), pp. 617–623. Available at: https://doi.org/10.1094/CC-83-0617.
Freeman, J. *et al.* (2016) 'Effect of suppression of arabinoxylan synthetic genes in wheat endosperm on chain length of arabinoxylan and extract viscosity', *Plant Biotechnology Journal*, 14(1), pp. 109–116. Available at: https://doi.org/10.1111/pbi.12361.
Freeman, J. *et al.* (2017) 'Feruloylation and structure of arabinoxylan in wheat endosperm cell walls from RNAi lines with suppression of genes responsible for backbone synthesis and decoration', *Plant Biotechnology Journal*, 15(11), pp. 1429–1438. Available at: https://doi.org/10.1111/pbi.12727.
Gebruers, K. *et al.* (2008) 'Variation in the content of dietary fiber and components thereof in wheats in the HEALTHGRAIN Diversity Screen', *Journal of agricultural and food chemistry*, 56(21), pp. 9740–9749. Available at: https://doi.org/10.1021/JF800975W.
Gibeaut, D.M. and Carpita, N.C. (1990) 'Separation of membranes by flotation centrifugation for in vitro synthesis of plant cell wall polysaccharides', *Protoplasma*, 156(1–2), pp. 82–93. Available at: https://doi.org/10.1007/BF01666509.

Gupta, R.K., Gangoliya, S.S. and Singh, N.K. (2013) 'Reduction of phytic acid and enhancement of bioavailable micronutrients in food grains', *Journal of Food Science and Technology*. Springer India, pp. 676–684. Available at: https://doi.org/10.1007/s13197-013-0978-y.
Hamant, O. and Traas, J. (2010) 'The mechanics behind plant development', *New Phytologist*, 185(2), pp. 369–385. Available at: https://doi.org/10.1111/J.1469-8137.2009.03100.X.
Harland, J. (2014) 'Authorised EU health claims for barley and oat beta-glucans', in *Foods, Nutrients and Food Ingredients with Authorised EU Health Claims*. Elsevier Ltd, pp. 25–45. Available at: https://doi.org/10.1533/9780857098481.2.25.
Harland, J. (2015) 'Authorised EU health claims for wheat bran fibre', in *Foods, Nutrients and Food Ingredients with Authorised EU Health Claims: Volume 2*. Elsevier, pp. 109–127. Available at: https://doi.org/10.1016/B978-1-78242-382-9.00006-2.
Haskå, L., Nyman, M. and Andersson, R. (2008) 'Distribution and characterisation of fructan in wheat milling fractions', *Journal of Cereal Science*, 48(3), pp. 768–774. Available at: https://doi.org/10.1016/J.JCS.2008.05.002.
Hassan, A.S. *et al.* (2017) 'A Genome Wide Association Study of arabinoxylan content in 2-row spring barley grain', *PloS one*, 12(8). Available at: https://doi.org/10.1371/JOURNAL.PONE.0182537.
Hernández-Espinosa, N. *et al.* (2022) 'Quantification of the total and water-extractable pentosan content in wheat flour using a higher-throughput version of the phloroglucinol colorimetric assay', *Zernodo* [Preprint]. Available at: https://doi.org/10.5281/ZENODO.6824411.
Höng, K. *et al.* (2021) 'The thionin family of antimicrobial peptides', *PLoS One*, 16(7), p. e0254549. Available at: https://doi.org/10.1371/journal.pone.0254549.
Hossain, K. *et al.* (2013) *Interdependence of cultivar and environment on fiber composition in wheat bran*, *525 AJCS*.
Hughes, P. *et al.* (2000) 'The cytotoxic plant protein, β-purothionin, forms ion channels in lipid membranes', *Journal of Biological Chemistry*, 275(2), pp. 823–827. Available at: https://doi.org/10.1074/jbc.275.2.823.
Ibba, M.I. *et al.* (2021) 'Genome-wide association analysis for arabinoxylan content in common wheat (T. aestivum L.) flour', *Journal of Cereal Science*, 98, p. 103166. Available at: https://doi.org/10.1016/j.jcs.2021.103166.
Itoh, N. *et al.* (2019) 'Next-generation sequencing-based bulked segregant analysis for QTL mapping in the heterozygous species Brassica rapa', 132, pp. 2913–2925. Available at: https://doi.org/10.1007/s00122-019-03396-z.
Jones, J.M. (2014) 'CODEX-aligned dietary fiber definitions help to bridge the "fiber gap"', *Nutrition Journal*, 13(1), p. 34. Available at: https://doi.org/10.1186/1475-2891-13-34.
Kaczmarczyk, M.M., Miller, M.J. and Freund, G.G. (2012) 'The health benefits of dietary fiber: Beyond the usual suspects of type 2 diabetes mellitus, cardiovascular disease and colon cancer', *Metabolism: Clinical and Experimental*, pp. 1058–1066. Available at: https://doi.org/10.1016/j.metabol.2012.01.017.
Keegstra, K. (2010) 'Plant cell walls', *Plant Physiology*, 154(2), pp. 483–486. Available at: https://doi.org/10.1104/pp.110.161240.
Kelleher, J. *et al.* (1984) 'Degradation of cellulose within the gastrointestinal tract in man', *Gut*, 25(8), pp. 811–815. Available at: https://doi.org/10.1136/gut.25.8.811.
Kellow, N.J. and Walker, K.Z. (2018) 'Authorised EU health claim for arabinoxylan', in *Foods, Nutrients and Food Ingredients with Authorised EU Health Claims*. Elsevier Inc., pp. 201–218. Available at: https://doi.org/10.1016/B978-0-08-100922-2.00013-9.
Kim, D., Langmead, B. and Salzberg, S.L. (2015) 'HISAT: a fast spliced aligner with low memory requirements', *Nature methods*, 12(4), pp. 357–360. Available at: https://doi.org/10.1038/NMETH.3317.
Kim, Y.Y. *et al.* (2012) 'A novel F-box protein represses endothecial secondary wall thickening for anther dehiscence in Arabidopsis thaliana', *Journal of plant physiology*, 169(2), pp. 212–216. Available at: https://doi.org/10.1016/J.JPLPH.2011.09.006.
Kiszonas, A.M., Courtin, C.M. and Morris, C.F. (2012) 'A Critical Assessment of the Quantification of Wheat Grain Arabinoxylans Using a Phloroglucinol Colorimetric Assay', *Cereal Chemistry Journal*, 89(3), pp. 143–150. Available at: https://doi.org/10.1094/CCHEM-02-12-0016-R.

Kozlova, L. V. *et al.* (2022) 'Identification of genes involved in the formation of soluble dietary fiber in winter rye grain and their expression in cultivars with different viscosities of wholemeal water extract', *The Crop Journal*, 10(2), pp. 532–549. Available at: https://doi.org/10.1016/J.CJ.2021.05.008.
de la Fuente Cantó, C. and Vigouroux, Y. (2022) 'Evaluation of nine statistics to identify QTLs in bulk segregant analysis using next generation sequencing approaches', *BMC Genomics*, 23(1), pp. 1–12. Available at: https://doi.org/10.1186/S12864-022-08718-Y/TABLES/1.
Lafiandra, D. and Shewry, P. (2014) 'Improving Cereal Grain Carbohydrates for Diet and Health', *Journal of Cereal Science*, 59. Available at: https://doi.org/10.1016/j.jcs.2014.01.001.
Laperche, A. *et al.* (2007) 'Using genotype x nitrogen interaction variables to evaluate the QTL involved in wheat tolerance to nitrogen constraints', *Theoretical and Applied Genetics*, 115(3), pp. 399–415. Available at: https://doi.org/10.1007/s00122-007-0575-4.
Li, Q. *et al.* (2020) 'Quantitative Trait Locus (QTLs) Mapping for Quality Traits of Wheat Based on High Density Genetic Map Combined With Bulked Segregant Analysis RNA-seq (BSR-Seq) Indicates That the Basic 7S Globulin Gene Is Related to Falling Number', *Frontiers in Plant Science*, 11, p. 600788. Available at: https://doi.org/10.3389/FPLS.2020.600788/BIBTEX.
Loosveld, A.M.A., Grobet, P.J. and Delcour, J.A. (1997) 'Contents and Structural Features of Water-Extractable Arabinogalactan in Wheat Flour Fractions', *Journal of Agricultural and Food Chemistry*, 45(6), pp. 1998–2002. Available at: https://doi.org/10.1021/jf960901k.
Love, M.I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), pp. 1–21. Available at: https://doi.org/10.1186/S13059-014-0550-8/FIGURES/9.
Lovegrove, A. *et al.* (2013) 'RNA interference suppression of genes in glycosyl transferase families 43 and 47 in wheat starchy endosperm causes large decreases in arabinoxylan content', *Plant Physiology*, 163(1), pp. 95–107. Available at: https://doi.org/10.1104/pp.113.222653.
Lovegrove, A., Wingen, L.U., Plummer, A., Wood, A., Passmore, D., Kosik, O., Freeman, J., Mitchell, Rowan A. C., *et al.* (2020) 'Identification of a major QTL and associated molecular marker for high arabinoxylan fibre in white wheat flour', *PLOS ONE*. Edited by A. Zhang, 15(2), p. e0227826. Available at: https://doi.org/10.1371/journal.pone.0227826.
Lovegrove, A., Wingen, L.U., Plummer, A., Wood, A., Passmore, D., Kosik, O., Freeman, J., Mitchell, Rowan A.C., *et al.* (2020a) 'Identification of a major QTL and associated molecular marker for high arabinoxylan fibre in white wheat flour', *PLoS ONE*, 15(2), pp. 1–13. Available at: https://doi.org/10.1371/journal.pone.0227826.
Lovegrove, A., Wingen, L.U., Plummer, A., Wood, A., Passmore, D., Kosik, O., Freeman, J., Mitchell, Rowan A.C., *et al.* (2020b) 'Identification of a major QTL and associated molecular marker for high arabinoxylan fibre in white wheat flour', *PLoS ONE*, 15(2), pp. 1–13. Available at: https://doi.org/10.1371/journal.pone.0227826.
Maes, C. and Delcour, J.A. (2002) 'Structural characterisation of water-extractable and water-unextractable arabinoxylans in wheat bran', *Journal of Cereal Science*, 35(3), pp. 315–326. Available at: https://doi.org/10.1006/jcrs.2001.0439.
Mata-Pérez, C. and Spoel, S.H. (2019) 'Thioredoxin-mediated redox signalling in plant immunity', *Plant Science*, 279, pp. 27–33. Available at: https://doi.org/10.1016/J.PLANTSCI.2018.05.001.
McRae, M.P. (2017) 'Health Benefits of Dietary Whole Grains: An Umbrella Review of Meta-analyses', *Journal of Chiropractic Medicine*. Elsevier USA, pp. 10–18. Available at: https://doi.org/10.1016/j.jcm.2016.08.008.
Metzger, M.B. *et al.* (no date) 'RING-type E3 ligases: Master manipulators of E2 ubiquitin-conjugating enzymes and ubiquitination'. Available at: https://doi.org/10.1016/j.bbamcr.2013.05.026.
Michelmore, R.W., Paran, I. and Kesseli, R. V. (1991) 'Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations', *Proceedings of the National Academy of Sciences of the United States of America*, 88(21), pp. 9828–9832. Available at: https://doi.org/10.1073/PNAS.88.21.9828.
Miller, T. *et al.* (2001) 'The Watkins collection of landrace derived wheats', *Wheat taxon legacy of John Percival. London: Linnean Society*, pp. 113–120.

Mio, K. *et al.* (2020) 'Effects of β-glucan Rich Barley Flour on Glucose and Lipid Metabolism in the Ileum, Liver, and Adipose Tissues of High-Fat Diet Induced-Obesity Model Male Mice Analyzed by DNA Microarray', *Nutrients*, 12(11), pp. 1–18. Available at: https://doi.org/10.3390/NU12113546.
Mio, K. *et al.* (2023) 'A single administration of barley β-glucan and arabinoxylan extracts reduces blood glucose levels at the second meal via intestinal fermentation', 87(1), pp. 99–107. Available at: https://doi.org/10.1093/bbb/zbac171.
Mitchell, R.A.C., Dupree, P. and Shewry, P.R. (2007) 'A Novel Bioinformatics Approach Identifies Candidate Genes for the Synthesis and Feruloylation of Arabinoxylan', *Plant Physiology*, 144(1), pp. 43 LP – 53. Available at: https://doi.org/10.1104/pp.106.094995.
Nagawa, S., Xu, T. and Yang, Z. (2010) 'RHO GTPase in plants: Conservation and invention of regulators and effectors', *Small GTPases*, 1(2), pp. 78–88. Available at: https://doi.org/10.4161/SGTP.1.2.14544.
Nauck, M.A. *et al.* (2017) 'Cardiovascular Actions and Clinical Outcomes With Glucagon-Like Peptide-1 Receptor Agonists and Dipeptidyl Peptidase-4 Inhibitors', *Circulation*, 136(9), pp. 849–870. Available at: https://doi.org/10.1161/CIRCULATIONAHA.117.028136.
O'Brien, J.A. *et al.* (2012) 'Reactive oxygen species and their role in plant defence and cell wall metabolism', *Planta*, 236(3), pp. 765–779. Available at: https://doi.org/10.1007/S00425-012-1696-9/FIGURES/3.
Oudgenoeg, G. *et al.* (2001) 'Peroxidase-mediated cross-linking of a tyrosine-containing peptide with ferulic acid', *Journal of agricultural and food chemistry*, 49(5), pp. 2503–2510. Available at: https://doi.org/10.1021/JF000906O.
Parveen, A. *et al.* (123AD) 'Genome-wide analysis of RING-type E3 ligase family identifies potential candidates regulating high amylose starch biosynthesis in wheat (Triticum aestivum L.)', *Scientific Reports* |, 11, p. 11461. Available at: https://doi.org/10.1038/s41598-021-90685-7.
Perretant, M. *et al.* (no date) *QTL analysis of bread-making quality in wheat using a doubled haploid population*, *Springer*.
Piro, M.C., Muylle, H. and Haesaert, G. (2023) 'Exploiting Rye in Wheat Quality Breeding: The Case of Arabinoxylan Content', *Plants*, 12(4), p. 737. Available at: https://doi.org/10.3390/PLANTS12040737.
Quraishi, U.M. *et al.* (2009) 'Genomics in cereals: From genome-wide conserved orthologous set (COS) sequences to candidate genes for trait dissection', *Functional and Integrative Genomics*, 9(4), pp. 473–484. Available at: https://doi.org/10.1007/s10142-009-0129-8.
Quraishi, U.M. *et al.* (2011) 'Combined meta-genomics analyses unravel candidate genes for the grain dietary fiber content in bread wheat (Triticum aestivum L.)', *Functional and Integrative Genomics*, 11(1), pp. 71–83. Available at: https://doi.org/10.1007/s10142-010-0183-2.
Ren, Y. *et al.* (2014) 'Site-directed mutagenesis of IRX9, IRX9L and IRX14 proteins involved in xylan biosynthesis: Glycosyltransferase activity is not required for IRX9 function in arabidopsis', *PLoS ONE*, 9(8), p. e105014. Available at: https://doi.org/10.1371/journal.pone.0105014.
Rennie, E.A. and Scheller, H.V. (2014) 'Xylan biosynthesis', *Current Opinion in Biotechnology*, pp. 100–107. Available at: https://doi.org/10.1016/j.copbio.2013.11.013.
Saulnier, L., Vigouroux, J. and Thibault, J.F. (1995) 'Isolation and partial characterization of feruloylated oligosaccharides from maize bran', *Carbohydrate Research*, 272(2), pp. 241–253. Available at: https://doi.org/10.1016/0008-6215(95)00053-V.
Shen, X. *et al.* (2003) 'Detection of Fusarium head blight resistance QTL in a wheat population using bulked segregant analysis', *Theoretical and Applied Genetics*, 106(6), pp. 1041–1047. Available at: https://doi.org/10.1007/S00122-002-1133-8/METRICS.
Shewry, P.R. *et al.* (2011) 'Relationship between the contents of bioactive components in grain and the release dates of wheat lines in the HEALTHGRAIN diversity screen', *Journal of Agricultural and Food Chemistry*, 59(3), pp. 928–933. Available at: https://doi.org/10.1021/JF103860X.
Shewry, P.R. and Hey, S.J. (2015) 'The contribution of wheat to human diet and health', *Food and Energy Security*. Wiley-Blackwell Publishing Ltd, pp. 178–202. Available at: https://doi.org/10.1002/FES3.64.
Shewry, P.R., Pellny, T.K. and Lovegrove, A. (2016) 'Is modern wheat bad for health?', *Nature Plants*, 2(7), pp. 1–3. Available at: https://doi.org/10.1038/NPLANTS.2016.97.

Speck, T. and Burgert, I. (2011) 'Plant Stems: Functional Design and Mechanics', *https://doi.org/10.1146/annurev-matsci-062910-100425*, 41, pp. 169–193. Available at: https://doi.org/10.1146/ANNUREV-MATSCI-062910-100425.
Strohmeier, M. *et al.* (2009) 'Molecular modeling of family GH16 glycoside hydrolases: Potential roles for xyloglucan transglucosylases/hydrolases in cell wall modification in the poaceae', *Protein Science*, 13(12), pp. 3200–3213. Available at: https://doi.org/10.1110/ps.04828404.
Stuper-Szablewska, K. *et al.* (2019) 'Quantitative profile of phenolic acids and antioxidant activity of wheat grain exposed to stress', *European Food Research and Technology*, 245(8), pp. 1595–1603. Available at: https://doi.org/10.1007/s00217-019-03262-8.
Suliman, M. *et al.* (2013) 'Identification of glycosyltransferases involved in cell wall synthesis of wheat endosperm', *Journal of Proteomics*, 78, pp. 508–521. Available at: https://doi.org/10.1016/j.jprot.2012.10.021.
Thondre, P.S., Shafat, A. and Clegg, M.E. (2013) 'Molecular weight of barley β-glucan influences energy expenditure, gastric emptying and glycaemic response in human subjects', *The British journal of nutrition*, 110(12), pp. 2173–2179. Available at: https://doi.org/10.1017/S0007114513001682.
Torres, M.A., Jones, J.D.G. and Dangl, J.L. (2006) 'Reactive oxygen species signaling in response to pathogens', *Plant Physiology*, 141(2), pp. 373–378. Available at: https://doi.org/10.1104/PP.106.079467.
Tosi, P. *et al.* (2018a) 'Gradients in compositions in the starchy endosperm of wheat have implications for milling and processing', *Trends in Food Science and Technology*, 82(February), pp. 1–7. Available at: https://doi.org/10.1016/j.tifs.2018.09.027.
Tosi, P. *et al.* (2018b) 'Gradients in compositions in the starchy endosperm of wheat have implications for milling and processing', *Trends in Food Science & Technology*, 82, pp. 1–7. Available at: https://doi.org/10.1016/J.TIFS.2018.09.027.
Urbanowicz, B.R. *et al.* (2014) 'Two Arabidopsis proteins synthesize acetylated xylan in vitro', *Plant Journal*, 80(2), pp. 197–206. Available at: https://doi.org/10.1111/TPJ.12643.
U.S. Department of Health and Human Services and U.S. Department of Agriculture (2020) *Dietary Guidelines for Americans 1015-2020*.
Wada, K. and Buchanan, B.B. (1981) 'Purothionin: a seed protein with thioredoxin activity', *FEBS Letters*, 124(2), pp. 237–240. Available at: https://doi.org/10.1016/0014-5793(81)80145-7.
Walkowiak, S. *et al.* (2020) 'Multiple wheat genomes reveal global variation in modern breeding', *Nature*, 588(7837), pp. 277–283. Available at: https://doi.org/10.1038/s41586-020-2961-x.
Wang, J. *et al.* (2023) 'A pangenome analysis pipeline provides insights into functional gene identification in rice', *Genome Biology*, 24(1), pp. 1–22. Available at: https://doi.org/10.1186/S13059-023-02861-9/FIGURES/7.
Wang, J. *et al.* (2024) 'Pangenome-Wide Association Study and Transcriptome Analysis Reveal a Novel QTL and Candidate Genes Controlling both Panicle and Leaf Blast Resistance in Rice', *Rice*, 17(1), p. 27. Available at: https://doi.org/10.1186/S12284-024-00707-X.
Ward, J.L. *et al.* (2008) 'The HEALTHGRAIN Cereal Diversity Screen: concept, results, and prospects', *Journal of agricultural and food chemistry*, 56(21), pp. 9699–9709. Available at: https://doi.org/10.1021/JF8009574.
Winfield, M.O. *et al.* (2018) 'High-density genotyping of the A.E. Watkins Collection of hexaploid landraces identifies a large molecular diversity compared to elite bread wheat', *Plant Biotechnology Journal*, 16(1), pp. 165–175. Available at: https://doi.org/10.1111/pbi.12757.
Wingen, L.U. *et al.* (2014a) 'Establishing the A. E. Watkins landrace cultivar collection as a resource for systematic gene discovery in bread wheat', *Theoretical and Applied Genetics*, 127(8), pp. 1831–1842. Available at: https://doi.org/10.1007/s00122-014-2344-5.
Wingen, L.U. *et al.* (2014b) 'Establishing the A. E. Watkins landrace cultivar collection as a resource for systematic gene discovery in bread wheat', *Theoretical and Applied Genetics*, 127(8), pp. 1831–1842. Available at: https://doi.org/10.1007/S00122-014-2344-5/FIGURES/3.
York, W.S. and O'Neill, M.A. (2008) 'Biochemical control of xylan biosynthesis - which end is up?', *Current opinion in plant biology*, 11(3), pp. 258–265. Available at: https://doi.org/10.1016/J.PBI.2008.02.007.

Zeng, W. *et al.* (2016) 'Asparagus IRX9, IRX10, and IRX14A are components of an active xylan backbone synthase complex that forms in the Golgi apparatus', *Plant Physiology*, 171(1), pp. 93–109. Available at: https://doi.org/10.1104/pp.15.01919.
Zhao, X. *et al.* (2018) 'Three UDP-xylose transporters participate in xylan biosynthesis by conveying cytosolic UDP-xylose into the Golgi lumen in Arabidopsis', *Journal of Experimental Botany*, 69(5), p. 1125. Available at: https://doi.org/10.1093/JXB/ERX448.
Zhou, X. *et al.* (2022) 'Phytoene Synthase: The Key Rate-Limiting Enzyme of Carotenoid Biosynthesis in Plants', *Frontiers in Plant Science*, 13, p. 884720. Available at: https://doi.org/10.3389/FPLS.2022.884720/BIBTEX.
Zhu, X. *et al.* (2018) 'Mutation of the OsSAC1 Gene, which Encodes an Endoplasmic Reticulum Protein with an Unknown Function, Causes Sugar Accumulation in Rice Leaves', *Plant & cell physiology*, 59(3), pp. 487–499. Available at: https://doi.org/10.1093/PCP/PCX203.

Abutair, A.S., Naser, I.A. and Hamed, A.T. (2016) 'Soluble fibers from psyllium improve glycemic response and body weight among diabetes type 2 patients (randomized control trial)', *Nutrition Journal*, 15(1), pp. 1–7. Available at: https://doi.org/10.1186/s12937-016-0207-4.
Alemui, S.K. *et al.* (2021) 'Genome-wide association mapping identifies yellow rust resistance loci in Ethiopian durum wheat germplasm', *PLoS ONE*, 16(5 May 2021), pp. 1–28. Available at: https://doi.org/10.1371/journal.pone.0243675.
Allen, A.M. *et al.* (2017) 'Characterization of a Wheat Breeders' Array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (*Triticum aestivum*)', *Plant Biotechnology Journal*, 15(3), pp. 390–401. Available at: https://doi.org/10.1111/pbi.12635.
Alonso, M.P. *et al.* (2021) 'QTL mapping of spike fertility index in bread wheat', *Crop Breeding and Applied Biotechnology*, 21(1), pp. 1–11. Available at: https://doi.org/10.1590/1984-70332021v21n1a3.
Ames, N. *et al.* (no date) 'A double-blind randomised controlled trial testing the effect of a barley product containing varying amounts and types of fibre on the postprandial glucose response of healthy volunteers'. Available at: https://doi.org/10.1017/S0007114515000367.
Anders, N. *et al.* (2012a) 'Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto xylan in grasses', *Proceedings of the National Academy of Sciences of the United States of America*, 109(3), pp. 989–993. Available at: https://doi.org/10.1073/pnas.1115858109.
Anders, N. *et al.* (2012b) 'Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto xylan in grasses', *Proceedings of the National Academy of Sciences of the United States of America*, 109(3), pp. 989–993. Available at: https://doi.org/10.1073/pnas.1115858109.
Anders, N. *et al.* (2012c) 'Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto xylan in grasses', *Proceedings of the National Academy of Sciences of the United States of America*, 109(3), pp. 989–993. Available at: https://doi.org/10.1073/pnas.1115858109.
Antoine, C. *et al.* (2004) 'Wheat bran tissue fractionation using biochemical markers', *Journal of Cereal Science*, 39(3), pp. 387–393. Available at: https://doi.org/10.1016/J.JCS.2004.02.001.
Balfourier, F. *et al.* (2019) 'Worldwide phylogeography and history of wheat genetic diversity', *Science Advances*, 5(5), p. eaav0536. Available at: https://doi.org/10.1126/sciadv.aav0536.
Barron, C., Surget, A. and Rouau, X. (2007) 'Relative amounts of tissues in mature wheat (Triticum aestivum L.) grain and their carbohydrate and phenolic acid composition', *Journal of Cereal Science*, 45(1), pp. 88–96. Available at: https://doi.org/10.1016/j.jcs.2006.07.004.
Bates, B. *et al.* (2014a) *National Diet and Nutrition Survey: Results from Years 1, 2, 3 and 4 (combined) of the Rolling Programme (2008/2009-2011/2012): A survey carried out on behalf of Public Health England and the Food Standards Agency*. Public Health England.
Bates, B. *et al.* (2014b) *National Diet and Nutrition Survey: Results from Years 1, 2, 3 and 4 (combined) of the Rolling Programme (2008/2009-2011/2012): A survey carried out on behalf of Public Health England and the Food Standards Agency*. Public Health England.

Baydoun, E.A., Waldron, K.W. and Brett, C.T. (1989) 'The interaction of xylosyltransferase and glucuronyltransferase involved in glucuronoxylan synthesis in pea (Pisum sativum) epicotyls.', *The Biochemical journal*, 257(3), pp. 853–858. Available at: https://doi.org/10.1042/bj2570853.
Boros, D. *et al.* (2002) 'Chromosome location of genes controlling the content of dietary fibre and arabinoxylans in rye', *Euphytica*, 128(1), pp. 1–8. Available at: https://doi.org/10.1023/A:1020639601959/METRICS.
Botticella, E., Savatin, D.V. and Sestili, F. (2021) 'The Triple Jags of Dietary Fibers in Cereals: How Biotechnology Is Longing for High Fiber Grains', *Frontiers in Plant Science*. Frontiers Media S.A. Available at: https://doi.org/10.3389/fpls.2021.745579.
De Brier, N. *et al.* (2015) 'The impact of pearling as a treatment prior to wheat roller milling on the texture and structure of bran-rich breakfast flakes', *LWT - Food Science and Technology*, 62(1), pp. 668–674. Available at: https://doi.org/10.1016/j.lwt.2014.08.015.
Brown, D.M. *et al.* (2005) 'Identification of novel genes in Arabidopsis involved in secondary cell wall formation using expression profiling and reverse genetics', *Plant Cell*, 17(8), pp. 2281–2295. Available at: https://doi.org/10.1105/tpc.105.031542.
Bulone, V., Schwerdt, J.G. and Fincher, G.B. (2019) 'Co-evolution of Enzymes Involved in Plant Cell Wall Metabolism in the Grasses', *Frontiers in Plant Science*. Frontiers Media S.A. Available at: https://doi.org/10.3389/fpls.2019.01009.
Cagirici, H.B. *et al.* (2021) 'Multiple variant calling pipelines in wheat whole exome sequencing', *International Journal of Molecular Sciences*, 22(19). Available at: https://doi.org/10.3390/IJMS221910400/S1.
Carpita, N.C. and Gibeaut, D.M. (1993) 'Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth', *The Plant journal : for cell and molecular biology*, 3(1), pp. 1–30. Available at: https://doi.org/10.1111/J.1365-313X.1993.TB00007.X.
De Carvalho, C.M. *et al.* (2017) 'Plasma glucose and insulin responses after consumption of breakfasts with different sources of soluble fiber in type 2 diabetes patients: A randomized crossover clinical trial', *American Journal of Clinical Nutrition*, 106(5), pp. 1238–1245. Available at: https://doi.org/10.3945/ajcn.117.157263.
Charmet, G. *et al.* (2009) 'Genetics of dietary fibre in bread wheat', *Euphytica*, 170(1–2), pp. 155–168. Available at: https://doi.org/10.1007/s10681-009-0019-0.
Chateigner-Boutin, A.L. *et al.* (2018) 'Ferulate and lignin cross-links increase in cell walls of wheat grain outer layers during late development', *Plant Science*, 276(September), pp. 199–207. Available at: https://doi.org/10.1016/j.plantsci.2018.08.022.
Chayut, N. *et al.* (2011) 'A bulk segregant transcriptome analysis reveals metabolic and cellular processes associated with Orange allelic variation and fruit β-carotene accumulation in melon fruit'. Available at: https://doi.org/10.1186/s12870-015-0661-8.
Cheng, S. *et al.* (2023) 'Harnessing Landrace Diversity Empowers Wheat Breeding for Climate Resilience', *bioRxiv*, p. 2023.10.04.560903. Available at: https://doi.org/10.1101/2023.10.04.560903.
Cherkaoui, M. *et al.* (2019) 'Cell Wall Proteome of Wheat Grain Endosperm and Outer Layers at Two Key Stages of Early Development', *International Journal of Molecular Sciences 2020, Vol. 21, Page 239*, 21(1), p. 239. Available at: https://doi.org/10.3390/IJMS21010239.
Chutkan, R. *et al.* (2012) 'Viscous versus nonviscous soluble fiber supplements: mechanisms and evidence for fiber-specific health benefits', *Journal of the American Academy of Nurse Practitioners*, 24(8), pp. 476–487. Available at: https://doi.org/10.1111/J.1745-7599.2012.00758.X.
Costa, M. *et al.* (2013) 'Expression-based and co-localization detection of arabinogalactan protein 6 and arabinogalactan protein 11 interactors in Arabidopsis pollen and pollen tubes', *BMC Plant Biology*, 13(1), p. 7. Available at: https://doi.org/10.1186/1471-2229-13-7.
Darvasi, A. and Soller, M. (1995) 'Advanced Intercross Lines, an Experimental Population for Fine Genetic Mapping', *LANDER and BOTSTEIN* [Preprint].
Delledonne, M. *et al.* (2001) 'Signal interactions between nitric oxide and reactive oxygen intermediates in the plant hypersensitive disease resistance response', *Proceedings of the National Academy of Sciences*, 98(23), pp. 13454–13459. Available at: https://doi.org/10.1073/PNAS.231178298.

Dost, K. and Tokul, O. (2006) 'Determination of phytic acid in wheat and wheat products by reverse phase high performance liquid chromatography', *Analytica Chimica Acta*, 558(1–2), pp. 22–27. Available at: https://doi.org/10.1016/j.aca.2005.11.035.
Douglas, S.G. (1981) 'A rapid method for the determination of pentosans in wheat flour', *Food Chemistry*, 7(2), pp. 139–145. Available at: https://doi.org/10.1016/0308-8146(81)90059-5.
Emily S, C. *et al.* (2013) 'Redefining agricultural yields: from tonnes to people nourished per hectare ', *Environmental Research Letters*, 8. Available at: https://doi.org/10.1088/1748-9326/8/3/034015.
Farre Martinez, A. *et al.* (2021) 'Resolving a QTL complex for height, heading, and grain yield on chromosome 3A in bread wheat', *Journal of Experimental Botany*, 72(8), pp. 2965–2978. Available at: https://doi.org/10.1093/jxb/erab058.
Finnie, S.M., Bettge, A.D. and Morris, C.F. (2006) 'Influence of Cultivar and Environment on Water-Soluble and Water-Insoluble Arabinoxylans in Soft Wheat', *Cereal Chemistry Journal*, 83(6), pp. 617–623. Available at: https://doi.org/10.1094/CC-83-0617.
Freeman, J. *et al.* (2016) 'Effect of suppression of arabinoxylan synthetic genes in wheat endosperm on chain length of arabinoxylan and extract viscosity', *Plant Biotechnology Journal*, 14(1), pp. 109–116. Available at: https://doi.org/10.1111/pbi.12361.
Freeman, J. *et al.* (2017) 'Feruloylation and structure of arabinoxylan in wheat endosperm cell walls from RNAi lines with suppression of genes responsible for backbone synthesis and decoration', *Plant Biotechnology Journal*, 15(11), pp. 1429–1438. Available at: https://doi.org/10.1111/pbi.12727.
Gebruers, K. *et al.* (2008) 'Variation in the content of dietary fiber and components thereof in wheats in the HEALTHGRAIN Diversity Screen', *Journal of agricultural and food chemistry*, 56(21), pp. 9740–9749. Available at: https://doi.org/10.1021/JF800975W.
Gibeaut, D.M. and Carpita, N.C. (1990) 'Separation of membranes by flotation centrifugation for in vitro synthesis of plant cell wall polysaccharides', *Protoplasma*, 156(1–2), pp. 82–93. Available at: https://doi.org/10.1007/BF01666509.
Gupta, R.K., Gangoliya, S.S. and Singh, N.K. (2013) 'Reduction of phytic acid and enhancement of bioavailable micronutrients in food grains', *Journal of Food Science and Technology*. Springer India, pp. 676–684. Available at: https://doi.org/10.1007/s13197-013-0978-y.
Hamant, O. and Traas, J. (2010) 'The mechanics behind plant development', *New Phytologist*, 185(2), pp. 369–385. Available at: https://doi.org/10.1111/J.1469-8137.2009.03100.X.
Harland, J. (2014) 'Authorised EU health claims for barley and oat beta-glucans', in *Foods, Nutrients and Food Ingredients with Authorised EU Health Claims*. Elsevier Ltd, pp. 25–45. Available at: https://doi.org/10.1533/9780857098481.2.25.
Harland, J. (2015) 'Authorised EU health claims for wheat bran fibre', in *Foods, Nutrients and Food Ingredients with Authorised EU Health Claims: Volume 2*. Elsevier, pp. 109–127. Available at: https://doi.org/10.1016/B978-1-78242-382-9.00006-2.
Haskå, L., Nyman, M. and Andersson, R. (2008) 'Distribution and characterisation of fructan in wheat milling fractions', *Journal of Cereal Science*, 48(3), pp. 768–774. Available at: https://doi.org/10.1016/J.JCS.2008.05.002.
Hassan, A.S. *et al.* (2017) 'A Genome Wide Association Study of arabinoxylan content in 2-row spring barley grain', *PloS one*, 12(8). Available at: https://doi.org/10.1371/JOURNAL.PONE.0182537.
Hernández-Espinosa, N. *et al.* (2022) 'Quantification of the total and water-extractable pentosan content in wheat flour using a higher-throughput version of the phloroglucinol colorimetric assay', *Zernodo* [Preprint]. Available at: https://doi.org/10.5281/ZENODO.6824411.
Höng, K. *et al.* (2021) 'The thionin family of antimicrobial peptides', *PLoS One*, 16(7), p. e0254549. Available at: https://doi.org/10.1371/journal.pone.0254549.
Hossain, K. *et al.* (2013) *Interdependence of cultivar and environment on fiber composition in wheat bran*, *525 AJCS*.
Hughes, P. *et al.* (2000) 'The cytotoxic plant protein, β-purothionin, forms ion channels in lipid membranes', *Journal of Biological Chemistry*, 275(2), pp. 823–827. Available at: https://doi.org/10.1074/jbc.275.2.823.

Ibba, M.I. *et al.* (2021) 'Genome-wide association analysis for arabinoxylan content in common wheat (T. aestivum L.) flour', *Journal of Cereal Science*, 98, p. 103166. Available at: https://doi.org/10.1016/j.jcs.2021.103166.
Itoh, N. *et al.* (2019) 'Next-generation sequencing-based bulked segregant analysis for QTL mapping in the heterozygous species Brassica rapa', 132, pp. 2913–2925. Available at: https://doi.org/10.1007/s00122-019-03396-z.
Jones, J.M. (2014) 'CODEX-aligned dietary fiber definitions help to bridge the "fiber gap"', *Nutrition Journal*, 13(1), p. 34. Available at: https://doi.org/10.1186/1475-2891-13-34.
Kaczmarczyk, M.M., Miller, M.J. and Freund, G.G. (2012) 'The health benefits of dietary fiber: Beyond the usual suspects of type 2 diabetes mellitus, cardiovascular disease and colon cancer', *Metabolism: Clinical and Experimental*, pp. 1058–1066. Available at: https://doi.org/10.1016/j.metabol.2012.01.017.
Keegstra, K. (2010) 'Plant cell walls', *Plant Physiology*, 154(2), pp. 483–486. Available at: https://doi.org/10.1104/pp.110.161240.
Kelleher, J. *et al.* (1984) 'Degradation of cellulose within the gastrointestinal tract in man', *Gut*, 25(8), pp. 811–815. Available at: https://doi.org/10.1136/gut.25.8.811.
Kellow, N.J. and Walker, K.Z. (2018) 'Authorised EU health claim for arabinoxylan', in *Foods, Nutrients and Food Ingredients with Authorised EU Health Claims*. Elsevier Inc., pp. 201–218. Available at: https://doi.org/10.1016/B978-0-08-100922-2.00013-9.
Kim, D., Langmead, B. and Salzberg, S.L. (2015) 'HISAT: a fast spliced aligner with low memory requirements', *Nature methods*, 12(4), pp. 357–360. Available at: https://doi.org/10.1038/NMETH.3317.
Kim, Y.Y. *et al.* (2012) 'A novel F-box protein represses endothecial secondary wall thickening for anther dehiscence in Arabidopsis thaliana', *Journal of plant physiology*, 169(2), pp. 212–216. Available at: https://doi.org/10.1016/J.JPLPH.2011.09.006.
Kiszonas, A.M., Courtin, C.M. and Morris, C.F. (2012) 'A Critical Assessment of the Quantification of Wheat Grain Arabinoxylans Using a Phloroglucinol Colorimetric Assay', *Cereal Chemistry Journal*, 89(3), pp. 143–150. Available at: https://doi.org/10.1094/CCHEM-02-12-0016-R.
Kozlova, L. V. *et al.* (2022) 'Identification of genes involved in the formation of soluble dietary fiber in winter rye grain and their expression in cultivars with different viscosities of wholemeal water extract', *The Crop Journal*, 10(2), pp. 532–549. Available at: https://doi.org/10.1016/J.CJ.2021.05.008.
de la Fuente Cantó, C. and Vigouroux, Y. (2022) 'Evaluation of nine statistics to identify QTLs in bulk segregant analysis using next generation sequencing approaches', *BMC Genomics*, 23(1), pp. 1–12. Available at: https://doi.org/10.1186/S12864-022-08718-Y/TABLES/1.
Lafiandra, D. and Shewry, P. (2014) 'Improving Cereal Grain Carbohydrates for Diet and Health', *Journal of Cereal Science*, 59. Available at: https://doi.org/10.1016/j.jcs.2014.01.001.
Laperche, A. *et al.* (2007) 'Using genotype x nitrogen interaction variables to evaluate the QTL involved in wheat tolerance to nitrogen constraints', *Theoretical and Applied Genetics*, 115(3), pp. 399–415. Available at: https://doi.org/10.1007/s00122-007-0575-4.
Li, Q. *et al.* (2020) 'Quantitative Trait Locus (QTLs) Mapping for Quality Traits of Wheat Based on High Density Genetic Map Combined With Bulked Segregant Analysis RNA-seq (BSR-Seq) Indicates That the Basic 7S Globulin Gene Is Related to Falling Number', *Frontiers in Plant Science*, 11, p. 600788. Available at: https://doi.org/10.3389/FPLS.2020.600788/BIBTEX.
Loosveld, A.M.A., Grobet, P.J. and Delcour, J.A. (1997) 'Contents and Structural Features of Water-Extractable Arabinogalactan in Wheat Flour Fractions', *Journal of Agricultural and Food Chemistry*, 45(6), pp. 1998–2002. Available at: https://doi.org/10.1021/jf960901k.
Love, M.I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), pp. 1–21. Available at: https://doi.org/10.1186/S13059-014-0550-8/FIGURES/9.
Lovegrove, A. *et al.* (2013) 'RNA interference suppression of genes in glycosyl transferase families 43 and 47 in wheat starchy endosperm causes large decreases in arabinoxylan content', *Plant Physiology*, 163(1), pp. 95–107. Available at: https://doi.org/10.1104/pp.113.222653.
Lovegrove, A., Wingen, L.U., Plummer, A., Wood, A., Passmore, D., Kosik, O., Freeman, J., Mitchell, Rowan A. C., *et al.* (2020) 'Identification of a major QTL and associated molecular marker for high

arabinoxylan fibre in white wheat flour', *PLOS ONE*. Edited by A. Zhang, 15(2), p. e0227826. Available at: https://doi.org/10.1371/journal.pone.0227826.
Lovegrove, A., Wingen, L.U., Plummer, A., Wood, A., Passmore, D., Kosik, O., Freeman, J., Mitchell, Rowan A.C., *et al.* (2020a) 'Identification of a major QTL and associated molecular marker for high arabinoxylan fibre in white wheat flour', *PLoS ONE*, 15(2), pp. 1–13. Available at: https://doi.org/10.1371/journal.pone.0227826.
Lovegrove, A., Wingen, L.U., Plummer, A., Wood, A., Passmore, D., Kosik, O., Freeman, J., Mitchell, Rowan A.C., *et al.* (2020b) 'Identification of a major QTL and associated molecular marker for high arabinoxylan fibre in white wheat flour', *PLoS ONE*, 15(2), pp. 1–13. Available at: https://doi.org/10.1371/journal.pone.0227826.
Maes, C. and Delcour, J.A. (2002) 'Structural characterisation of water-extractable and water-unextractable arabinoxylans in wheat bran', *Journal of Cereal Science*, 35(3), pp. 315–326. Available at: https://doi.org/10.1006/jcrs.2001.0439.
Mata-Pérez, C. and Spoel, S.H. (2019) 'Thioredoxin-mediated redox signalling in plant immunity', *Plant Science*, 279, pp. 27–33. Available at: https://doi.org/10.1016/J.PLANTSCI.2018.05.001.
McRae, M.P. (2017) 'Health Benefits of Dietary Whole Grains: An Umbrella Review of Meta-analyses', *Journal of Chiropractic Medicine*. Elsevier USA, pp. 10–18. Available at: https://doi.org/10.1016/j.jcm.2016.08.008.
Metzger, M.B. *et al.* (no date) 'RING-type E3 ligases: Master manipulators of E2 ubiquitin-conjugating enzymes and ubiquitination'. Available at: https://doi.org/10.1016/j.bbamcr.2013.05.026.
Michelmore, R.W., Paran, I. and Kesseli, R. V. (1991) 'Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations', *Proceedings of the National Academy of Sciences of the United States of America*, 88(21), pp. 9828–9832. Available at: https://doi.org/10.1073/PNAS.88.21.9828.
Miller, T. *et al.* (2001) 'The Watkins collection of landrace derived wheats', *Wheat taxon legacy of John Percival. London: Linnean Society*, pp. 113–120.
Mio, K. *et al.* (2020) 'Effects of β-glucan Rich Barley Flour on Glucose and Lipid Metabolism in the Ileum, Liver, and Adipose Tissues of High-Fat Diet Induced-Obesity Model Male Mice Analyzed by DNA Microarray', *Nutrients*, 12(11), pp. 1–18. Available at: https://doi.org/10.3390/NU12113546.
Mio, K. *et al.* (2023) 'A single administration of barley β-glucan and arabinoxylan extracts reduces blood glucose levels at the second meal via intestinal fermentation', 87(1), pp. 99–107. Available at: https://doi.org/10.1093/bbb/zbac171.
Mitchell, R.A.C., Dupree, P. and Shewry, P.R. (2007) 'A Novel Bioinformatics Approach Identifies Candidate Genes for the Synthesis and Feruloylation of Arabinoxylan', *Plant Physiology*, 144(1), pp. 43 LP – 53. Available at: https://doi.org/10.1104/pp.106.094995.
Nagawa, S., Xu, T. and Yang, Z. (2010) 'RHO GTPase in plants: Conservation and invention of regulators and effectors', *Small GTPases*, 1(2), pp. 78–88. Available at: https://doi.org/10.4161/SGTP.1.2.14544.
Nauck, M.A. *et al.* (2017) 'Cardiovascular Actions and Clinical Outcomes With Glucagon-Like Peptide-1 Receptor Agonists and Dipeptidyl Peptidase-4 Inhibitors', *Circulation*, 136(9), pp. 849–870. Available at: https://doi.org/10.1161/CIRCULATIONAHA.117.028136.
O'Brien, J.A. *et al.* (2012) 'Reactive oxygen species and their role in plant defence and cell wall metabolism', *Planta*, 236(3), pp. 765–779. Available at: https://doi.org/10.1007/S00425-012-1696-9/FIGURES/3.
Oudgenoeg, G. *et al.* (2001) 'Peroxidase-mediated cross-linking of a tyrosine-containing peptide with ferulic acid', *Journal of agricultural and food chemistry*, 49(5), pp. 2503–2510. Available at: https://doi.org/10.1021/JF000906O.
Parveen, A. *et al.* (123AD) 'Genome-wide analysis of RING-type E3 ligase family identifies potential candidates regulating high amylose starch biosynthesis in wheat (Triticum aestivum L.)', *Scientific Reports* |, 11, p. 11461. Available at: https://doi.org/10.1038/s41598-021-90685-7.
Perretant, M. *et al.* (no date) *QTL analysis of bread-making quality in wheat using a doubled haploid population*, *Springer*.
Piro, M.C., Muylle, H. and Haesaert, G. (2023) 'Exploiting Rye in Wheat Quality Breeding: The Case of Arabinoxylan Content', *Plants*, 12(4), p. 737. Available at: https://doi.org/10.3390/PLANTS12040737.

Quraishi, U.M. *et al.* (2009) 'Genomics in cereals: From genome-wide conserved orthologous set (COS) sequences to candidate genes for trait dissection', *Functional and Integrative Genomics*, 9(4), pp. 473–484. Available at: https://doi.org/10.1007/s10142-009-0129-8.
Quraishi, U.M. *et al.* (2011) 'Combined meta-genomics analyses unravel candidate genes for the grain dietary fiber content in bread wheat (Triticum aestivum L.)', *Functional and Integrative Genomics*, 11(1), pp. 71–83. Available at: https://doi.org/10.1007/s10142-010-0183-2.
Ren, Y. *et al.* (2014) 'Site-directed mutagenesis of IRX9, IRX9L and IRX14 proteins involved in xylan biosynthesis: Glycosyltransferase activity is not required for IRX9 function in arabidopsis', *PLoS ONE*, 9(8), p. e105014. Available at: https://doi.org/10.1371/journal.pone.0105014.
Rennie, E.A. and Scheller, H.V. (2014) 'Xylan biosynthesis', *Current Opinion in Biotechnology*, pp. 100–107. Available at: https://doi.org/10.1016/j.copbio.2013.11.013.
Saulnier, L., Vigouroux, J. and Thibault, J.F. (1995) 'Isolation and partial characterization of feruloylated oligosaccharides from maize bran', *Carbohydrate Research*, 272(2), pp. 241–253. Available at: https://doi.org/10.1016/0008-6215(95)00053-V.
Shen, X. *et al.* (2003) 'Detection of Fusarium head blight resistance QTL in a wheat population using bulked segregant analysis', *Theoretical and Applied Genetics*, 106(6), pp. 1041–1047. Available at: https://doi.org/10.1007/S00122-002-1133-8/METRICS.
Shewry, P.R. *et al.* (2011) 'Relationship between the contents of bioactive components in grain and the release dates of wheat lines in the HEALTHGRAIN diversity screen', *Journal of Agricultural and Food Chemistry*, 59(3), pp. 928–933. Available at: https://doi.org/10.1021/JF103860X.
Shewry, P.R. and Hey, S.J. (2015) 'The contribution of wheat to human diet and health', *Food and Energy Security*. Wiley-Blackwell Publishing Ltd, pp. 178–202. Available at: https://doi.org/10.1002/FES3.64.
Shewry, P.R., Pellny, T.K. and Lovegrove, A. (2016) 'Is modern wheat bad for health?', *Nature Plants*, 2(7), pp. 1–3. Available at: https://doi.org/10.1038/NPLANTS.2016.97.
Speck, T. and Burgert, I. (2011) 'Plant Stems: Functional Design and Mechanics', *https://doi.org/10.1146/annurev-matsci-062910-100425*, 41, pp. 169–193. Available at: https://doi.org/10.1146/ANNUREV-MATSCI-062910-100425.
Strohmeier, M. *et al.* (2009) 'Molecular modeling of family GH16 glycoside hydrolases: Potential roles for xyloglucan transglucosylases/hydrolases in cell wall modification in the poaceae', *Protein Science*, 13(12), pp. 3200–3213. Available at: https://doi.org/10.1110/ps.04828404.
Stuper-Szablewska, K. *et al.* (2019) 'Quantitative profile of phenolic acids and antioxidant activity of wheat grain exposed to stress', *European Food Research and Technology*, 245(8), pp. 1595–1603. Available at: https://doi.org/10.1007/s00217-019-03262-8.
Suliman, M. *et al.* (2013) 'Identification of glycosyltransferases involved in cell wall synthesis of wheat endosperm', *Journal of Proteomics*, 78, pp. 508–521. Available at: https://doi.org/10.1016/j.jprot.2012.10.021.
Thondre, P.S., Shafat, A. and Clegg, M.E. (2013) 'Molecular weight of barley β-glucan influences energy expenditure, gastric emptying and glycaemic response in human subjects', *The British journal of nutrition*, 110(12), pp. 2173–2179. Available at: https://doi.org/10.1017/S0007114513001682.
Torres, M.A., Jones, J.D.G. and Dangl, J.L. (2006) 'Reactive oxygen species signaling in response to pathogens', *Plant Physiology*, 141(2), pp. 373–378. Available at: https://doi.org/10.1104/PP.106.079467.
Tosi, P. *et al.* (2018a) 'Gradients in compositions in the starchy endosperm of wheat have implications for milling and processing', *Trends in Food Science and Technology*, 82(February), pp. 1–7. Available at: https://doi.org/10.1016/j.tifs.2018.09.027.
Tosi, P. *et al.* (2018b) 'Gradients in compositions in the starchy endosperm of wheat have implications for milling and processing', *Trends in Food Science & Technology*, 82, pp. 1–7. Available at: https://doi.org/10.1016/J.TIFS.2018.09.027.
Urbanowicz, B.R. *et al.* (2014) 'Two Arabidopsis proteins synthesize acetylated xylan in vitro', *Plant Journal*, 80(2), pp. 197–206. Available at: https://doi.org/10.1111/TPJ.12643.
U.S. Department of Health and Human Services and U.S. Department of Agriculture (2020) *Dietary Guidelines for Americans 1015-2020*.
Wada, K. and Buchanan, B.B. (1981) 'Purothionin: a seed protein with thioredoxin activity', *FEBS Letters*, 124(2), pp. 237–240. Available at: https://doi.org/10.1016/0014-5793(81)80145-7.

Walkowiak, S. *et al.* (2020) 'Multiple wheat genomes reveal global variation in modern breeding', *Nature*, 588(7837), pp. 277–283. Available at: https://doi.org/10.1038/s41586-020-2961-x.
Wang, J. *et al.* (2023) 'A pangenome analysis pipeline provides insights into functional gene identification in rice', *Genome Biology*, 24(1), pp. 1–22. Available at: https://doi.org/10.1186/S13059-023-02861-9/FIGURES/7.
Wang, J. *et al.* (2024) 'Pangenome-Wide Association Study and Transcriptome Analysis Reveal a Novel QTL and Candidate Genes Controlling both Panicle and Leaf Blast Resistance in Rice', *Rice*, 17(1), p. 27. Available at: https://doi.org/10.1186/S12284-024-00707-X.
Ward, J.L. *et al.* (2008) 'The HEALTHGRAIN Cereal Diversity Screen: concept, results, and prospects', *Journal of agricultural and food chemistry*, 56(21), pp. 9699–9709. Available at: https://doi.org/10.1021/JF8009574.
Winfield, M.O. *et al.* (2018) 'High-density genotyping of the A.E. Watkins Collection of hexaploid landraces identifies a large molecular diversity compared to elite bread wheat', *Plant Biotechnology Journal*, 16(1), pp. 165–175. Available at: https://doi.org/10.1111/pbi.12757.
Wingen, L.U. *et al.* (2014a) 'Establishing the A. E. Watkins landrace cultivar collection as a resource for systematic gene discovery in bread wheat', *Theoretical and Applied Genetics*, 127(8), pp. 1831–1842. Available at: https://doi.org/10.1007/s00122-014-2344-5.
Wingen, L.U. *et al.* (2014b) 'Establishing the A. E. Watkins landrace cultivar collection as a resource for systematic gene discovery in bread wheat', *Theoretical and Applied Genetics*, 127(8), pp. 1831–1842. Available at: https://doi.org/10.1007/S00122-014-2344-5/FIGURES/3.
York, W.S. and O'Neill, M.A. (2008) 'Biochemical control of xylan biosynthesis - which end is up?', *Current opinion in plant biology*, 11(3), pp. 258–265. Available at: https://doi.org/10.1016/J.PBI.2008.02.007.
Zeng, W. *et al.* (2016) 'Asparagus IRX9, IRX10, and IRX14A are components of an active xylan backbone synthase complex that forms in the Golgi apparatus', *Plant Physiology*, 171(1), pp. 93–109. Available at: https://doi.org/10.1104/pp.15.01919.
Zhao, X. *et al.* (2018) 'Three UDP-xylose transporters participate in xylan biosynthesis by conveying cytosolic UDP-xylose into the Golgi lumen in Arabidopsis', *Journal of Experimental Botany*, 69(5), p. 1125. Available at: https://doi.org/10.1093/JXB/ERX448.
Zhou, X. *et al.* (2022) 'Phytoene Synthase: The Key Rate-Limiting Enzyme of Carotenoid Biosynthesis in Plants', *Frontiers in Plant Science*, 13, p. 884720. Available at: https://doi.org/10.3389/FPLS.2022.884720/BIBTEX.
Zhu, X. *et al.* (2018) 'Mutation of the OsSAC1 Gene, which Encodes an Endoplasmic Reticulum Protein with an Unknown Function, Causes Sugar Accumulation in Rice Leaves', *Plant & cell physiology*, 59(3), pp. 487–499. Available at: https://doi.org/10.1093/PCP/PCX203.