

June 2018



## Student Report No. SR45

Understanding the genetics of wheat yield to deploy high and stable yielding wheat varieties across UK environments

Jemima Brinton<sup>1</sup> and Cristobal Uauy<sup>1</sup>

<sup>1</sup> John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK

**Supervisor:** Cristobal Uauy

This is the final report of a PhD project (21130023) that ran from Oct 2014 to Sep 2017. The work was funded by the John Innes Foundation and a contract for £54,000 from AHDB Cereals & Oilseeds.

While the Agriculture and Horticulture Development Board seeks to ensure that the information contained within this document is accurate at the time of printing, no warranty is given in respect thereof and, to the maximum extent permitted by law, the Agriculture and Horticulture Development Board accepts no liability for loss, damage or injury howsoever caused (including that caused by negligence) or suffered directly or indirectly in relation to information and opinions contained in or omitted from this document.

Reference herein to trade names and proprietary products without stating that they are protected does not imply that they may be regarded as unprotected and thus free for general use. No endorsement of named products is intended, nor is any criticism implied of other alternative, but unnamed, products.

AHDB Cereals & Oilseeds is a part of the Agriculture and Horticulture Development Board (AHDB).



# CONTENTS

<b>1</b>	<b>ABSTRACT</b> .....	<b>6</b>
<b>2</b>	<b>INTRODUCTION</b> .....	<b>7</b>
	<b>2.1 Genetic control of grain weight</b> .....	<b>7</b>
	2.1.1 Understanding of the genetic control of grain size in wheat .....	8
	<b>2.2 The 5A and 6A QTL for grain weight</b> .....	<b>9</b>
	2.2.1 Identification of the 5A QTL .....	9
	2.2.2 Identification of the 6A QTL .....	9
	2.2.3 TaGW2-A as a potential candidate gene underlying the 6A QTL .....	10
	<b>2.3 Thesis aims</b> .....	<b>11</b>
<b>3</b>	<b>MATERIALS AND METHODS</b> .....	<b>12</b>
	<b>3.1 Characterisation of 6A and 5A Near Isogenic Lines</b> .....	<b>12</b>
	3.1.1 Plant material and growth .....	12
	3.1.2 Phenotyping .....	12
	3.1.3 Carpel/grain developmental time courses .....	13
	3.1.4 Cell size measurements .....	14
	3.1.5 Statistical analysis .....	15
	<b>3.2 Fine mapping of the 5A and 6A QTL</b> .....	<b>15</b>
	3.2.1 Plant material and growth .....	15
	3.2.2 Grain phenotyping .....	16
	3.2.3 Marker development .....	16
	3.2.4 Physical positions .....	18
	3.2.5 DNA extraction and KASP genotyping .....	18
	3.2.6 Exome capture for haplotype analysis .....	18
	3.2.7 Statistical analysis .....	19
	<b>3.3 Comparative transcriptomics of 5A NILs</b> .....	<b>19</b>
	3.3.1 Plant material .....	19
	3.3.2 RNA extraction and sequencing .....	19
	3.3.3 Read alignment and differential expression analysis .....	20
	3.3.4 GO term enrichment .....	20

3.3.5	Functional annotation .....	20
3.3.6	Identification of transcription factor binding sites .....	21
3.3.7	Enrichment testing .....	21
<b>4</b>	<b>RESULTS.....</b>	<b>22</b>
<b>4.1</b>	<b>Characterisation of 6A and 5A Near Isogenic Lines .....</b>	<b>22</b>
4.1.1	6A QTL for grain width.....	22
4.1.2	5A QTL for grain length .....	31
<b>4.2</b>	<b>Fine mapping of the 5A and 6A QTL .....</b>	<b>44</b>
4.2.1	Genetic mapping of the 6A QTL for grain width .....	44
4.2.2	Genetic mapping of the 5A QTL for grain length .....	49
<b>4.3</b>	<b>Comparative transcriptomics of 5A NILs.....</b>	<b>60</b>
4.3.1	RNA-sequencing of 5A near isogenic lines .....	60
4.3.2	Comparison between Chinese Spring reference transcriptomes.....	63
4.3.3	Many DE transcripts during early grain development are shared between NILs.....	71
4.3.4	DE transcripts between NILs are concentrated on chromosome 5A .....	80
4.3.5	DE transcripts outside of chromosome 5A are enriched in specific transcription factor binding sites.....	82
4.3.6	Functional annotation of DE transcripts .....	85
<b>5</b>	<b>DISCUSSION .....</b>	<b>95</b>
<b>5.1</b>	<b>Mechanisms and genes underlying the 6A and 5A QTL.....</b>	<b>95</b>
5.1.1	Genes and pathways underlying the 6A QTL.....	95
5.1.2	Genes and pathways underlying the 5A QTL.....	98
5.1.3	Maternal control of grain size.....	101
5.1.4	Importance of early grain development.....	102
<b>5.2</b>	<b>Potential consequences of increasing grain size and pleiotropic effects of the 5A and 6A QTL.....</b>	<b>103</b>
5.2.1	Pleiotropic effects on yield components.....	103
5.2.2	Pleiotropic developmental effects .....	104
5.2.3	Pleiotropic effects on grain nutrient composition .....	104
5.2.4	Understanding the causes of pleiotropic effects.....	105

<b>5.3</b>	<b>Combining beneficial alleles.....</b>	<b>105</b>
5.3.1	Combining homoeologues .....	105
5.3.2	Combining components of pathways involved in grain size regulation .....	106
5.3.3	Combining grain size genes with other aspects of plant development .....	106
<b>5.4</b>	<b>Concluding statement.....</b>	<b>107</b>
<b>6</b>	<b>REFERENCES .....</b>	<b>107</b>

## 1 Abstract

There is an urgent need to increase crop yields to address food insecurity. Grain weight, determined by grain length and width, is an important component of final grain yield component. However, our understanding of the mechanisms that control grain weight in polyploid wheat is limited. The overall aim of this thesis was to understand the mechanisms that control grain length and width in hexaploid wheat through the characterisation of two previously identified grain weight quantitative trait loci (QTL) on chromosomes 5A and 6A.

Using near isogenic lines (NILs) we found that the 5A and 6A QTL act through different mechanisms to increase grain weight. The 5A QTL acts post-fertilisation, primarily to increase grain length (4.0%) through increased pericarp cell size. The 5A QTL also has a pleiotropic effect on grain width (1.5%) during late grain development. The 6A QTL acts during very early grain development, perhaps pre-fertilisation, and specifically increases final grain width (2.3%).

Fine-mapping reduced the QTL mapping intervals and revealed complex underlying genetic architectures. The 6A QTL mapped to a large linkage block in the centromeric region of chromosome 6A containing the known grain size gene, *TaGW2-A*, although we provide evidence to suggest that this is not the causal gene underlying the 6A QTL. Fine-mapping of the 5A QTL suggests that two tightly linked genes with an additive effect on grain length underlie the locus. A haplotype analysis suggests that the 5A QTL is not fixed in UK germplasm.

The corresponding physical intervals for both the 6A and 5A QTL remain large and contain several hundred genes, making speculation on candidates for the causal genes difficult. A transcriptomics study with the 5A NILs provided insight into the genes and pathways that are differentially regulated and hence may play a role in controlling the differences in grain weight. The markers and germplasm developed within this thesis had been published and have been made available to UK breeding companies to ensure quick uptake of this knowledge and transfer into improved wheat varieties.

## 2 Introduction

### 2.1 Genetic control of grain weight

Thousand grain weight (TGW) is largely defined by the size of individual grains and can be broken down further into the morphometric components grain length, width, height and area, which are under independent genetic control (Gegas *et al.*, 2010). These grain size parameters are mainly controlled by the coordination of cell proliferation and expansion processes.

In rice, over 400 grain weight quantitative trait loci (QTL) have been identified, and several of the underlying genes have been cloned (reviewed in Xing & Zhang, 2010; Huang *et al.*, 2013). Studies in the model species, Arabidopsis, have also provided a deep molecular insight into the control of seed size (reviewed in Li & Li, 2015; Li & Li, 2016). These studies and others have revealed that seed/grain size is controlled by genes with a diverse range of molecular functions, some examples of which are described below.

Transcription factors (TFs) belonging to many different families have been shown to be involved in the control of seed/grain size, for example, the rice *SQUAMOSA PROMOTER-BINDING LIKE* (*SPL*) TF, *OsSPL16*. *OsSPL16* was cloned as the gene underlying the rice *GRAIN WIDTH 8* (*GW8*) QTL and positively regulates grain size through the promotion of cell proliferation (Wang *et al.*, 2012). Similarly, the Arabidopsis TF, *AINTEGUMENTA* (*ANT*) also promotes cell proliferation, acting as a positive regulator of seed size (Mizukami & Fischer, 2000). TFs that act to regulate seed/grain size through the regulation of cell expansion have also been identified. *APETALA2* (*AP2*) and the WRKY TF, *TRANSPARENT TESTA GLABRA2*, both act as negative regulators of seed size by limiting cell expansion in the integument in Arabidopsis (Johnson *et al.*, 2002; Garcia *et al.*, 2005; Ohto *et al.*, 2005).

Genes involved in the ubiquitin pathway are also important regulators of seed/grain size in many plant species (reviewed in Li & Li, 2014a). This pathway acts to modify target proteins by the addition of a small protein called ubiquitin (Ub) through the sequential action of three enzymes: E1 (Ub activase), E2 (Ub conjugase) and E3 (Ub ligase). This modification has important regulatory functions in many cellular processes in plants and often involves the modified protein being targeted for degradation by the 26S proteasome (Hershko & Ciechanover, 1998). For example, *GW2*, a RING-type E3 Ub ligase, was cloned as the gene underlying a major rice grain weight QTL and negatively regulates grain width by limiting cell division (Song, XJ *et al.*, 2007). Orthologues of *GW2* in other species including Arabidopsis, wheat and maize also negatively regulate seed/grain weight (Li *et al.*, 2010; Xia, Tian *et al.*, 2013; Simmonds *et al.*, 2016) suggesting that this mechanism may be conserved across species. Downstream targets of the Arabidopsis *GW2* orthologue, *DA2*, have been identified that also regulate seed size, such as *DA1* and *UBIQUITIN SPECIFIC PROTEASE 15* (*UBP15*). *DA1* and *UBP15* interact genetically and physically and both regulate cell proliferation in the integument, however, *DA1* acts as a negative regulator whilst *UBP15* is a positive regulator (Liu *et al.*, 2008; Du *et al.*, 2014). *UBP15* is actually a

deubiquitinating enzyme and other genes with deubiquitination activity have also been identified as regulators of grain size, such as *WIDE AND THICK GRAIN 1 (WTG1)*, which regulates grain size and shape in rice mainly through cell expansion (Huang *et al.*, 2017).

Components of several different signalling pathways have also been shown to play roles in the control of seed/grain size. Several studies have demonstrated roles for components of the G-protein signalling pathway, in which heterotrimeric G-protein complexes act with membrane bound G-protein coupled receptors to transduce extracellular signals to intracellular components (Trusov & Botella, 2016). Heterotrimeric G-protein complexes consist of three subunits:  $G_{\alpha}$ ,  $G_{\beta}$  and  $G_{\gamma}$  and roles in seed/grain size regulation have been identified for all subunits in rice and Arabidopsis (reviewed in Botella, 2012). However, it is not clear if function is completely conserved across species. For example, an Arabidopsis  $G_{\gamma}$  subunit, *AGG3*, positively regulates seed size (Fang *et al.*, 2012), whilst the rice  $G_{\gamma}$  subunits, *DEP1* and *GS3* appear to be negative regulators of seed size (Fan *et al.*, 2006; Huang *et al.*, 2009). Phytohormone signalling is also important in the control of seed/grain size with roles being demonstrated for auxin, brassinosteroid and cytokinin biosynthesis and signalling components (Riefler *et al.*, 2006; Schruoff *et al.*, 2006; Jiang *et al.*, 2013). Other important signalling components have also been identified, for example *KLUH*, an Arabidopsis cytochrome P450, positively regulates seed size through promoting cell proliferation in the integuments (Adamski *et al.*, 2009) and this function appears to be conserved in both wheat and rice (Ma *et al.*, 2016). Genes affecting epigenetic status have also been shown to have important roles in the control of seed/grain size (Xiao *et al.*, 2006).

Many of the components described above have been shown to act maternally to affect the final seed/grain size (reviewed in Li & Li, 2015) and it has been proposed in several species that the maternal outer tissues (i.e. seed coat or pericarp) set an upper limit to the final size of the seed/grain by physically restricting endosperm growth (Adamski *et al.*, 2009; Hasan *et al.*, 2011; Xia, T. *et al.*, 2013).

### **2.1.1 Understanding of the genetic control of grain size in wheat**

Despite the advances in Arabidopsis and rice, our understanding of the genetic mechanisms controlling grain size remains limited in wheat. Comparative genomics approaches and association studies have provided some insight (Ma *et al.*, 2016; Simmonds *et al.*, 2016) and quantitative trait loci (QTL) associated with grain size and shape components (grain area, length and width) have been identified on almost every wheat chromosome (Brescghello & Sorrells, 2007; Gegas *et al.*, 2010; Simmonds *et al.*, 2014; Farré *et al.*, 2016; Kumar *et al.*, 2016; Brinton *et al.*, 2017). However, few of these QTL have been validated, none have been cloned and little is understood about the underlying mechanisms.

One of the major challenges to cloning grain size QTL in wheat and understanding the underlying mechanisms is the subtle nature of the effects compared to QTL in diploid species such as rice.

Grain weight QTL in rice often have effects of > 20% (Song, XJ *et al.*, 2007), whilst grain size QTL in wheat usually have effects of ~ 5 % (Uauy, 2017a) It has been proposed that the subtlety of these effects in wheat is due to functional redundancy between homoeologues resulting in the effects of variation in a single gene being masked by the effects of the remaining functional copies. Indeed, variation in the *GW2* gene in rice leads to grain weight differences of over 50% whereas a similar mutant in a single wheat homoeologues affects TGW by only 7 % in wheat (Song, X-J *et al.*, 2007; Simmonds *et al.*, 2016).

## **2.2 The 5A and 6A QTL for grain weight**

Previously in the lab, two distinct major wheat grain weight QTL were identified on chromosomes 5A and 6A (henceforth referred to as the 5A QTL and 6A QTL, respectively; Simmonds *et al.*, 2014; Brinton *et al.*, 2017). Both QTL were identified in doubled haploid (DH) populations between UK hexaploid winter wheat cultivars and validated using near isogenic lines (NILs).

### **2.2.1 Identification of the 5A QTL**

The 5A QTL was identified in a DH population developed between the UK cultivars ‘Charger’ and ‘Badger’ (CxB). The CxB DH population was evaluated for final yield and TGW across twelve environments: at least two years (yr) at five different locations (2 x England (3 yr), 1 x Scotland (2yr), 1 x France (2 yr) and 1 x Germany (2 yr)). A QTL analysis identified a region on chromosome 5A that was consistently associated with TGW, significant in seven out of twelve environments (based on the log-of-odds (LOD) score) and explaining 15.5 % of the phenotypic variation. The QTL interval was confirmed in a multi-trait multi-environment (MTME) analysis, with at least one marker in the QTL region being significantly associated with TGW in all twelve environments. In the CxB DH population, the 5A QTL increased TGW by 5.5 % with Badger offering the increasing allele.

Overall, there was a significant correlation between TGW and final grain yield across all environments, but a yield QTL only collocated with the 5A TGW in two of the twelve environments in the QTL analysis. However, MTME analysis for yield showed significant association between yield and at least one marker in the 5A QTL interval in seven out of twelve environments. It was concluded that the 5A QTL interval is associated with a consistent effect on TGW that often, but not always, translates to an increase in final grain yield. In both the QTL analysis and MTME analysis the TGW effect co-located with an association with (Brinton *et al.*, 2017).

### **2.2.2 Identification of the 6A QTL**

The 6A QTL was identified in a DH population between the UK cultivars ‘Spark’ and ‘Rialto’ (SxR) and was evaluated in the same twelve environments detailed above for the CxB DH population. A

QTL analysis identified several TGW QTL in the SxR DH population present at least five environments, but the TGW QTL on chromosome 6A was collocated with a QTL for final grain yield. Across environments, there was a significant correlation between TGW and final grain yield in the SxR DH lines. MTME analysis found that markers within the 6A QTL interval were significantly associated with TGW and yield in ten and nine out of twelve environments, respectively. In the SxR DH population, the 6A QTL increased TGW by 4.5 % and final yield by 3.8% with Rialto providing the increasing allele in both cases (Simmonds *et al.*, 2014). Interestingly, *TaGW2-A*, the A genome wheat orthologue of *GW2* (rice E3 Ub ligase, described above), was located within the 6A QTL mapping interval (Simmonds *et al.*, 2014).

### 2.2.3 *TaGW2-A* as a potential candidate gene underlying the 6A QTL

At the beginning of my PhD, several studies had investigated the role of *TaGW2-A* in the control of grain size in wheat but contradictory results had been reported.

Multiple association studies had identified an A/G promoter SNP at the -593 bp position of *TaGW2-A* (known as Hap-P2) but had produced contradictory results. One study found an association between the A allele and increased grain weight (Su *et al.*, 2011), whilst another identified the G allele as increasing grain weight (Zhang, X *et al.*, 2013). Contradictory results had also been produced as to whether the function of rice *GW2* as a negative regulator of grain weight is conserved in wheat. A natural missense mutation in exon 8 of *TaGW2-A* (Yang *et al.*, 2012) and downregulation of *TaGW2* expression by RNAi (Hong *et al.*, 2014) were both associated with an increase in grain weight, suggesting that *TaGW2-A* functions as a negative regulator of grain size in wheat. However, a separate RNAi study found that suppression of *TaGW2* expression resulted in smaller grains, suggesting positive regulation of grain size (Bednarek *et al.*, 2012). Therefore, although the evidence strongly suggested that *TaGW2-A* is involved in the control of grain size, the precise function was not clear. Numerous studies had identified grain weight QTL on wheat chromosome 6A and alluded to *TaGW2-A* as the possible causal gene (Mir *et al.*, 2012; Zhang, K *et al.*, 2013; Williams & Sorrells, 2014) but none had conclusively shown whether or not this was the case.

Although the parents of the SxR DH population, Spark and Rialto, do not have any coding region polymorphisms in *TaGW2-A*, they do have the A/G -593 bp promoter SNP (Spark-A, Rialto-G; Simmonds *et al.*, 2014). Given the association of *TaGW2-A* with final grain size, its location within the 6A QTL mapping interval and the presence of the promoter SNP, we hypothesised that *TaGW2-A* could be a candidate for the causal gene underlying the 6A QTL.

## 2.3 Thesis aims

The overall aim of this thesis is to understand the mechanisms that control grain length and width in hexaploid wheat through the characterisation of the 5A and 6A QTL. Specifically, this thesis will combine phenotypic characterisation, genetic mapping and transcriptomics to answer the following questions:

- Do the 5A and 6A QTL increase grain weight via the same or different mechanisms?
- What are the genes/pathways underlying the 5A and 6A QTL?
- Is *TaGW2-A* the gene underlying the 6A QTL?

### 3 Materials and methods

#### 3.1 Characterisation of 6A and 5A Near Isogenic Lines

##### 3.1.1 Plant material and growth

The 5A, 6A and *TaGW2-A* NILs used in this chapter were generated by James Simmonds and are described in Brinton *et al.*, 2017, Simmonds *et al.*, 2014 and Simmonds *et al.*, 2016, respectively. All NILs were evaluated at Church Farm in Norwich with exact numbers of NILs grown in each year outlined in Table 3.1. All NILs were grown in large-scale yield plots (1.1 × 6 m) and a randomized complete block design was used with five replications.

*Table 3.1: Summary of NILs grown in each year at Church Farm*

Year	5A NILs	6A NILs	<i>TaGW2-A</i> NILs
2012	10 BC <sub>2</sub>	-	-
2013	10 BC <sub>2</sub>	-	-
2014	12 BC <sub>4</sub>	7 BC <sub>4</sub>	-
2015	4 BC <sub>4</sub>	4 BC <sub>4</sub>	4 BC <sub>2</sub>
2016	4 BC <sub>4</sub>	4 BC <sub>4</sub>	4 BC <sub>4</sub>

##### 3.1.2 Phenotyping

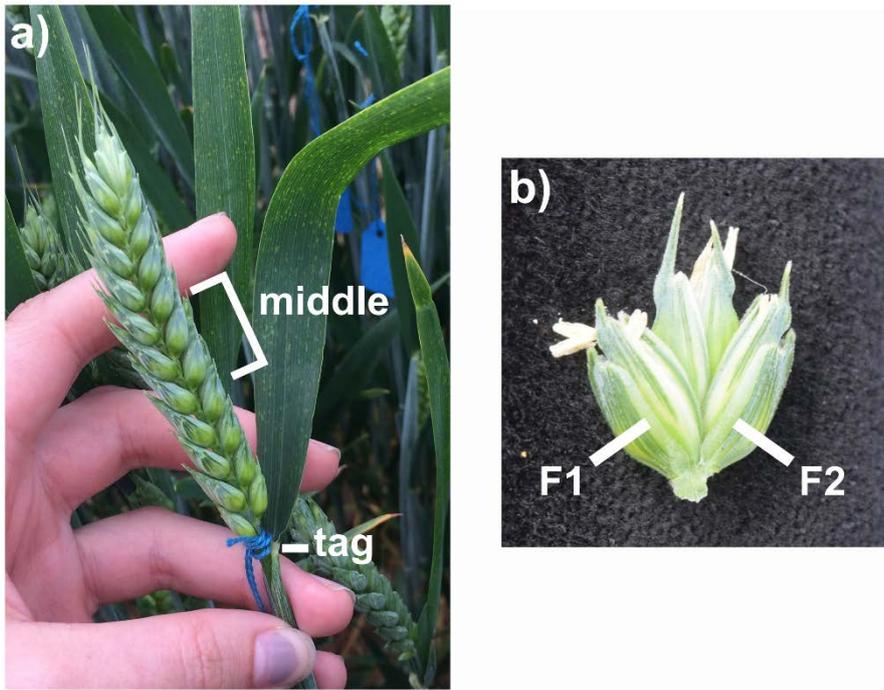
Grain morphometric measurements (grain width, length, area) and thousand grain weight (TGW) were recorded on the MARVIN grain image analyser (GTA Sensorik GmbH, Germany) using approximately 400 grains obtained from the harvested grain samples of each plot. The plot average was used in the statistical analyses. Individual grain data from each plot sample was also extracted to examine distributions of grain size in the 5A and 6A NILs. Final grain yield was adjusted by plot size and moisture content. Other spike yield components and developmental traits measured include:

- Spikelet number (all spikelets on the spike)
- Viable spikelets (all spikelets containing grains)
- Grain number per spike (Total grains from a single spike)
- Seeds per spikelet (Total grains per spike/number of viable spikelets)
- Spike yield (Total weight of all seeds per spike)
- Days to heading (days from sowing until 75% ear emergence of 75% of plot)
- Days to maturity (days from sowing until 75% plot senesced)
- Tiller number

- Crop height

### 3.1.3 Carpel/grain developmental time courses

For the 5A and 6A carpel/grain developmental time courses, BC<sub>4</sub> NILs grown in 2014-2016 were used. For both QTL, two independent NILs carrying the negative allele (2 x 5A<sup>-</sup> or 2 x 6A<sup>-</sup>) and two independent NILs carrying the positive QTL allele (2 x 5A<sup>+</sup> or 2 x 6A<sup>+</sup>) were used. The same NILs were used in all three years. For the *TaGW2-A* time courses, NILs grown in 2015 (BC<sub>2</sub>) and 2016 (BC<sub>4</sub>) were used. Again, two independent NILs were used for each genotype: two NILs carrying the wild type allele of *TaGW2-A* (2x *TaGW2-A*) and two NILs carrying the non-functional A-genome allele (2x *gw2-A*). In all experiments, 65 wheat inflorescences (referred to as ear or spike) per NIL were tagged across up to five blocks in the field at full ear emergence (peduncle just visible; Figure 3.1a) to ensure sampling at the same developmental stage. Ten spikes per NIL, per block, were sampled at each time point (i.e. 50 total spikes from the 65 tagged spikes). Exact time points taken are detailed in figure legends of the time courses (Figure 4.2, Figure 4.5). Spikes were kept on ice and taken to JIC for dissection. Ten carpels/grains were sampled from each spike from the outer florets (positions F1 and F2; Figure 3.1b) of spikelets located in the middle of the spike (Figure 3.1a) and placed in 2 mL Eppendorf tubes. Carpels/grains were weighed to obtain fresh weight (FW) and assessed for morphometric parameters (carpel/grain area, length and width) on the MARVIN grain image analyser. Measurements were taken within 3 hours of dissection from the spike and kept at 4°C in the intervening period to avoid water loss. Carpels/grains were then dried at 37 °C to constant weight (dry weight; DW). For each block at each time point, a total of ~100 carpels/grains were sampled (10 spikes per block x 10 carpels/grains per spike) per NIL. However, for the statistical analysis the average of the ~100 carpels/grain from each NIL within each block was used as the phenotypic value as the individual grains and spikes were considered as subsamples.



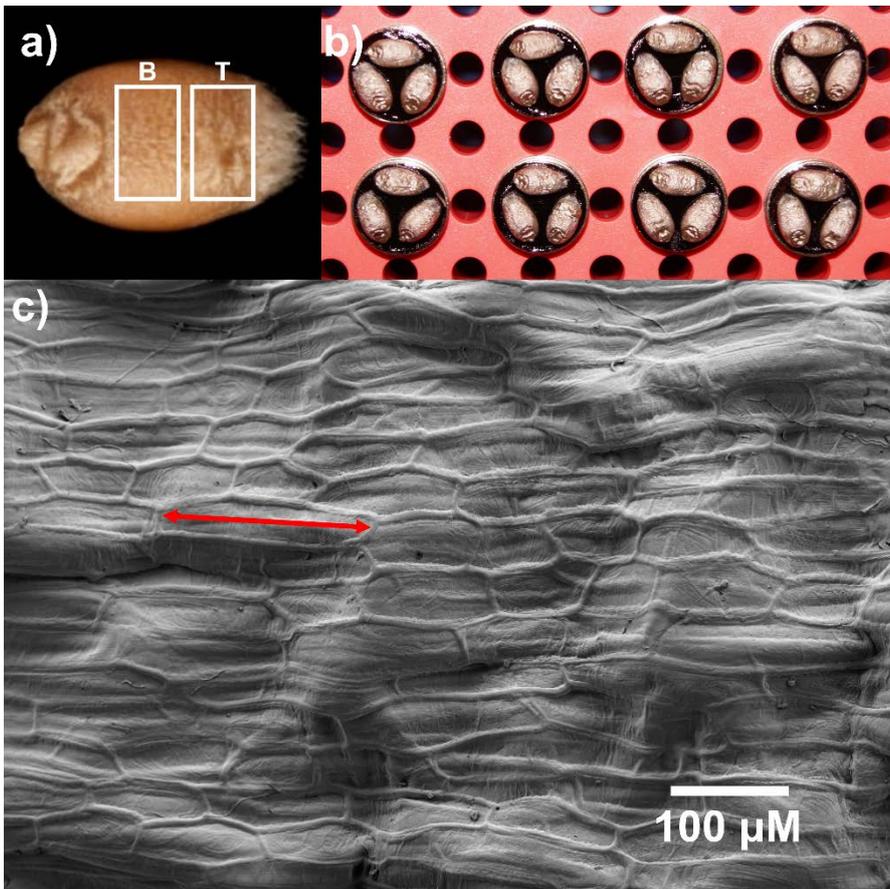
**Figure 3.1: Sampling strategy for the carpel/grain development time courses**

a) Spikes were tagged at full ear emergence (peduncle just visible). Grains were sampled from the middle of the spike. b) Grains were sampled from the outer florets of spikelets (floret 1 (F1) and floret 2 (F2)).

### 3.1.4 Cell size measurements

One representative 5A- and 5A+ BC<sub>4</sub> NIL was used for cell size measurements. This pair of NILs were selected based on the consistency of the grain length effect across previous years. For each NIL, nine grains of average grain length were selected from the whole harvest sample from each block (groups 5A-/5A+ average). For the 5A- NIL, an additional nine grains were selected that had grain lengths equivalent to the average of the 5A+ NIL sample (5A- large). For the 5A+ NIL, an additional nine grains were selected that had grain lengths equivalent to the average of the 5A- NIL sample (5A+ small). Grains of average length from three blocks of the 2016 harvest samples were also selected (nine grains from each block per genotype).

Grains were stuck crease-down on to 12.5 mm diameter aluminium specimen stubs using 12 mm adhesive carbon tabs (both Agar Scientific), sputter coated with gold using an Agar high resolution sputter coater (Figure 3.1b) and imaged using a Zeiss Supra 55 SEM. The surface (pericarp) of each grain was imaged in the top and bottom (embryo) half of the grain (Figure 3.1a, T and B, respectively), with images taken in at least three positions in each half. All images were taken at a magnification of 500x. Cell length was measured manually using the Fiji distribution of ImageJ (Schindelin *et al.*, 2012) (Figure 3.2c). Cell number was estimated for each grain using average cell length/grain length. For the statistical analyses, the average cell length of each individual grain was considered as a subsample within the block.



**Figure 3.2: Scanning Electron Microscopy imaging for pericarp cell size measurements**

a) Example grain showing the bottom (B) and top (T) half as used for imaging. b) Grains stuck crease down and sputter-coated with gold to be imaged. c) Example scanning electron microscopy image taken for cell size measuring. Red arrow indicates how cell length was measured. Image taken at 500x magnification.

### 3.1.5 Statistical analysis

The NILs were evaluated using two-way ANOVAs across all years with the model including the interaction between environment and the genotype. For the evaluation of individual years the block and genotype were included in the model. Similarly, two-way ANOVAs, including genotype and block, were conducted for the developmental time courses and cell size measurements. Analyses were performed using R v3.2.5.

## 3.2 Fine mapping of the 5A and 6A QTL

### 3.2.1 Plant material and growth

The 6A recombinant populations used in this chapter were generated by James Simmonds alongside the development of 6A NILs, described in Simmonds *et al.*, 2014. For the original 6A population, 212 BC<sub>4</sub>F<sub>2</sub> plants were screened for recombination between markers *gwm334* and *gwm570*, encompassing the 6A genetic map developed during the initial identification of the 6A QTL (Simmonds *et al.*, 2014). 67 recombinants were identified and self-pollinated to generate homozygous BC<sub>4</sub>F<sub>3</sub> recombinant inbred lines (RILs). The larger 6A RIL population was generated

within this PhD in the same way, but screening a larger number of BC<sub>4</sub>F<sub>2</sub> plants (2,674). These plants were screened for recombination between a narrower marker interval (*BS00010933-BS00066623*) identifying 892 recombinants. Further development of this population was carried out during the PhD and is therefore described in the results section.

The 5A RIL populations used in this chapter were also generated by James Simmonds alongside 5A NIL development, described in Brinton *et al.*, 2017. Screening of 170 BC<sub>4</sub>F<sub>2</sub> plants identified 60 recombinants between *gwm293* and *gwm186*, the markers used for the selection of NILs.

Recombinant plants were self-pollinated to develop homozygous BC<sub>4</sub>F<sub>3</sub> RILs. The larger 5A RIL population was developed in the same way, but screening a larger number of BC<sub>4</sub>F<sub>2</sub> plants (1,140) and using a slightly narrower marker interval (*BS00075504* and *BS00183958*). 310 recombinant plants were identified. Again, further development of this population was carried out during the PhD and is described in the results section.

All RIL populations were evaluated at Church Farm in Norwich (52.628 N, 1.171 E). Subsets of the original 6A RIL population were evaluated in five trials across four years: large-scale yield plots (1.1 x 6m) in 2013-2015 and an additional trial of 1.1 x 1m plots in 2015. In all five trials, a randomised complete block design was used with at least five replications. The exact 6A RILs used in each trial are detailed in Table 3.1 (see Results section). The larger 6A RIL population was evaluated in 2016. RILs were grown in single 1m rows with up to three replications depending on seed availability. Subsets of the original 5A population were evaluated in four trials across three years: 1.1 x 1m plots in 2014 and 2015 and 1.1 x 6m plots in 2015 and 2016. In all four trials, a randomised complete block design was used with at least five replications. The exact details of 5A RILs assessed in each trial are outlined in Table 4.11 (see Results section). The larger 5A RIL population was evaluated in 2016. RILs were grown in single 1m rows, replicated up to five times depending on seed availability.

### **3.2.2 Grain phenotyping**

Grain morphometric measurements (grain width, length, area) and thousand grain weight (TGW) were recorded on the MARVIN grain image analyser (GTA Sensorik GmbH, Germany). For all full plots (1.1 x 6m and 1.1 x 1m) approximately 400 grains obtained from the combine harvested grain samples were used. For single rows, ten representative spikes were harvested from each row. The ten spikes were threshed together and the grains obtained from these samples were assessed.

### **3.2.3 Marker development**

Genetic maps were available for both the 6A and 5A original RIL populations at the start of the PhD (details in Simmonds *et al.*, 2014; Brinton *et al.*, 2017). However, these did not provide sufficient marker density across the intervals of interest and additional markers were developed. With the exception of a single marker, SNP markers used to genotype the RIL populations fall into four categories (BS, BA, JB\_RNASeq and JBHap markers) which are described below.

### 3.2.3.1 BS and BA markers

BS (Bristol SNP) markers were developed based on data from 90K iSelect array genotyping of BC<sub>4</sub> 6A and 5A NILs (Simmonds *et al.*, 2014; Brinton *et al.*, 2017). BA (Bristol Axiom) markers were developed based on data from 820k Axiom array genotyping of parental varieties of the QTL: Spark (6A-), Rialto (6A+), Charger (5A-) and Badger (5A+) (Winfield *et al.*, 2016). KASP primers for all SNPs in the iSelect and Axiom arrays have been designed previously by Ricardo Ramirez-Gonzalez using Polymarker (Ramirez-Gonzalez *et al.*, 2015) and are publicly available at <http://polymarker.tgac.ac.uk/>. Initially, BS and BA markers across the 6A and 5A QTL intervals were selected based on the predicted genetic positions of markers (Chapman *et al.*, 2015). However, with the release of more contiguous genome assemblies, markers were selected based on their physical positions across the intervals with respect to the reference sequence (details of how markers were positioned are below (3.2.4)).

### 3.2.3.2 JB\_RNASeq markers

JB\_RNASeq (Jemima Brinton RNASeq) markers used to genotype the 5A RILs were designed using RNA-seq data from a pair of 5A NILs. Twelve RNA samples from grains were sequenced: one 5A- and one 5A+ NIL, each at two time points and with three biological replicates. The RNA-seq experiment and detailed methods including RNA extraction and sequencing are described in detail in the next chapter. Specifically for the SNP identification, RNA-seq reads were aligned to the Chinese Spring Chromosome Survey Sequence cDNA reference (CSS; IWGSC, 2014) downloaded from *Ensembl* plants release 29. Read alignment was performed using kallisto-0.42.3 (Bray *et al.*, 2016) with default parameters, 30 bootstraps (-b 30) and the -pseudobam option. Pseudobam files for each genotype (5A- and 5A+) were merged to generate a single BAM file for each genotype. SNP calling with respect to Chinese Spring was performed using the samtools-0.1.19 mpileup command followed by the bcftools-1.2 call command (Li *et al.*, 2009). Samtools mpileup was used with the -Agf options: -A includes improperly paired reads, -g computes the genotype likelihoods and outputs them in binary call format (BCF) and -f specifies a reference fasta file. The bcftools call command was used with -O u (to give an uncompressed output, essential for downstream processing) and -c (to call SNPs using Bayesian inference) options. BCF files were converted to variant call format (VCF) using bcftools view and VCF files were filtered with Samtools vcfutils.pl using -d 10 -a 9 options to output SNPs with a minimum read depth of 10 and a minimum alternate read number of 9. A grep command was used to extract only SNPs with an allele frequency of 1 ('AF1=1') to filter for homozygous SNPs only. SNPs located in the 5A mapping interval were extracted and compared between genotypes to identify SNPs that were unique to either the 5A- or 5A+ NIL. This identified 145 SNPs between NILs in 34 gene models. However, after manual inspection of BAM files, only SNPs in four of the genes looked to be real. Common reasons for discarding SNPs included small regions of mis-mapping or the SNP being present in both NILs but filtered out of the output for one NIL due to low read depth. All four SNPs were

validated experimentally using KASP assays (designed using polymarker (Ramirez-Gonzalez *et al.*, 2015) which were subsequently used as markers JBRNA\_Seq1-4 (Appendix 2). JBRNA\_Seq1, 2 and 4 were predicted to be non-synonymous SNPs resulting in missense mutations in the 5A-NIL. The three genes (1: Traes\_5AL\_6401EFD6F, 2: Traes\_5AL\_AEB344EBB, 3 Traes\_5AL\_632F49251) were predicted to encode a TATA binding protein, an Fe-S cluster protein and P-loop NTPase, respectively.

### **3.2.3.3 JBHap markers**

The JBHap (Jemima Brinton Haplotype) markers were developed based on the haplotype analysis conducted across the 5A interval (described below). KASP assays were designed for 22 SNPs defining haplotypes across the 5A grain length mapping interval (Appendix 2).

### **3.2.3.4 Hap-P2 marker**

Hap-P2 is an A/G SNP at the -593 bp position in the promoter of *TaGW2\_A* and the original marker was designed as a cleaved amplified polymorphism sequence (CAPS) marker by Su *et al.*, 2011. For ease of genotyping, a KASP assay for the Hap-P2 SNP was designed using Polymarker and used to genotype the 6A RIL populations (Appendix 2).

## **3.2.4 Physical positions**

To obtain physical locations, SNPs were positioned with respect to the recently released Chinese Spring sequence (International Wheat Genome Sequencing Consortium (IWGSC) RefSeq v.1.0; <https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies>). Physical positions of all iSelect and Axiom SNPs were obtained using BLASTN (Altschul *et al.*, 1990) to align the surrounding sequence (201 bp) to the RefSeq v.1.0 assembly, provided by Ricardo Ramirez-Gonzalez and available at <http://www.wheat-training.com/useful-wheat-links/>. The positions of all additional SNPs were determined in a similar way by using BLASTN to align 100-300 bp of surrounding sequence to RefSeq v1.0. Positions of TGACv1 gene models in RefSeq v.1.0 were obtained using GMAP (Wu & Watanabe, 2005) retaining the best hit position and using a 95% minimum similarity cut-off (David Swarbreck and Gemy Kaithakottil, Earlham Institute).

## **3.2.5 DNA extraction and KASP genotyping**

DNA extraction and KASP genotyping were performed as previously described (Pallotta *et al.*, 2003; Trick *et al.*, 2012).

## **3.2.6 Exome capture for haplotype analysis**

Exome capture data for 20 UK wheat varieties was provided by Philippa Borrill. Alignment of data and SNP calling with respect to the CSS reference (IWGSC, 2014) were also performed by Philippa Borrill. Briefly, reads were aligned to the CSS reference using bowtie2 with the very-sensitive-local option (Langmead & Salzberg, 2012) followed by SNP calling using freebayes (Garrison & Marth, 2012) with the following options: --use-best-n-alleles 2 (only allow sites with up

to two alleles), --min-mapping-quality 7 (only use reads with MAPQ>7) and --min-base-quality 20 (only use bases with quality > 20). Details of how SNPs defining haplotypes across the 5A grain length interval were identified are detailed in the Results section. The position of SNPs with respect to the IWGSC RefSeq v1.0 were determined as described above (3.2.4).

### **3.2.7 Statistical analysis**

RILs were evaluated using two-way ANOVAs. For the original 6A and 5A RIL populations, the model included the trial as a factor in the model. When individual trials were evaluated the field block (replicate) was included as a factor in the model. Similarly, for the larger RIL populations (assessed in a single trial) the field block was included as a factor in the model. When RIL groups were assessed, independent RILs within each group were considered as replicates within the model. For the larger RIL populations, individual RILs belonging to a single family were considered as replicates of a single independent RIL. RIL groups were assigned to parental genotypes using a *post hoc* Dunnett's test to compare with control groups. The specific control groups used for each comparison are described in the results section. All statistical analyses were performed using Minitab® Statistical Software.

## **3.3 Comparative transcriptomics of 5A NILs**

### **3.3.1 Plant material**

The 5A BC<sub>4</sub> NILs used in this chapter were characterised and have been described previously (Brinton *et al.*, 2017). One genotype each for the 5A- (Charger allele, short grains) and 5A+ NIL (Badger allele, long grains) were used (the same NIL pair as used for the cell size measurements). Plants were sampled at 4 (time point 1: T1) and 8 (time point 2: T2) days post anthesis (dpa) during the 2014 developmental time course outlined in (Brinton *et al.*, 2017). Briefly, plants were grown in 1.1 x 6 m plots (experimental units) in a complete randomised block design with five replications, and spikes were tagged at full ear emergence. The three blocks with the most similar flowering time were used for sampling. For each genotype, three grains from three separate spikes from different plants within the experimental unit were sampled. Each biological replicate, therefore, consisted of the pooling of nine grains per genotype. Grains were sampled from the outer florets (positions F1 and F2) from the middle section of each of the three spikes. Grains were removed from the spikes in the field, immediately frozen in liquid nitrogen and stored at -80°C. In total, three biological replicates (from the three blocks in the field) were sampled for each NIL at each time point.

### **3.3.2 RNA extraction and sequencing**

For each biological replicate, the nine grains were pooled and ground together under liquid nitrogen. RNA was extracted in RE buffer (0.1 M Tris pH 8.0, 5 mM EDTA pH8.0, 0.1 M NaCl, 0.5% SDS, 1% β-mercaptoethanol) with Ambion Plant RNA Isolation Aid (Thermo Fisher

Scientific). The supernatant was extracted with 1:1 acidic Phenol (pH 4.3):Chloroform. RNA was precipitated at -80°C by addition of Isopropanol and 3M NA Acetate (pH 5.2). The RNA pellet was washed twice in 70% Ethanol and resuspended in RNase-free water. RNA was DNase treated and purified using RNeasy Plant Mini kit (Qiagen) according to the manufacturer's instructions. RNA QC, library construction and sequencing were performed by the Earlham Institute, Norwich. Library construction was performed on a PerkinElmer Sciclone using the TruSeq RNA protocol v2 (Illumina 15026495 Rev.F). Libraries were pooled (2 pools of 6) and sequenced on 2 lanes of a HiSeq 2500 (Illumina) in High Output mode using 100bp paired end reads and V3 chemistry. Initial quality assessment of the reads was performed using fastQC (Andrews, 2010).

### **3.3.3 Read alignment and differential expression analysis**

Reads were aligned to two reference sequences from the same wheat variety, Chinese Spring: the Chromosome Survey Sequence (CSS; (IWGSC, 2014) downloaded from *Ensembl* plants release 29) and the TGACv1 reference sequence (Clavijo *et al.*, 2017b). Read alignment and expression quantification were performed using kallisto-0.42.3 (Bray *et al.*, 2016) with default parameters, 30 bootstraps (-b 30) and the -pseudobam option. Kallisto has previously been shown to be suitable for the alignment of wheat transcriptome data in a homoeolog specific manner (Borrill *et al.*, 2016). Differential expression analysis was performed using sleuth-0.28.0 (Pimentel *et al.*, 2017) with default parameters. Transcripts with a false-discovery rate (FDR) adjusted p-value (q value) < 0.05 were considered as differentially expressed. Transcripts with a mean abundance of < 0.5 tpm (transcripts per million) in all four conditions were considered not expressed and were therefore, excluded from further analyses.

For each condition, the mean tpm of all three biological replicates was calculated. All heatmaps display mean expression values as normalised tpm, on a scale of 0 to 1, with 1 being the highest expression value of the transcript. Read coverage for gene models was obtained using bedtools-2.24.0 genome cov (Quinlan & Hall, 2010) for each pseudobam file and then combined to get a total coverage value of each position. Coverage across a gene model was plotted as relative coverage on a scale of 0 to 1, with 1 being equivalent to the highest level of coverage for the gene model in question.

### **3.3.4 GO (gene ontology) term enrichment**

The R package Goseq v1.26 was used (Young *et al.*, 2010) to test for enrichment of GO terms in specific groups of DE (differentially expressed) transcripts. Over-represented GO terms with a Benjamini Hochberg FDR adjusted p-value of < 0.05 were considered to be significantly enriched.

### **3.3.5 Functional annotation**

Functional annotations of transcripts were obtained from the TGACv1 annotation (Clavijo *et al.*, 2017b). Additionally, for coding transcripts BLASTP against the non-redundant NCBI protein database and conserved domain database were performed, in each case the top hit based on e-

value was retained. In cases where all three annotations were in agreement, the TGAC annotation is reported. In cases where the three annotations produced differing results, all annotations are reported. Orthologues in other species such as *Arabidopsis* and rice were obtained from *Ensembl* plants release 36. Eight of the 112 DE transcripts had no annotation or protein sequence similarity with other species. The remaining 104 DE transcripts were manually categorised based on their predicted function. Transcripts that fell into a category of size 1 were classed as 'other'. For the non-coding transcripts, BLASTN was used to identify potential miRNA precursors using a set of conserved and wheat specific miRNA sequences obtained from Sun *et al.*, 2014. The -task blastn-short option of BLAST for short sequences was used and only hits of the full length of the miRNA sequence with no mismatches as were considered as potential precursors. The psRNAtarget tool (<http://plantgrn.noble.org/psRNAtarget/>) was used to determine the miRNA targets.

### 3.3.6 Identification of transcription factor binding sites

1,000 bp of sequence upstream of the cDNA start site was extracted to search for transcription factor binding sites (TFBS). Transcripts with < 1,000 bp upstream in the reference sequence were not used in the analysis. The FIMO tool from the MEME suite (v 4.11.4; (Grant *et al.*, 2011)) was used with a position weight matrix (PWM) obtained from plantPAN 2.0 (<http://plantpan2.itps.ncku.edu.tw/>; (Chow *et al.*, 2016)). FIMO was run with a p value threshold of <1e-4 (default), an increased max-stored-scores of 1,000,000 to account for the size of the dataset, and a -motif-pseudo of 1e-8 as recommended for use with PWMs (Peng *et al.*, 2016). The background model was generated using the fasta-get-markov command of MEME on all extracted promoter sequences.

### 3.3.7 Enrichment testing

Fisher's exact test was performed to test for enrichment of different categories of transcripts relative to all expressed transcripts using R-3.2.5. For functional annotation categories, enrichment testing was only performed on categories that could be extracted using GO terms and key words based on their annotation in the TGAC reference. Only DE transcripts that could be extracted using this method were used in the enrichment tests. For example, 12 DE transcripts identified were associated with ubiquitin. The annotation of these transcripts was obtained through a combination of the TGAC annotation and manual annotation. However, only seven of these transcripts could be extracted using GO terms and key words from the whole reference annotation. Therefore, only seven transcripts were used for the enrichment test.

## 4 Results

### 4.1 Characterisation of 6A and 5A Near Isogenic Lines

#### 4.1.1 6A QTL for grain width

##### 4.1.1.1 6A NILs have a 4.4% difference in TGW

Across three years of replicated field trials, 6A+ NILs had significantly increased TGW compared with 6A- NILs (4.39%;  $P < 0.001$ ; Table 4.1), ranging from 1.38% to 7.42% in individual years. However, when years were analysed individually, the increase in TGW was non-significant in 2016 (1.38%,  $P = 0.33$ ). Across all three years, the increase in TGW was associated with a 2.25% increase in plot yield, although this was non-significant across years ( $P = 0.42$ ) and in each year individually (Table 4.1).

*Table 4.1: Mean Thousand Grain Weight (TGW), yield and grain morphometric parameters of 6A NILs*

Year	Genotype	TGW (g)	Yield (kg/plot)	Grain area (mm <sup>2</sup> )	Grain length (mm)	Grain width (mm)
2014	6A-	43.20	5.98	19.49	6.34	3.82
	6A+	45.29	6.08	20.01	6.37	3.91
		4.84% <sup>***</sup>	0.46% <sup>NS</sup>	2.65% <sup>***</sup>	0.48% <sup>NS</sup>	2.27% <sup>***</sup>
2015	6A-	38.22	6.66	15.66	5.94	3.26
	6A+	41.06	6.97	16.34	5.99	3.37
		7.42% <sup>***</sup>	4.71% <sup>NS</sup>	4.37% <sup>***</sup>	0.77% <sup>NS</sup>	3.35% <sup>***</sup>
2016	6A-	45.14	5.75	20.06	6.37	3.93
	6A+	45.77	5.76	20.29	6.33	3.99
		1.38% <sup>NS</sup>	0.08% <sup>NS</sup>	1.13% <sup>NS</sup>	-0.62% <sup>NS</sup>	1.49% <sup>*</sup>
Overall	6A-	42.19	6.13	18.40	6.22	3.67
	6A+	44.04	6.27	18.88	6.23	3.76
		4.39% <sup>***</sup>	2.25% <sup>NS</sup>	2.58% <sup>***</sup>	0.20% <sup>NS</sup>	2.31% <sup>***</sup>

<sup>1</sup> %s indicate amount gained in 6A+ NILs compared with 6A- NILs. Superscripts indicate significance determined by ANOVA for either each year, or across all years (final row) i.e. NS = Non-significant, \* =  $P < 0.05$ , \*\*\* =  $P < 0.001$

To identify potential pleiotropic effects of the QTL that could account for the absence of a significant yield effect, ten representative spikes from each plot of 6A NILs were assessed for a series of spike yield components (Table 4.2). Components included spikelet number, seeds per spikelet, grain number per spike and spike yield, although not all measurements were taken in all three years. In 2014, grain samples were compromised due to high levels of bunt infection at Church Farm and in 2016, the spikelet number counting was performed incorrectly. Across two years, there was significant decrease in the number of viable spikelets in 6A+ NILs (-2.17% equivalent to 0.44 spikelets per spike;  $P = 0.001$ ), although this was driven by a strong effect in 2014. The decrease in viable spikelet number (and spikelet number overall) appears to have been compensated for in 2015 by an increase in the number of seeds per spikelet (3.98%) which resulted in one extra grain per spike (2.86%) in 6A+ NILs, although neither were significant (Table 4.2). Unfortunately, no grain number data is available for 2014 (when spikelet number was significantly reduced) so it is unclear whether this would have resulted in a significant reduction in grain number. Despite the fact that the grain number differences were not significant in 2015, the tendency towards more grains per spike, combined with 8.7% higher TGW ( $P < 0.001$ ) in the ten spike sample, resulted in significantly higher spike yield in 6A+ NILs (11.9%,  $P = 0.001$ ). This translated into a higher overall plot yield (4.7%; Table 4.1), although this was not statistically significant.

**Table 4.2: Spike yield components of ten representative single ear samples (SES) of 6A- and 6A+ BC<sub>4</sub> NILs**

Year	Genotype	Spikelet number	Viable Spikelets	Grain number per spike	Spike yield (g/spike)	Seeds per spikelet	SES-TGW (g)	SES-Grain Area (mm <sup>2</sup> )	SES-Grain length (mm)	SES-Grain width (mm)
2014	6A-	22.12	21.00	-	-	-	-	-	-	-
	6A+	21.24	20.35	-	-	-	-	-	-	-
		-3.98% <sup>***</sup>	-3.10% <sup>**</sup>							
2015	6A-	21.53	19.15	33.52	1.265	1.75	37.80	19.11	6.39	3.71
	6A+	21.13	18.93	34.48	1.416	1.82	41.09	19.89	6.41	3.85
		-1.86% <sup>NS</sup>	-1.15% <sup>NS</sup>	2.86% <sup>NS</sup>	11.90% <sup>**</sup>	3.98% <sup>NS</sup>	8.71% <sup>***</sup>	4.06% <sup>**</sup>	0.31% <sup>MS</sup>	3.65% <sup>**</sup>
Overall	6A-	21.83	20.08	-	-	-	-	-	-	-
	6A+	21.19	19.64	-	-	-	-	-	-	-
		-2.94% <sup>***</sup>	-2.17% <sup>**</sup>							

%s indicate amount gained in 6A+ NILs compared with 6A- NILs. Superscripts indicate significance determined by ANOVA for either each year, or across all years (final row). i.e. NS = Non-significant, \*\* < 0.01, \*\*\* < 0.001. SES = Single Ear Samples

6A NILs were also measured in the field for a series of developmental traits. 6A+ NILs flowered c. one day earlier than 6A- NILs across years (measured as days to heading;  $P = 0.01$ ; Table 4.3). 6A+ NILs also senesced c. one day later than 6A- NILs (measured as days to maturity;  $P < 0.006$ ) although this was only measured in a single year (Table 4.3). No consistent significant effects were observed across years for crop height and tiller number, although there was a significant reduction in tiller number in 6A+ NILs in 2014. These results suggest that the 6A QTL acts to increase TGW in a stable manner across years, but the effects on final yield may be modulated by environmental interactions and negative effects on components such as spikelet number and tiller number.

**Table 4.3: Developmental traits of 6A BC<sub>4</sub> NILs**

Year	Genotype	Days to heading	Days to maturity	Tiller number	Crop Height (cm)
2014	6A-	243.80	296.00	82.10	75.70
	6A+	242.95	296.70	76.53	75.13
		-0.85**	0.70**	-5.57**	-0.58 <sup>NS</sup>
2015	6A-	250.90	NA	133.75	84.75
	6A+	250.00	NA	136.10	83.75
		-0.90*		2.35 <sup>NS</sup>	-1.00 <sup>NS</sup>
2016	6A-	250.38	NA	NA	NA
	6A+	249.89	NA	NA	NA
		-0.49 <sup>NS</sup>			
Overall	6A-	248.36	NA	107.93	80.23
	6A+	247.61	NA	106.31	79.44
		-0.75*		-0.75 <sup>NS</sup>	-0.75 <sup>NS</sup>

Differences indicate amount gained in 5A+ NILs compared with 5A- NILs. Superscripts indicate significance determined by ANOVA for either each year, or across all years (final row) i.e. NS = Non-significant, \* =  $P < 0.05$ , \*\* =  $P < 0.01$

#### **4.1.1.2 Grain width underlies the increase in TGW in 6A+ NILs**

Grain morphometric parameters (grain area, length and width) of 6A NILs were measured to understand the contribution of the individual components to the overall increase in TGW (Table 4.1). 6A+ NILs had significantly increased grain area ( $P < 0.001$ ) and grain width ( $P < 0.001$ ) compared to 6A- NILs. No significant grain length differences were observed in any year. 6A+ NILs had 2.31% wider grains across all years ranging from 1.49-3.35% in individual years. Grain area differences ranged from 1.13 – 4.37 %, although the difference was non-significant in 2016, reminiscent of the non-significant TGW increase in 2016. These results were based on whole plot samples and were confirmed in ten representative ear samples taken before harvest (Table 4.2). The absence of any significant grain length effect suggests that grain width is the main factor underlying the increase in grain area and TGW in 6A+ NILs.

#### **4.1.1.3 The 6A QTL affects grains uniformly within the spike**

Distributions of grain width were compared between 6A NILs using measurements from individual seeds to determine whether the 6A QTL has a uniform effect on all grains within the spike. Violin plots of grain width showed some variation in distribution shape between years (Figure 4.1). However, distribution shapes within years were very similar between 6A- and 6A+ NILs suggesting that the QTL has a uniform and stable effect across the whole spike and within spikelets. In all years, the 6A+ distributions were shifted higher reflecting the higher average grain width and illustrating the fact that 6A+ NILs had both larger numbers of wider grains and fewer thinner grains than 6A- NILs. Note that individual distributions are not completely normally distributed since the plots are based on the multiple independent NILs used for each genotype.



**Figure 4.1: Distribution of grain width of 6A NILs from whole plot samples**

Violin plots showing the distribution of individual seed measurements of grain width across three field experiments of BC<sub>4</sub> 6A near isogenic lines (NILs). Orange plots = 6A+ NILs, grey = 6A- NILs. All within year comparisons were significant (2014, 2015:  $P < 0.001$ ; 2016:  $P = 0.03$ ).

#### **4.1.1.4 GW2-A NILs show phenotypic differences compared to 6A NILs**

The A genome copy of the E3 ubiquitin ligase *TaGW2* (*TaGW2-A*) genetically mapped to the original 6A QTL region (Simmonds *et al.*, 2014) and has previously been associated with the control of grain size (Su *et al.*, 2011; Bednarek *et al.*, 2012; Yang *et al.*, 2012; Zhang, X *et al.*, 2013; Hong *et al.*, 2014). Therefore, we hypothesised that *TaGW2-A* could be a candidate gene for the 6A QTL (as discussed in the introduction). To test this hypothesis phenotypically, *TaGW2-A* NILs were assessed for grain weight and morphometric parameters and carpel/grain development time courses were conducted in 2015 and 2016. The results from 2015 have been published in Simmonds *et al.* (2016).

**4.1.1.4.1** GW2-A NILs have 6.7% higher TGW, driven by both grain length and width. Across two years of field trials, *gw2-A* (mutant) NILs had 6.65% ( $P < 0.001$ ) higher TGW than *GW2-A* (WT) NILs, ranging from 6.17-7.11% in each year. This was larger than the TGW differences observed between 6A NILs across years (4.4% higher TGW in 6A+ NILs across three years, 4.2 % in 2015-2016; Table 4.1). Similarly to the 6A NILS, no significant differences in yield were observed in either year.

Across years, *gw2-A* (mutant) NILs had significantly increased grain length (1.74%,  $P < 0.001$ ; Table 4.4), which combined to give a 3.57% ( $P < 0.001$ ) increase in grain area compared to *GW2-A* (WT) NILs. This was in contrast to 6A NILs, which showed significant differences in grain width and area, but no significant differences in grain length in each of the three years tested (Table 4.1). This would support the hypothesis that the 6A effect is distinct from *TaGW2-A*.

Given that the differences in grain length and grain width between *TaGW2-A* NILs were of a similar magnitude, these results suggest that the increase in TGW in *gw2-a* (mutant) NILs is driven by a combination of increases in both grain width and grain length.

**Table 4.4: Mean Thousand Grain Weight (TGW), yield and grain morphometric parameters of GW2-A NILs**

Year	Genotype	TGW (g)	Yield (kg/plot)	Grain area (mm <sup>2</sup> )	Grain length (mm)	Grain width (mm)
2015	GW2-A (WT)	43.869	5.195	20.275	6.698	3.699
	gw2-A (Mut)	46.573	5.328	20.909	6.785	3.767
		6.17% <sup>***</sup>	2.56% <sup>NS</sup>	3.13% <sup>**</sup>	1.30% <sup>*</sup>	1.84% <sup>**</sup>
2016	GW2-A (WT)	45.859	5.612	21.058	6.676	3.896
	gw2-A (Mut)	49.118	5.642	21.901	6.822	3.975
		7.11% <sup>***</sup>	0.53% <sup>NS</sup>	4.00% <sup>***</sup>	2.18% <sup>***</sup>	2.03% <sup>**</sup>
Overall	GW2-A (WT)	44.864	5.404	20.666	6.687	3.797
	gw2-A (Mut)	47.846	5.485	21.405	6.804	3.871
		6.65% <sup>***</sup>	1.51% <sup>NS</sup>	3.57% <sup>***</sup>	1.74% <sup>***</sup>	1.94% <sup>**1</sup>

<sup>1</sup> %s indicate amount gained in *gw2-a* (mutant) NILs compared with *GW2-A* (WT) NILs. Superscripts indicate significance determined by ANOVA for either each year, or across all years (final row). ie. NS = Non-significant, \* = P < 0.05, \*\* = P < 0.01, \*\*\* = P < 0.001. 2015 = BC<sub>2</sub>-NILs, 2016 = BC<sub>4</sub>-NILs.

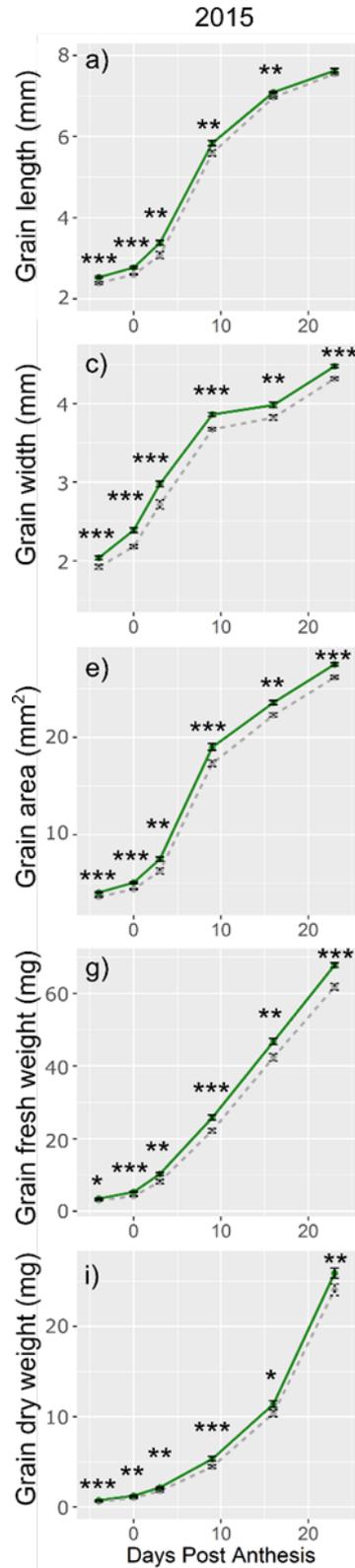
#### 4.1.1.4.2 *TaGW2-A* acts before fertilisation

Time courses of carpel/grain development were conducted in 2015. In 2015, samples were taken at -4, 0, 3, 9, 16 and 23 dpa. In 2015, *gw2-A* (Mutant) NILs had significantly increased carpel length (5.5%), width (6.2%) and area (12.1%) at the first time point (-4 dpa) (Table 4.5; Figure 4.2). These differences were maintained for the duration of the time course with the exception of carpel/grain length, which became non-significant by the final time point (P = 0.267). The differences in carpel/grain size components translated to increases in both carpel/FW and DW, which were significantly increased in *gw2-A* (Mutant) NILs across the whole time course. *TaGW2-A* acts before anthesis to influence carpel length and width, which combine to modulate carpel area and weight, and ultimately final grain weight.

**Table 4.5: Differences between *TaGW2-A* NILs during carpel/grain development time courses**

Year	Days Post Anthesis	Length (%)	Width (%)	Area (%)	FW (%)	DW (%)
2015	-4	5.55% <sup>&lt;0.001</sup>	6.22% <sup>&lt;0.001</sup>	12.13% <sup>&lt;0.001</sup>	14.22% <sup>0.02</sup>	27.99% <sup>&lt;0.001</sup>
	0	6.66% <sup>&lt;0.001</sup>	9.58% <sup>&lt;0.001</sup>	15.13% <sup>&lt;0.001</sup>	27.34% <sup>&lt;0.001</sup>	24.49% <sup>0.002</sup>
	3	9.81% <sup>0.004</sup>	9.43% <sup>&lt;0.001</sup>	18.95% <sup>0.002</sup>	25.41% <sup>0.005</sup>	20.90% <sup>0.003</sup>
	9	4.23% <sup>0.005</sup>	5.03% <sup>&lt;0.001</sup>	9.52% <sup>&lt;0.001</sup>	15.27% <sup>&lt;0.001</sup>	18.86% <sup>&lt;0.001</sup>
	16	1.78% <sup>0.005</sup>	4.11% <sup>0.001</sup>	5.61% <sup>0.001</sup>	10.00% <sup>0.002</sup>	9.48% <sup>0.013</sup>
	23	1.02% <sup>0.267</sup>	3.64% <sup>&lt;0.001</sup>	5.09% <sup>&lt;0.001</sup>	9.78% <sup>&lt;0.001</sup>	7.66% <sup>0.003</sup>

%s indicate amount gained in *TaGW2-A* (WT) NILs compared with *gw2-A* (Mutant) NILs. Superscripts are the ANOVA P values of the comparison between *TaGW2-A* (WT) and *gw2-A* (Mutant) NILs.



**Figure 4.2: Carpel/grain development time course of GW2-A NILs**

Grain/carpel length (a), width (c), area (e), fresh weight (g) and dry weight (i) of GW2-A (WT; grey, dashed line) and gw2-a (mutant; green, solid line) BC<sub>2</sub> near isogenic lines (NILs) during grain/carpel development in 2015 field trials. 2015 samples: -4, 0 (anthesis), 3, 9, 16 and 23 dpa; \* = P < 0.05, \*\* = P < 0.01, \*\*\* = P < 0.001. Error bars show one standard error above and below the mean.

#### **4.1.2 5A QTL for grain length**

All results described here relating to the 5A QTL have been published in Brinton *et al.*, 2017. Results from the BC<sub>2</sub> NILs (2012-2013) were obtained by James Simmonds prior to the start of the PhD, but had not been previously published. These results were, therefore, analysed alongside results obtained during the PhD (BC<sub>4</sub> NILs, 2014-2016).

##### **4.1.2.1 5A NILs have a 6.9% difference in TGW**

Across five years of replicated field trials, 5A+ NILs showed an average increase in TGW of 6.92% ( $P < 0.001$ ) ranging from 4.00 to 9.28% (Table 4.6), and significant in all years. The difference in TGW was associated with a yield increase of 1.28% in 5A+ NILs across all years, although this effect was not significant ( $P = 0.093$ ). The effect varied across years with a significant yield increase of 2.17% ( $P = 0.046$ ) in 2014 and non-significant effects of between 0.02 to 1.72% in the other four years.

**Table 4.6: Mean Thousand Grain Weight (TGW), yield and grain morphometric parameters of 5A NILs**

Year	Genotype	TGW (g)	Yield (kg/plot)	Grain area (mm <sup>2</sup> )	Grain length (mm)	Grain width (mm)
2012	5A-	38.027	4.408	18.755	6.625	3.475
	5A+	41.554	4.437	19.930	6.900	3.557
		9.28% <sup>***</sup>	0.66% <sup>NS</sup>	6.26% <sup>***</sup>	4.15% <sup>***</sup>	2.35% <sup>**</sup>
2013	5A-	40.772	6.157	19.969	6.705	3.674
	5A+	43.544	6.159	20.979	6.963	3.727
		6.80% <sup>***</sup>	0.02% <sup>NS</sup>	5.06% <sup>***</sup>	3.86% <sup>***</sup>	1.44% <sup>***</sup>
2014	5A-	47.368	6.495	21.493	6.798	3.930
	5A+	50.729	6.636	22.579	7.063	3.979
		7.09% <sup>***</sup>	2.17% <sup>*</sup>	5.05% <sup>***</sup>	3.90% <sup>***</sup>	1.25% <sup>**</sup>
2015	5A-	42.734	7.582	18.044	6.426	3.479
	5A+	46.201	7.712	19.293	6.730	3.554
		8.11% <sup>***</sup>	1.72% <sup>NS</sup>	6.93% <sup>***</sup>	4.72% <sup>***</sup>	2.16% <sup>***</sup>
2016	5A-	49.292	5.974	19.829	6.580	3.735
	5A+	51.266	6.064	20.610	6.816	3.745
		4.00% <sup>*</sup>	1.50% <sup>NS</sup>	3.94% <sup>**</sup>	3.58% <sup>***</sup>	0.27% <sup>NS</sup>
Overall	5A-	43.639	6.123	19.618	6.627	3.659
	5A+	46.659	6.201	20.678	6.894	3.712
		6.92% <sup>***</sup>	1.28% <sup>NS</sup>	5.41% <sup>***</sup>	4.04% <sup>***</sup>	1.45% <sup>***1</sup>

<sup>1</sup> %s indicate amount gained in 5A+ NILs compared with 5A- NILs. Superscripts indicate significance determined by ANOVA for either each year, or across all years (final row). ie. NS = Non-significant, \* = P < 0.05, \*\* = P < 0.01, \*\*\* = P < 0.001. 2012-13 = BC<sub>2</sub>-NILs, 2014-16 = BC<sub>4</sub>-NILs.

5A NILs were measured for a series of spike yield component traits to determine possible pleiotropic effects associated with the 5A+ TGW effect. Within most years, there was no significant effect of the 5A+ allele on spike yield components such as spikelet number, seeds per spikelet or grain number per spike (Table 4.7). However, when all years were analysed together, there was a significant reduction in grain number (-3.55%,  $P = 0.04$ ) and seeds per spikelet (-3.37%,  $P = 0.015$ ) associated with the 5A+ QTL. This statistical significance was driven by a particularly strong negative effect in 2016 as grain number and seeds per spikelet were non-significant in the preceding four seasons (2012-15). Overall, however, the 5A+ QTL is associated with a consistent small decrease in these spike yield components. Taking into account the 6.92% effect of the 5A+ QTL on TGW and the tendency for decreases in some spike yield components, the overall spike yield increased by 2.33% ( $P = 0.032$ ) across the five years. However, similar to grain number and seeds per spikelet, the statistical significance is driven by a single year (2014) despite overall positive effects in another three years (2012, 2013, and 2015).

**Table 4.7: Spike yield components of ten representative single ear samples of 5A- and 5A+ NILs**

Year	Genotype	Spikelet number	Viable Spikelets	Spike Length	Grain number per spike	Spike yield (g/spike)	Seeds per spikelet	SES-TGW (g)	SES-Grain Area (mm <sup>2</sup> )	SES-Grain length (mm)	SES-Grain width (mm)
2012	5A-	24.69	22.33	9.41	64.15	2.49	2.60	38.70	19.71	6.66	3.62
	5A+	24.87	22.40	9.64	63.29	2.58	2.55	40.61	20.59	6.88	3.66
		0.73% <sup>NS</sup>	0.36% <sup>NS</sup>	2.43% <sup>NS</sup>	-1.34% <sup>NS</sup>	3.83% <sup>NS</sup>	-1.82% <sup>NS</sup>	4.92% <sup>*</sup>	4.46% <sup>***</sup>	3.32% <sup>***</sup>	1.26% <sup>NS</sup>
2013	5A-	21.93	20.79	9.50	65.07	2.84	2.97	43.65	20.35	6.66	3.78
	5A+	22.00	20.64	9.63	62.72	2.90	2.85	46.33	21.33	6.92	3.83
		0.30% <sup>NS</sup>	-0.72% <sup>NS</sup>	1.32% <sup>NS</sup>	-3.60% <sup>NS</sup>	2.16% <sup>NS</sup>	-3.92% <sup>NS</sup>	6.14% <sup>***</sup>	4.84% <sup>***</sup>	3.88% <sup>***</sup>	1.09% <sup>**</sup>
2014	5A-	21.54	20.42	-	84.06	4.11	3.90	48.90	21.55	6.74	3.94
	5A+	21.59	20.31	-	82.43	4.36	3.81	52.90	22.76	7.02	4.01
		0.21% <sup>NS</sup>	-0.53% <sup>NS</sup>		-1.94% <sup>NS</sup>	6.02% <sup>**</sup>	-2.24% <sup>NS</sup>	8.17% <sup>***</sup>	5.60% <sup>***</sup>	4.08% <sup>***</sup>	1.73% <sup>***</sup>
2015	5A-	20.65	18.27	-	54.83	2.56	3.00	46.74	19.06	6.63	3.59
	5A+	20.52	18.09	-	53.89	2.64	2.98	48.97	20.12	6.91	3.63
		-0.61% <sup>NS</sup>	-0.96% <sup>NS</sup>		-1.71% <sup>NS</sup>	3.03% <sup>NS</sup>	-0.84% <sup>NS</sup>	4.77% <sup>**</sup>	5.54% <sup>***</sup>	4.33% <sup>***</sup>	1.01% <sup>NS</sup>
2016	5A-	23.25	22.55	-	83.77	3.86	3.72	46.04	19.59	6.63	3.65
	5A+	23.27	22.35	-	77.04	3.75	3.45	48.68	20.35	6.78	3.71
		0.10% <sup>NS</sup>	-0.90% <sup>NS</sup>		-8.04% <sup>**</sup>	-2.90% <sup>NS</sup>	-7.24% <sup>**</sup>	5.72% <sup>**</sup>	3.86% <sup>**</sup>	2.21% <sup>***</sup>	1.56% <sup>*</sup>
Overall	5A-	22.41	20.87	9.46	70.37	3.17	3.24	44.81	20.05	6.66	3.72
	5A+	22.45	20.76	9.63	67.88	3.25	3.13	47.49	21.03	6.90	3.77
		0.17% <sup>NS</sup>	-0.53% <sup>NS</sup>	1.87% <sup>*</sup>	-3.55% <sup>*</sup>	2.33% <sup>*</sup>	-3.37% <sup>*</sup>	6.00% <sup>***</sup>	4.87% <sup>***</sup>	3.57% <sup>***</sup>	1.34% <sup>***</sup>

%s indicate amount gained in 5A+ NILs compared with 5A- NILs. Superscripts indicate significance determined by ANOVA for either each year, or across all years (final row) i.e. NS = Non-significant, \* = P < 0.05, \*\* = P < 0.01, \*\*\* = P < 0.001. 2012-13 = BC<sub>2</sub>-NILs, 2014-16 = BC<sub>4</sub>-NILs.

5A NILs were also measured for several developmental traits (Table 4.8). There was a significant reduction of 4 tillers per m across two years ( $P = 0.008$ ) and a 1 cm increase in crop height in a single year ( $P = 0.038$ ) in the 5A+ NILs (Table 4.8). No effect was seen for spikelet number, days to heading, and days to maturity or HLW (Table 4.8). Taken together, these results suggest that the 5A+ QTL has a consistent positive effect on TGW and that the effects on yield are modulated by a series of smaller compensating negative effects on yield components such as grain number, seeds per spike and tiller number.

**Table 4.8: Developmental traits of 5A NILs**

Year	Genotype	Days to heading	Days to maturity	Tiller number	Crop Height	HLW
2012	5A-	250.2	318.5	126.7	76.9	NA
	5A+	250.5	317.7	121.2	77.9	NA
		0.3*	-0.8*	-5.5*	1.0*	
2013	5A-	250.2	298.9	NA	NA	NA
	5A+	250.5	298.7	NA	NA	NA
		0.3 <sup>NS</sup>	-0.2 <sup>NS</sup>			
2014	5A-	236.4	293.3	69.3	NA	78.3
	5A+	236.5	293.3	66.9	NA	77.5
		0.1 <sup>NS</sup>	0.0 <sup>NS</sup>	-2.4 <sup>NS</sup>		-0.96%***
2015	5A-	246.5	NA	NA	NA	79.0
	5A+	246.0	NA	NA	NA	79.8
		-0.5 <sup>NS</sup>				0.96% <sup>NS</sup>
2016	5A-	242.5	NA	NA	NA	75.6
	5A+	242.8	NA	NA	NA	74.6
		0.3 <sup>NS</sup>				-1.37% <sup>NS</sup>
Overall	5A-	245.2	303.6	98.0	76.9	77.6
	5A+	245.2	303.2	94.0	77.9	77.3
		0.1 <sup>NS</sup>	-0.3 <sup>NS</sup>	-4.0**	NA	-0.44% <sup>NS</sup>

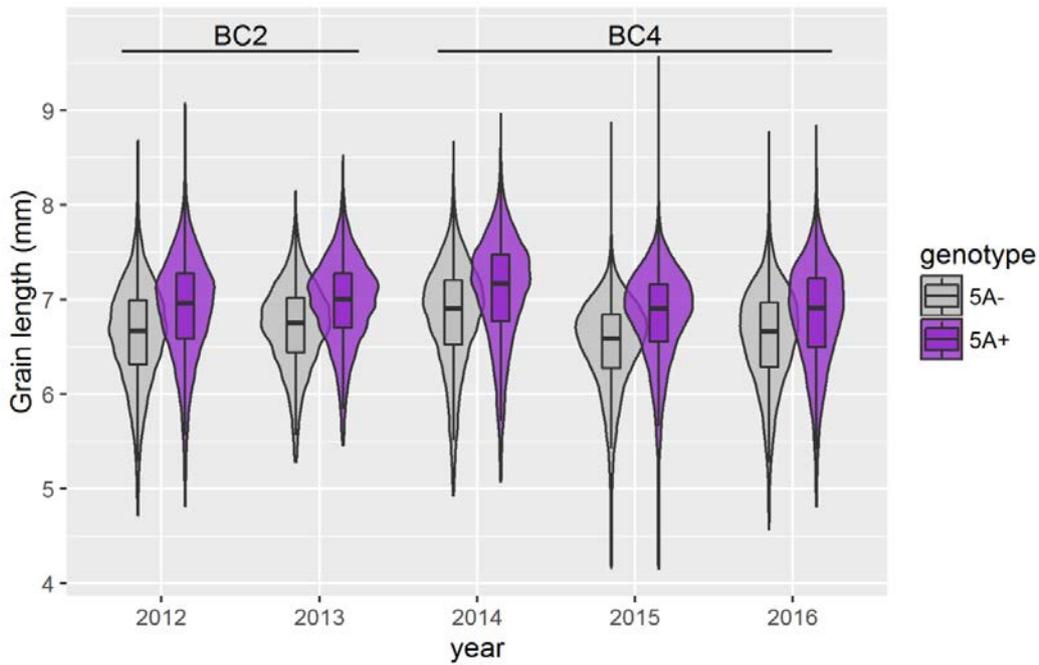
Differences indicate amount gained in 5A+ NILs compared with 5A- NILs. Superscripts indicate significance determined by ANOVA for either each year, or across all years (final row) i.e. NS = Non-significant, \* =  $P < 0.05$ , \*\* =  $P < 0.01$ , \*\*\* =  $P < 0.001$ . 2012-13 = BC<sub>2</sub>-NILs, 2014-16 = BC<sub>4</sub>-NILs. HLW = hectolitre weight.

#### **4.1.2.2 The TGW increase in 5A+ NILs is primarily due to increased grain length**

NILs were assessed for grain morphometric parameters (length, width and area) (Table 4.6). 5A+ NILs had significantly increased grain length ( $P < 0.001$ ), width ( $P < 0.001$ ) and area ( $P < 0.001$ ) compared to 5A- NILs across all years with the exception of width in 2016. On average, the 5A+ QTL increased grain length by 4.04% ( $P < 0.001$ ), ranging from 3.58 to 4.72% ( $P < 0.001$  in all years). Unlike the *TaGW2-A* NILs which had equivalent differences in length and width (Table 1.5), the 5A effect on width was smaller than that on length, averaging 1.45% ( $P < 0.001$ ; range 0.27 to 2.35%) and significant in four out of five years (Table 4.6). The effects on length and width combined to increase grain area by an average of 5.41% ( $P < 0.001$ ), significant in all five years. These results were based on combine harvested grain samples and were also confirmed in ten representative single ear samples taken before harvest. TGW of the ten spikes correlated strongly with the whole plot samples ( $r = 0.84$ ,  $P < 0.001$ ) and showed a similar difference between NILs (6.00%,  $P < 0.001$ ; Table 4.7). Across datasets, the effect of the 5A+ QTL on grain length was more than twice the size of the effect on grain width. This fact, together with the more consistent effect on grain length across years (Coefficient of variation length = 10.6%; width = 55.3%; TGW = 27.8%) suggests that the increase in grain length is the main factor driving the increase in grain area and TGW.

#### **4.1.2.3 The 5A QTL has a uniform effect on grains within the spike**

Violin plots for grain length showed variation in the shape of the distribution of individual seeds among years (Figure 4.3). However, within years the 5A- and 5A+ grain length distributions were very similar in shape, suggesting the 5A QTL affects all grains uniformly and in a stable manner across the whole spike and within spikelets, similar to the 6A QTL. In all years, the 5A+ grain length distributions were shifted higher than the 5A- NILs with an increase in longer grains and fewer shorter grains, in addition to the higher average grain length (Figure 4.3). Grain width distributions were also very similar in shape within years, but had a less pronounced shift between NILs (Figure 4.4) consistent with the overall smaller effect of the 5A QTL on grain width.



**Figure 4.3: Distribution of grain length of NILs from whole plot samples**

Violin plots showing the distribution of individual seed measurements of grain length across the five field experiments of BC<sub>2</sub> (2012-2013) and BC<sub>4</sub> (2014-2016) near isogenic lines (NILs). Purple = 5A+ NILs, grey plots = 5A- NILs. All within year comparisons between NILs were significant ( $P < 0.001$ ).

**Figure 4.4: Distribution of grain width of NILs from whole plot samples**

Violin plots showing the distribution of individual seed measurements of grain width across the five field experiments of BC<sub>2</sub> (2012-2013) and BC<sub>4</sub> (2014-2016) near isogenic lines (NILs). Purple = 5A+ NILs, grey plots = 5A- NILs. 2012-2015 within year comparisons between NILs were significant ( $P < 0.01$ ). The 2016 comparison between NILs was non-significant.

#### **4.1.2.4 The 5A QTL region acts during grain development to increase grain length**

Grain development time courses of two 5A- and two 5A+ BC4 NILs were conducted to determine when differences in grain morphometric parameters (grain length, width and area) between NILs are first established. Grain fresh weight (FW) and dry weight (DW) were also measured. Grains were sampled in 2014, 2015 and 2016 from field plots at anthesis (with the exception of 2014) and at five further time points across grain development until the difference in grain size had been fully established. Exact time points are detailed in Figure 4.5 and Table 4.9. Similar profiles were observed in all years, with the first morphometric parameter to show a significant difference being grain length (Figure 4.5a-c) and any differences in grain width were not observed until the final time point (Figure 4.5d-f).

In 2014, the first significant difference in grain length was observed at 8 dpa with 5A+ NILs having 3.3% longer grains than 5A- NILs ( $P = 0.001$ ; Table 4.9, Figure 4.5a). This was later in grain development than when the first significant differences in grain size were observed in both the 6A and *TaGW2-A* NILs. 5A+ grains remained significantly longer until the final time point (27 dpa; 4.9 % increase,  $P < 0.001$ ; Figure 4.5a). Similarly, differences in grain area were observed at 8 dpa (5.6%,  $P = 0.001$ ) and maintained until the final time point (6.7%,  $P < 0.001$ ). A significant difference in grain width was observed at the final time point only (2.6%,  $P = 0.001$ ). Grains from 5A+ NILs were also significantly heavier at 8 dpa (FW: 8.3%,  $P = 0.003$ ; DW: 6.8%,  $P = 0.02$ ). 5A+ grains remained heavier until the final time point, although the differences at 12 dpa were non-significant (FW: 8.3%,  $P = 0.061$ , DW: 7.5 %,  $P = 0.258$ ).

In 2015, the first significant difference in grain length was observed at 12 dpa with 5A+ NILs having 1.5% longer grains than 5A- NILs ( $P = 0.034$ ). Although this was four days later than the first grain length difference in 2014, the mean grain lengths were similar at these time points in the two years (2014: 5A+ = 6.62 mm, 5A- = 6.41 mm; 2015: 5A+ = 6.50 mm, 5A- = 6.40 mm). The 2015 grain length effect increased to 4.4 % at 19 dpa ( $P < 0.001$ ) and was maintained at the final time point (26 dpa; 4.5 % increase,  $P < 0.001$ ; Figure 4.5b). Significant differences in grain area were detected at 19 dpa (5.7 % increase;  $P < 0.001$ ; Figure 4.5h) and this difference was maintained at the final time point (6.1 %,  $P < 0.001$ ). No significant effects on grain width were observed until 26 dpa when 5A+ NILs increased grain width by 1.7 % ( $P = 0.015$ ; Figure 4.5e). By the final time point 5A+ NILs also had significantly heavier grains (FW: 7.13 %,  $P < 0.001$ ; DW: 3.7%,  $P = 0.01$ ; Figure 4.5k, n).

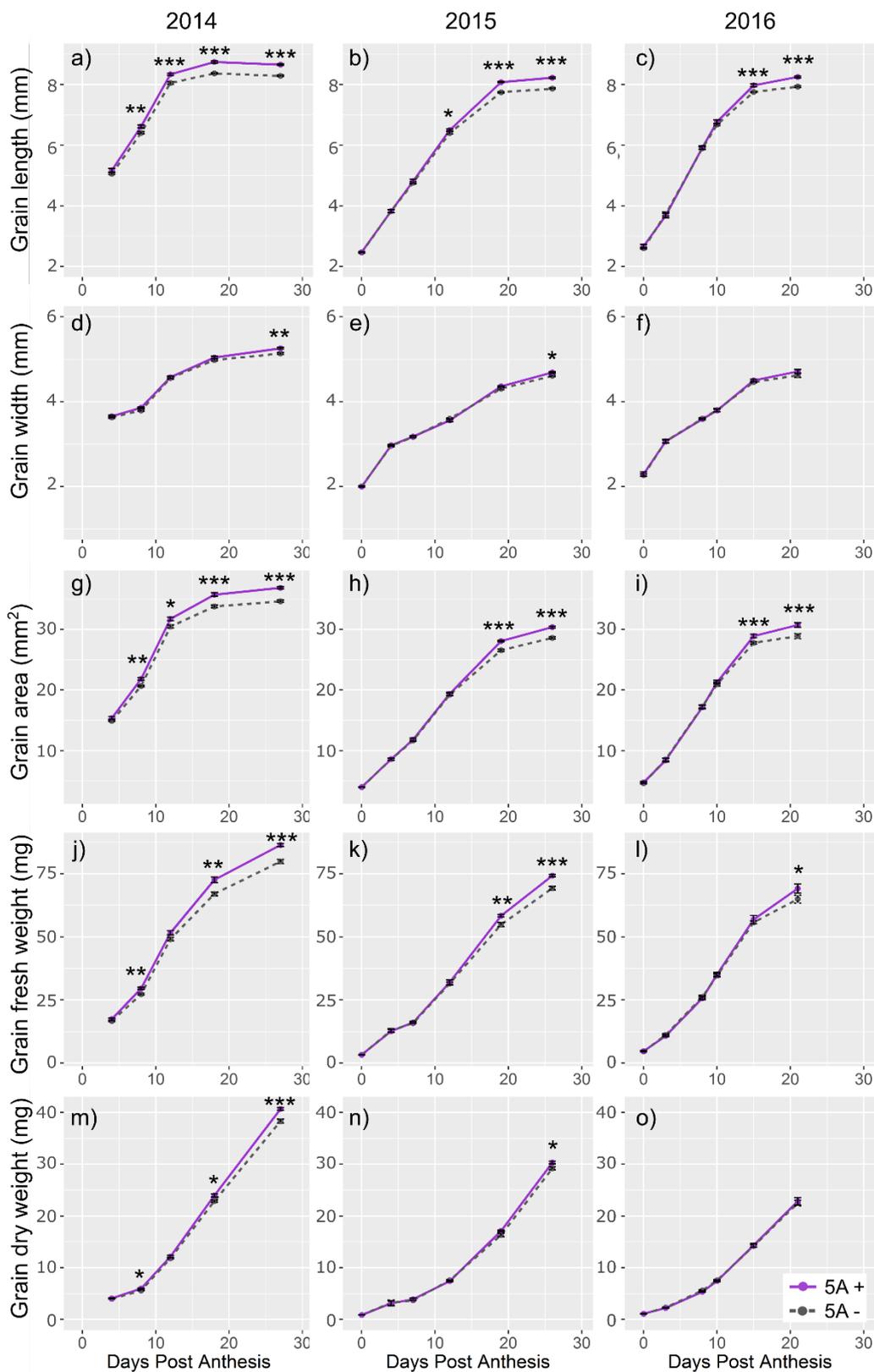
In 2016, the first differences in both grain length (2.9 %,  $P < 0.001$ ) and grain area (4.2%,  $P < 0.001$ ) were observed at 15 dpa (Figure 4.5c, i). These differences increased to 4.0% (grain length,  $P < 0.001$ ) and 6.3% (grain area,  $P < 0.001$ ) at the final time point (21 dpa). There was also a 6.4% increase in the fresh weight of 5A+ grains at the final time point ( $P = 0.019$ ). No significant differences were observed at any time point in grain width or dry weight in 2016, reminiscent of the non-significant difference in the grain width of mature grains in 2016 (Table 4.6).

In all years, the grain size and dry weight effects observed were consistent with the differences observe in mature grains (Table 4.6). The fact that the effects on width, area and weight were all observed after the first significant difference on grain length in all three years further supports grain length as the main factor driving the increase in grain weight in 5A+ NILs.

**Table 4.9: Differences between 5A NILs of grain size and weight parameters during grain development time courses**

Year	Days Post Anthesis	Length (%)	Width (%)	Area (%)	FW (%)	DW (%)
2014	4	2.28 <sup>0.130</sup>	0.77 <sup>0.466</sup>	3.24 <sup>0.0695</sup>	4.57 <sup>0.0764</sup>	2.32 <sup>0.430</sup>
	8	3.33 <sup>0.00146</sup>	1.69 <sup>0.0512</sup>	5.57 <sup>0.001</sup>	8.31 <sup>0.004</sup>	6.82 <sup>0.020</sup>
	12	3.57 <sup>&lt;0.001</sup>	0.43 <sup>0.670</sup>	4.12 <sup>0.0132</sup>	5.07 <sup>0.062</sup>	2.84 <sup>0.258</sup>
	18	4.48 <sup>&lt;0.001</sup>	1.16 <sup>0.162</sup>	5.72 <sup>&lt;0.001</sup>	8.33 <sup>0.003</sup>	4.73 <sup>0.031</sup>
	27	4.46 <sup>&lt;0.001</sup>	2.47 <sup>0.001</sup>	6.52 <sup>&lt;0.001</sup>	8.47 <sup>&lt;0.001</sup>	6.30 <sup>&lt;0.001</sup>
2015	0	0.64 <sup>0.386</sup>	-0.23 <sup>0.734</sup>	0.80 <sup>0.560</sup>	1.58 <sup>0.576</sup>	1.65 <sup>0.613</sup>
	4	0.06 <sup>0.993</sup>	0.41 <sup>0.751</sup>	0.03 <sup>0.983</sup>	-1.79 <sup>0.771</sup>	-2.88 <sup>0.876</sup>
	7	1.22 <sup>0.404</sup>	0.24 <sup>0.777</sup>	1.57 <sup>0.493</sup>	1.95 <sup>0.505</sup>	-2.20 <sup>0.472</sup>
	12	1.49 <sup>0.035</sup>	-1.05 <sup>0.237</sup>	0.63 <sup>0.651</sup>	1.35 <sup>0.684</sup>	-2.29 <sup>0.317</sup>
	19	4.35 <sup>&lt;0.001</sup>	1.26 <sup>0.090</sup>	5.74 <sup>&lt;0.001</sup>	6.27 <sup>0.006</sup>	4.72 <sup>0.077</sup>
	26	4.48 <sup>&lt;0.001</sup>	1.66 <sup>0.015</sup>	6.06 <sup>&lt;0.001</sup>	7.13 <sup>&lt;0.001</sup>	3.71 <sup>0.01</sup>
2016	0	2.65 <sup>0.191</sup>	1.16 <sup>0.527</sup>	3.20 <sup>0.304</sup>	2.22 <sup>0.626</sup>	2.80 <sup>0.462</sup>
	3	-1.21 <sup>0.352</sup>	-0.03 <sup>0.926</sup>	-1.16 <sup>0.555</sup>	-3.14 <sup>0.345</sup>	-3.51 <sup>0.088</sup>
	8	0.40 <sup>0.743</sup>	-0.35 <sup>0.644</sup>	-0.63 <sup>0.750</sup>	-2.10 <sup>0.463</sup>	-4.61 <sup>0.168</sup>
	10	1.35 <sup>0.144</sup>	-0.14 <sup>0.919</sup>	1.51 <sup>0.455</sup>	1.56 <sup>0.592</sup>	-1.44 <sup>0.602</sup>
	15	2.88 <sup>&lt;0.001</sup>	0.88 <sup>0.118</sup>	4.15 <sup>&lt;0.001</sup>	2.10 <sup>0.379</sup>	1.06 <sup>0.705</sup>
	21	4.02 <sup>&lt;0.001</sup>	1.97 <sup>0.063</sup>	6.30 <sup>&lt;0.001</sup>	6.37 <sup>0.019</sup>	1.48 <sup>0.551</sup>

%s indicate amount gained in 5A+ NILs compared with 5A- NILs. Superscripts are the ANOVA P values of the comparison between 5A+ and 5A- NILs.

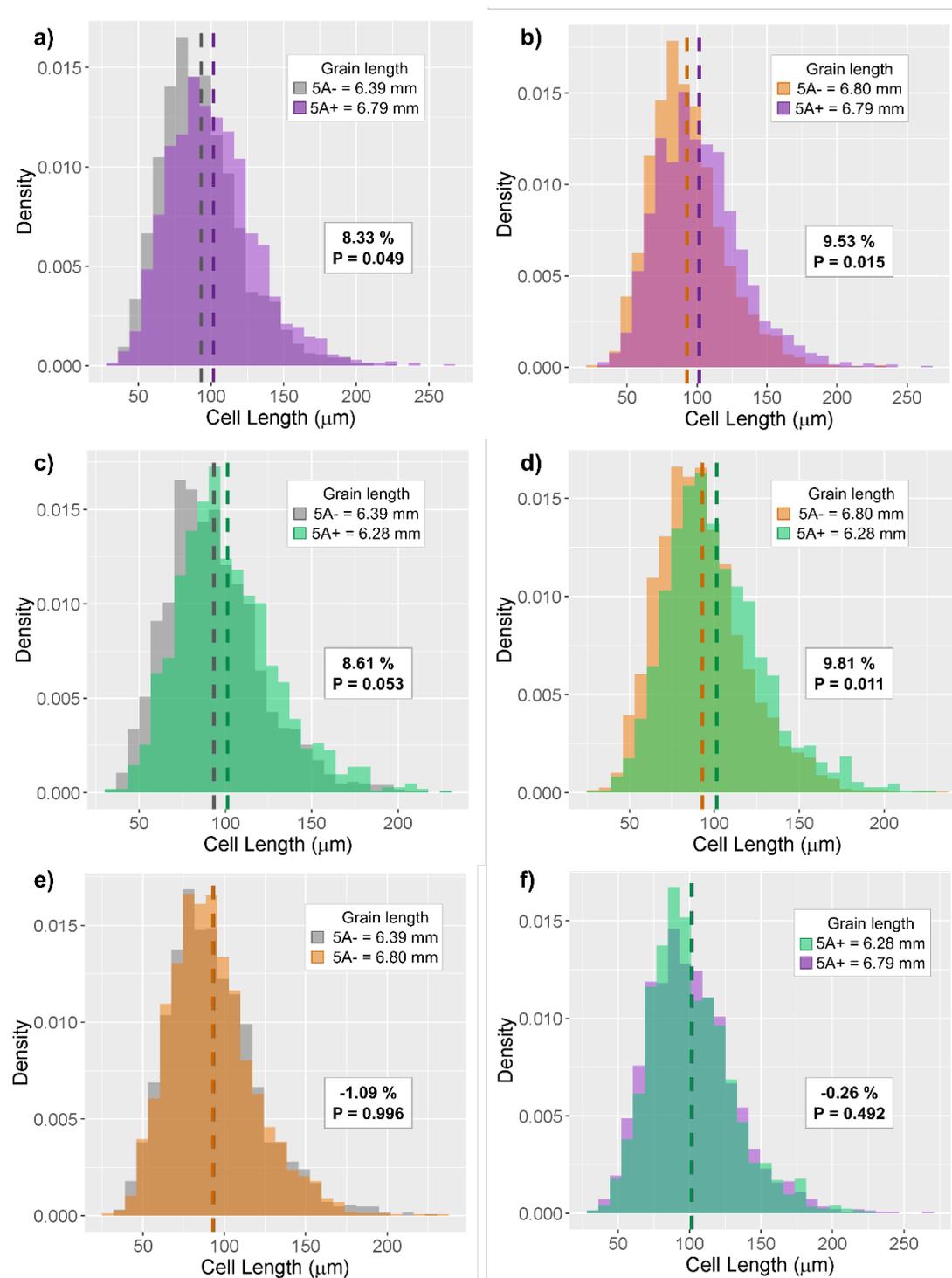


**Figure 4.5: Grain developmental time courses of 5A NILs**

Grain length (a, b, c), width (d, e, f), area (g, h, i), fresh weight (j, k, l) and dry weight (m, n, o) of 5A- (grey, dashed line) and 5A+ (purple, solid line) BC<sub>4</sub> near isogenic lines (NILs) during grain development in 2014-2016 field trials. 2014 samples: 4, 8, 12, 18 and 27 days post anthesis (dpa); 2015 samples: 0 (anthesis), 4, 7, 12, 19 and 26 dpa; 2016 samples: 0, 3, 8, 10, 15 and 21 dpa. \* =  $P < 0.05$ , \*\* =  $P < 0.01$ , \*\*\* =  $P < 0.001$ . Error bars show one standard error above and below the mean.

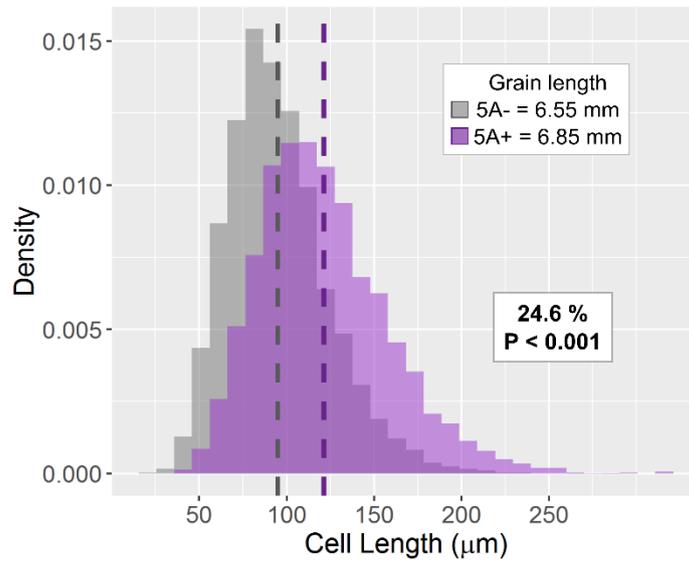
#### **4.1.2.5 5A+ NILs have increased pericarp cell length independent of absolute grain length**

Grain size is influenced by both cell proliferation and cell expansion. To understand which of these processes the 5A QTL affects, scanning electron microscopy (SEM) was used to image pericarp cells and determine cell size of BC<sub>4</sub> 5A- and 5A+ grains. Mature grains from the 2015 field experiment were selected from a 5A- and 5A+ NIL pair based on their grain length and using a variety of criteria to allow for distinct comparisons (Figure 4.6). The first comparison was between grains of average length from the 5A- and 5A+ NIL distributions (Figure 4.6a). Average 5A+ grains had an 8.33 % significant increase in mean cell length ( $P = 0.049$ ) compared to average 5A- grains and this was reflected in a shift in the whole distribution of 5A+ cell lengths (Figure 4.6a). Next, cell lengths in grains of the same size from 5A- and 5A+ NILs were compared. Relatively long grains from the 5A- NIL distribution (Figure 4.6b; orange) that had the same grain length as the average 5A+ grains were selected. This comparison showed that 5A+ grains still had longer cells (9.53%,  $P = 0.015$ ) regardless of the fact that the grain length of the two groups were the same (6.8 mm; Figure 4.6b). The opposite comparison was also made by selecting relatively short grains from the 5A+ NIL distribution (Figure 4.6c; green) and comparing them with average 5A- grains. Similar to before, the 5A+ grains had longer cells (8.61%), although this effect was borderline non-significant ( $P = 0.053$ ; Figure 4.6c). Finally, a comparison of long 5A- grains and short 5A+ grains again showed that cells were longer in 5A+ grains (9.81%,  $P = 0.011$ ; Figure 4.6d), even though the 5A+ grains used in this comparison were 7.65% shorter than the 5A- grains. Within genotype comparisons of cell length between grains of different lengths showed no significant differences in mean cell length (Figure 4.6e, f). The results were confirmed in 2016 where average 5A+ grains had a 24.6 % significant increase in mean cell length compared to average 5A- grains ( $P < 0.001$ ; Figure 4.7). These results indicate that the 5A+ region from Badger increases the length of pericarp cells independent of absolute grain length. In 2015, average length grains of both 5A- and 5A+ NILs had the same number of cells (calculated as grain length / mean cell length; Figure 4.8a). However, in 2016, 5A- NILs had significantly more cells than 5A+ NILs (19.8 %,  $P < 0.001$ ; Figure 4.8b).



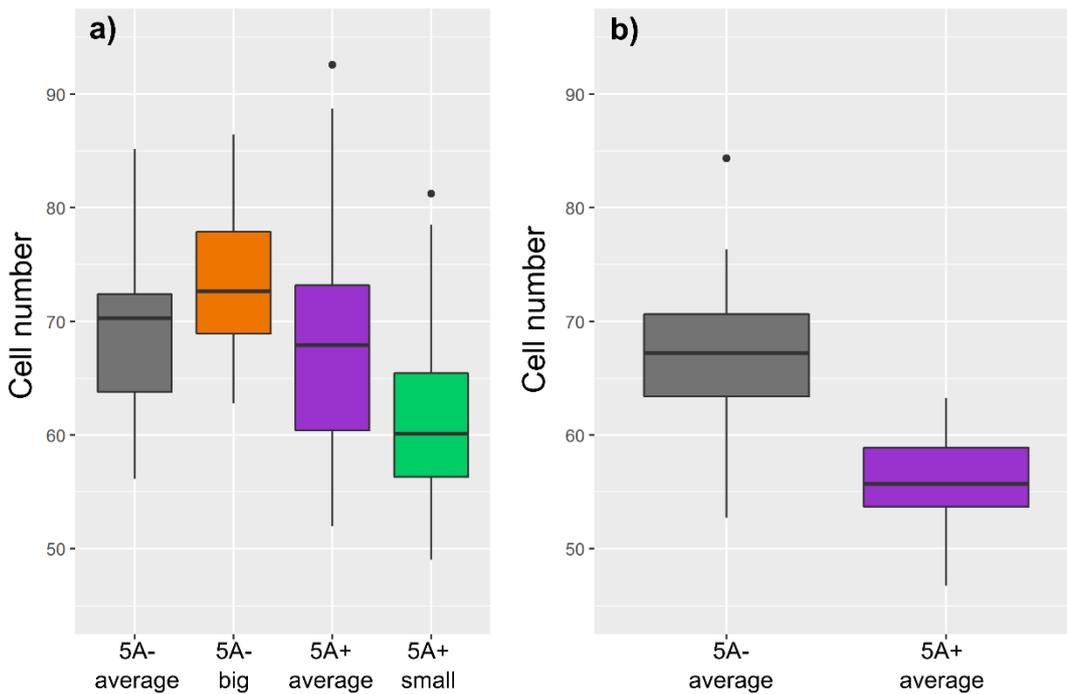
**Figure 4.6: Comparisons of pericarp cell length in 5A NILs (2015)**

Density plots of cell length measurements from 27 grains per genotype group in 2015; dashed line represents the mean. “Grain length” insets show the average grain length of each group of grains used for measurements. In panels a-d the increase in cell length of 5A+ near isogenic lines (NILs) relative to cell length of 5A- grains is shown as a percentage along with the P values calculated using ANOVA to compare means of the two groups displayed. In panels e-f the percentage indicates the increase in cell length of the group with longer grains relative to the group with shorter grains. a) Grains of average length from 5A- and 5A+ NILs, b) average 5A+ grains and equivalent 5A- grains, c) average 5A- grains and equivalent 5A+ grains, d) long 5A- grains (length equivalent to average 5A+ grains) and short 5A+ grains (grain length equivalent to average 5A- grains), e) average 5A- grains and long 5A- grains, f) short 5A+ grains and average 5A+ grains.



**Figure 4.7: Comparison of pericarp cell length in 5A NILs (2016)**

Density plots of cell length measurements, dashed line represents the mean. “Grain length” inset shows the average grain length of each group of grains used for measurements. Grains used were of average length from 5A- and 5A+. The increase in cell length of 5A+ NILs relative to cell length of 5A- grains is shown as a percentage along with the P value calculated using ANOVA to compare means of the two groups displayed.



**Figure 4.8: Comparison of pericarp cell number in 5A NILs (2015 and 2016)**

Boxplots show the distribution of cell number (calculated as grain length/mean cell length) in the different groups of grains from which pericarp cell size was measured in 2015 (a) and 2016 (b). In 2015, there was no significant difference between 5A- average and 5A+ average cell numbers, whereas in 2016 the difference was significant ( $P < 0.001$ )

## 4.2 Fine mapping of the 5A and 6A QTL

### 4.2.1 Genetic mapping of the 6A QTL for grain width

#### 4.2.1.1 Grain width maps to a 4.6 cM interval on chromosome 6A

A set of 67 RILs with recombination between the microsatellite markers *gwm334* and *gwm570* were used initially to fine map the grain width QTL on chromosome 6A. These markers define the bounds of the genetic map of chromosome 6A developed in Simmonds *et al.*, 2014 with the identification of the 6A thousand grain weight (TGW) QTL. These 67 RILs were identified in a screen of 212 plants performed by James Simmonds, defining an interval of 15.8 cM. The recombination events in individual RILs were defined by addition of 41 SNP markers across the 15.8 cM interval (details in methods; a). This identified two linkage blocks within the interval, comprised 13 (Linkage block 1; Figure 4.9a; circles) and twelve (Linkage block 2; Figure 4.9a; squares) markers each. Linkage block 2 contains the Hap-P2 marker, a marker previously described which maps the position of *TaGW2-A* (Su *et al.*, 2011) a proposed candidate gene for the 6A TGW QTL.

A subset of 41 RILs showing recombination between *BS00003635* and *BS00003835* were grown in five replicated field trials (6m plots 2013-2016, 1m plots 2015) and phenotyped for grain weight and grain morphometric parameters. These markers encompass the introgressed region of the 6A NILs, and hence the interval to which the 6A TGW effect was initially mapped (Simmonds *et al.*, 2014). 38 of these RILs were unambiguously assigned to 13 distinct RIL groups based on their genotype at the 39 markers within the interval between *BS00003635* and *BS00003835* (SR Gr1.1-13; Figure 4.9b). RILs with either the Spark (6A-; S-control) or Rialto (6A+; R-control) genotype across the entire interval were selected as controls. Grain width was used as the grain morphometric parameter for mapping as it had previously been defined as the factor underlying the TGW difference in 6A NILs (Chapter 2; Simmonds *et al.*, 2014). Across all five trials and within each trial individually, there were significant differences in grain width observed between RIL groups ( $P < 0.001$ ). Across all trials, the R-Control had 4.18% wider grains than the S-Control, ranging from 2.50% to 5.65% in individual trials, consistent with the grain width differences observed between 6A NILs (Chapter 2). Each RIL group was classified to a parental type (Spark, 6A-; Rialto, 6A+) using Dunnett's tests to both the R- and S-controls. For example, RIL groups were classified as Spark-like if they were both significantly different to the R-control and non-significantly different to the S-control and *vice versa*. Of the thirteen RIL groups, eleven were unambiguously assigned as either Spark or Rialto-like (Figure 4.9c; grey and orange, respectively). Two of the groups (SR Gr1.2 and SR Gr1.9) were significantly different from both the S- and R-controls and therefore, were classified as intermediate types (Figure 4.9c; hatched). Using this method, the grain width was mapped to the 4.6 cM interval between *BS00066522* and *BS00066623* (Figure 4.9a; green markers). The critical RIL groups defining this interval (SR Gr1.3,8,10) are indicated with green arrows in Figure 4.9c. The interval between *BS00066522* and

*BS00066623* encompassed 26 additional markers, however, 25 of these belong to Linkage groups 1 and 2 and notably the interval contained Hap-P2 (the *TaGW2-A* marker). Only two RIL groups had recombination within this interval: SR Gr1.2 (two independent RILs) which has recombination between *BA00363556* and Linkage group 2 (squares) and SR Gr1.9 (one RIL) which has recombination between Linkage group 1 (circles) and *BA00363556*. However, both these groups were classified as intermediate types and therefore, could not be used to define the interval further. Looking at the classification of each of the lines and trials individually (Table 4.10) shows that the classification of lines in both SR Gr1.2 and 1.9 was variable across trials. For example, in SR Gr1.2, SR21 was classified as S in 2013 and 2015 6m plots, SR in 2014 6m and 2015 1m and R in 2016 6m plots. SR Gr1.9 showed a slight tendency towards an S-like classification but was still variable (S in 2013-2015 6m plots, SR in 2015 1m plots and R in 2016 6m plots). Interestingly, no RIL groups were classified as S in the 2016 trial (only SR and R classifications could be assigned), however, reanalysing the data across trials without the 2016 data still resulted in the same overall classification of RIL groups (data not shown). The grain width interval on chromosome 6A could therefore, not be defined further than the 4.6 cM interval between *BS00066522* and *BS00066623* due to limited recombination within this RIL population.

Using the same approach to fine map the 6A TGW effect (as opposed to the grain width effect) resulted in only four of thirteen RIL groups being unambiguously assigned to a parental type (Figure 4.10). Using these four lines, the TGW effect can be positioned between *BS00010933* and *BS00066623*. However, the Dunnett's tests did not identify the S- and R-controls as significantly different from each other and therefore these results are not reliable. This highlights the importance of mapping using the grain width phenotype due to the increased phenotypic stability compared with TGW. All subsequent genetic mapping was therefore performed using grain width only.

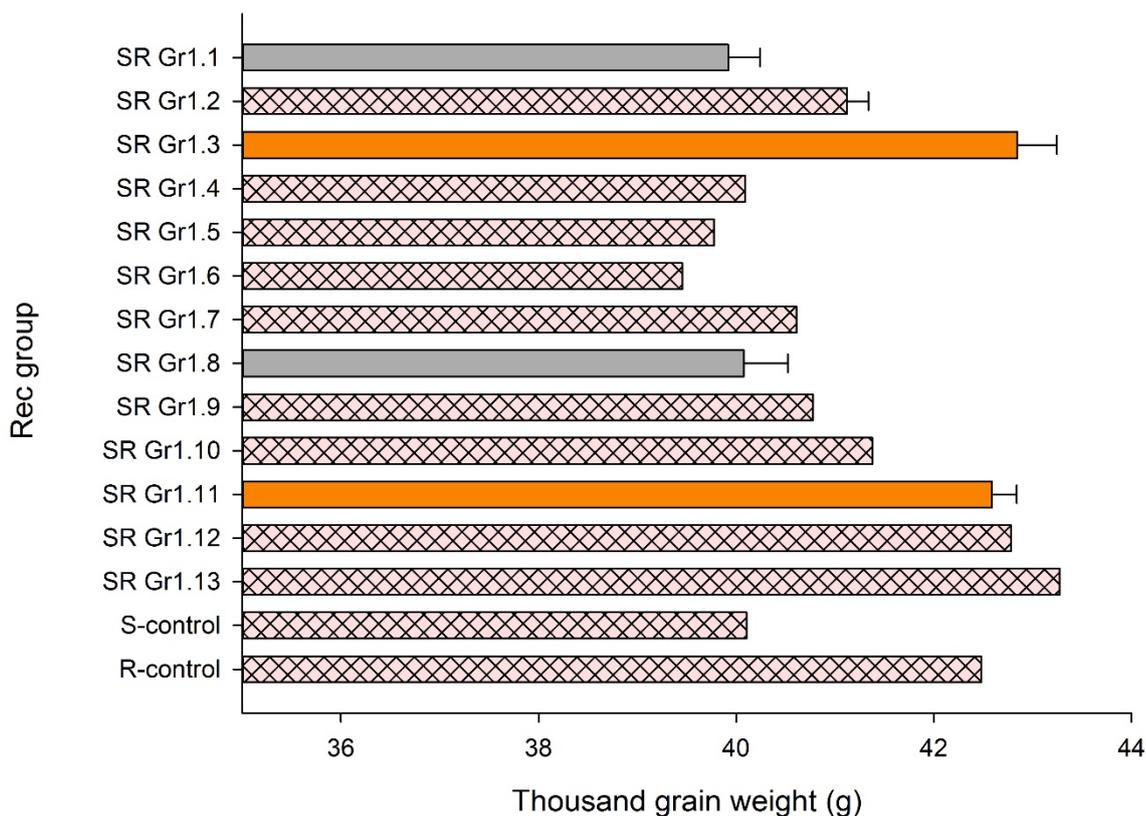


**Table 4.10: ANOVA adjusted mean grain width and Dunnett classification of BC<sub>4</sub> RILs used for initial fine mapping of the 6A grain width QTL**

		2013 (6m plots)		2014 (6m plots)		2015 (6m plots)		2015 (1m plots)		2016 (6m plots)		Overall		
RIL group	Class	RIL	Width (mm)	Class	Width (mm)	Class	Width (mm)	Class						
SR Gr1.1	S	SR6	3.555	S	3.878	S	3.274	S	3.394	SR	3.951	SR	3.607	SR
		SR19	-	-	-	-	-	-	3.340	SR	-	-	3.573	S
		SR23	-	-	-	-	-	-	3.348	SR	-	-	3.581	S
		SR25	-	-	-	-	-	-	3.271	S	-	-	3.505	S
		SR35	-	-	-	-	-	-	3.281	S	-	-	3.515	S
		SR36	-	-	-	-	-	-	3.309	S	-	-	3.542	S
		SR38	-	-	-	-	-	-	3.347	SR	-	-	3.580	S
		SR44	-	-	-	-	-	-	3.306	S	-	-	3.539	S
		SR46	-	-	-	-	-	-	3.336	SR	-	-	3.569	S
		SR66	-	-	-	-	-	-	3.334	SR	-	-	3.567	S
		SR67	-	-	-	-	-	3.307	S	-	-	3.541	S	
SR Gr1.2	SR	SR1	3.595	S	3.848	S	3.333	SR	3.388	SR	3.928	SR	3.620	SR
		SR21	3.505	S	3.890	SR	3.303	S	3.372	SR	3.960	R	3.614	SR
SR Gr1.3	R	SR2	3.683	R	3.914	SR	3.356	R	3.478	R	4.036	R	3.690	R
		SR3	3.625	SR	3.925	SR	3.357	R	3.432	R	4.064	R	3.677	R
		SR4	3.601	S	3.889	SR	3.320	S	3.437	R	4.015	R	3.649	SR
		SR12	3.735	R	3.951	R	3.389	R	3.507	R	4.027	R	3.718	R
		SR13	-	-	-	-	-	-	3.422	R	-	-	3.656	R
SR Gr1.4	S	SR10	3.502	S	3.852	S	3.264	S	-	-	-	-	3.570	S
SR Gr1.5	S	SR63	-	-	-	-	-	3.301	S	-	-	3.534	S	
SR Gr1.6	S	SR45	-	-	-	-	-	3.293	S	-	-	3.527	S	
SR Gr1.7	S	SR57	-	-	-	-	-	3.342	SR	-	-	3.575	S	
SR Gr1.8	S	SR14	3.572	S	3.844	S	3.298	S	3.370	SR	3.942	SR	3.602	S
		SR54	-	-	-	-	-	-	3.284	S	3.910	SR	3.542	S
SR Gr1.9	SR	SR15	3.599	S	3.879	S	3.304	S	3.422	R	3.955	SR	3.630	SR
SR Gr1.10	R	SR39	-	-	-	-	3.381	R	3.383	SR	3.991	R	3.652	R
SR Gr1.11	R	SR17	3.707	R	3.938	R	3.427	R	3.504	R	4.054	R	3.722	R

		2013 (6m plots)		2014 (6m plots)		2015 (6m plots)		2015 (1m plots)		2016 (6m plots)		Overall		
RIL group	Class	RIL	Width (mm)	Class	Width (mm)	Class	Width (mm)	Class						
		SR22	-	-	-	-	-	-	3.483	R	-	-	3.717	R
		SR24	-	-	-	-	-	-	3.458	R	-	-	3.691	R
		SR27	-	-	-	-	-	-	3.411	SR	-	-	3.644	R
		SR28	-	-	-	-	-	-	3.443	R	-	-	3.677	R
		SR32	-	-	-	-	-	-	3.389	SR	-	-	3.623	SR
		SR51	-	-	-	-	-	-	3.447	R	-	-	3.680	R
		SR52	-	-	-	-	-	-	3.464	R	-	-	3.697	R
		SR55	-	-	-	-	-	-	3.420	R	-	-	3.653	R
		SR58	-	-	-	-	-	-	3.439	R	-	-	3.672	R
SR Gr1.12	R	SR30	3.690	R	3.943	R	3.409	R	-	-	-	-	3.712	R
SR Gr1.13	R	SR9	3.750	R	3.984	R	3.393	R	-	-	-	-	3.739	R
S-Control (6A-)		SR10C	3.521	S	3.860	S	3.252	S	3.302	S	3.873	S	3.558	S
R-Control (6A+)		SR9C	3.720	R	3.956	R	3.427	R	3.447	R	3.997	R	3.707	R

Width (mm) are the ANOVA adjusted means of grain width in each trial (or overall in the final column) each incorporating at least five replicates. Classifications were assigned using Dunnett's test to compare each line to a control (S-Control (6A-; narrow grains) and R-Control (6A+; wide grains)): S = significantly different from the R-Control and not significantly different from the S-Control; R = significantly different from the S-Control and no significantly different from the R-Control; SR = not significantly different from the S-Control or the R-Control.



**Figure 4.10: ANOVA adjusted mean thousand grain weight of the original 6A BC<sub>4</sub>RIL groups**

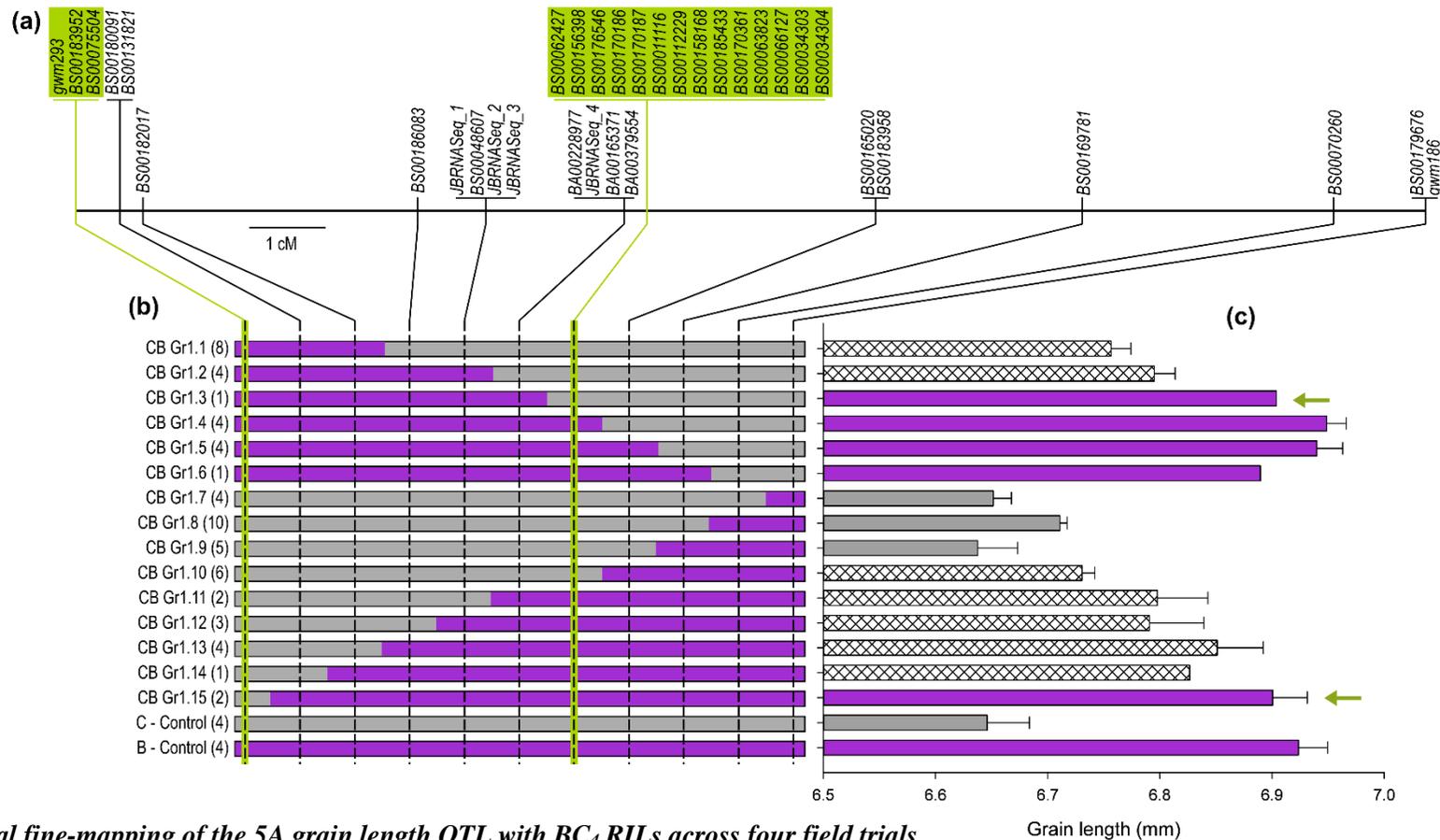
ANOVA adjusted mean thousand grain weight (TGW) for each 6A RIL group across five field trials (6m plots 2013-2016 and 1m plots 2015). Error bars are the standard error of all lines within the RIL group. Bars are coloured according to classification as R-control (6A+; orange) or S-control (6A-; grey) like according to Dunnett's test. Hatched bars were classified as intermediate.

#### 4.2.2 Genetic mapping of the 5A QTL for grain length

##### 4.2.2.1 Grain length maps to a 6.6 cM interval on chromosome 5A

A set of 60 BC<sub>4</sub>RILs showing recombination between *gwm293* and *gwm186* were used to fine-map the grain length interval on chromosome 5A. These markers were selected as they were used for generation of the 5A NILs and therefore, encompass the interval to which the 5A TGW and grain length effect were initially mapped. These 60 RILs were identified from a screen of 170 plants (performed by James Simmonds) defining a genetic distance of 17.65 cM between *gwm293* and *gwm186*.

The genotypes of the 60 RILs were further defined by the addition of 33 SNP markers across the interval between *gwm293* and *gwm186*. Genotyping with these 33 SNP markers defined the recombination events in the 60 RILs and, similar to the 6A interval discussed previously, revealed a linkage block of 14 markers along with several smaller groups of genetically linked markers (Figure 4.11a).



**Figure 4.11: Initial fine-mapping of the 5A grain length QTL with BC<sub>4</sub> RILs across four field trials**

(a) Genetic map of the 5A QTL mapping interval based on the original set of BC<sub>4</sub> recombinant inbred lines (RILs) displayed in (b). Markers highlighted in green are the flanks for the fine-mapped grain length interval defined by this population. (b) Graphical genotypes of RIL groups with the number in brackets indicating the number of independent RILs in each RIL group. RILs were grouped based on their genotypes defined by having either the Charger-like (grey; 5A-) or Badger-like (purple; 5A+) allele at each marker shown across the interval. (c) ANOVA adjusted mean grain lengths for each RIL group across four field trials (1m plots 2014-2015 and 6m plots 2015-2016). Error bars are the standard error of all lines within the RIL group. Bars are coloured according to classification as C-control (5A-; grey) or B-control (5A+; purple) like according to a Dunnett's test. Hatched bars were classified as intermediate. Green arrows indicate critical RIL groups that define the green highlighted markers as the flanks.

59 of the 60 RILs were unambiguously assigned to 15 RIL groups based on the genotype of each of the 33 markers across the interval between *gwm293* and *gwm186*, shown graphically in Figure 4.11b. Lines with either the Charger (5A-; C-control) or Badger allele (5A+; B-control) across all markers within the whole interval were used as controls. Additionally, in some trials 5A NILs were also used as controls (exact lines used in each trial are detailed in (Table 4.11)). The 59 RILs were grown and phenotyped for grain morphometric parameters across four field trials: 2014-2015 1m plots and 2015-2016 6m plots. As grain length is the main driver of the TGW difference between 5A NILs (Brinton *et al.*, 2017), the grain length phenotype was used for fine mapping.

Across trials and within each trial, significant differences in grain length between RILs were identified ( $P < 0.001$ ). Overall, the B-control group had 4.18% longer grains than C-control group (ranging 3.58 – 4.95%), reflective of the grain length differences observed between 5A NILs (4.04%, Table 4.6). In the same way as described previously for the 6A RILs, *post hoc* Dunnett's tests were used to classify RIL groups to a parental type. RIL groups significantly different from the B-control group and non-significantly different from the C-control group were classed as Charger-like (5A-, short grains; Figure 4.11c grey bars). RIL groups significantly different from the C-control group and non-significantly different from the B-control group were classed as Badger-like (5A+, long grains; Figure 4.11c purple bars). RIL groups that did not satisfy both conditions were classed as intermediate (CB; Figure 4.11c hatched bars). In this way, eight of the RIL groups could be assigned to a parental type, whilst the remaining seven groups were classed as intermediate. The eight groups that could be classed as either Charger or Badger-like defined the grain length effect to 7.49 cM interval between two groups of linked markers (Figure 4.11a; green markers). The left flank included *gwm293* (the original left hand flank of the introgressed interval) and the right hand flank consisted of 14 linked markers.

Six RIL groups had recombination between the 10 markers located within the 7.49 cM interval but all six were classed as intermediate and therefore, could not be used to further fine map the grain length phenotype. However, looking at the individual RILs that comprised the six intermediate RIL groups showed that individual RILs had a range of classifications within a group, but within each RIL itself the classifications were relatively stable across trials (Table 4.11). In other words, unlike in the initial 6A fine mapping where the intermediate groups (SR Gr1.2 and 1.9; Table 4.10) consisted of RIL lines that were themselves classed as intermediate, in this case with the 5A RILs intermediate groups were often classed as such because they contained RIL lines that had different classifications. For example, CB Gr1.12 was classed as intermediate (CB) and contained three independent RILs: HR-CB5, HR-CB30 and HR-29. HR-CB5 was classed as a B-type across all trials and in each of the four trials individually. HR-CB30 was classed as CB overall and in three of the four trials (B in 2016). Finally, HR-CB29 was classed as C-type overall and in two of the three trials in which it was grown (CB in 2014).

These phenotypic differences within a RIL group could be explained by the individual RILs within a group having different recombination events but the marker density across the interval was not

high enough to identify this. For example, all lines in CB Gr1.12 have recombination between *BS00186083* and *JBRNASeq\_1*, but the exact location of the recombination may be different in each of the three RILs. Further fine-mapping was therefore performed using the original individual RIL lines from RIL groups with recombination across the fine-mapped 7.46 cM interval. Figure 4.12 shows the 14 RILs with recombination across the 7.46 cM interval that could be assigned unambiguously as C or B types. Using these 14 RILs, the grain length effect was fine-mapped to a slightly narrower 6.59 cM interval between *BS00182017* and a group of four linked markers (*BA00228977*, *JBRNASeq\_4*, *BA00165371*, *BA00379554*; Figure 4.12a, blue markers). Several RILs classified C or B had recombination within this interval but three of these RILs suggested conflicting mapping positions. HR-CB9 placed the grain length phenotype to the left of *BS00186083*, whilst HR-CB5 and HR-58 mapped grain length to the right of *BS00186083*. The grain length phenotype could therefore not be mapped to a narrower interval using this population. It is also worth noting that eleven of the individual RILs were classed themselves as intermediate.

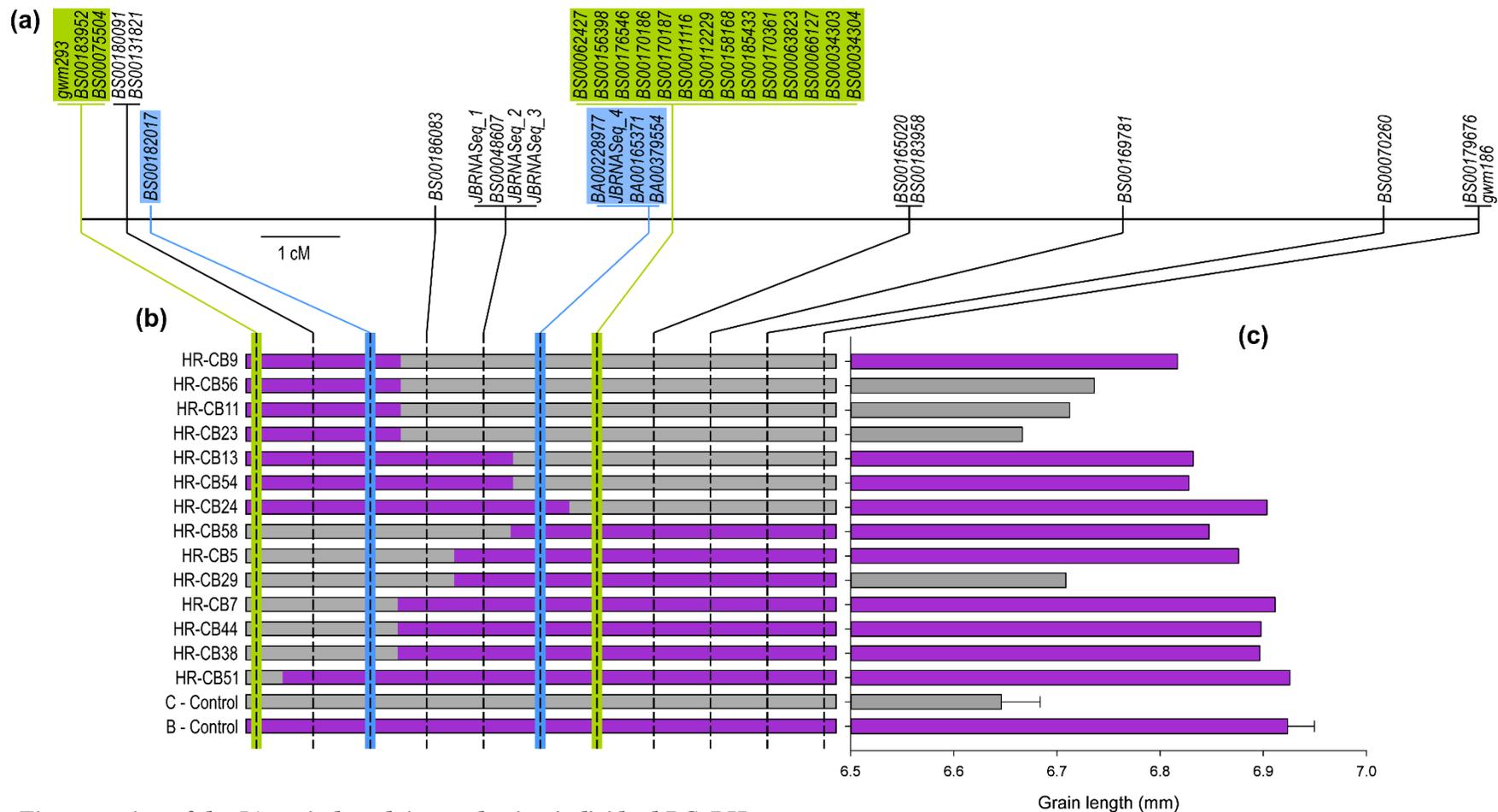
**Table 4.11: ANOVA adjusted mean grain length and Dunnett's test classification of individual BC<sub>4</sub> RILs used for initial 5A fine-mapping**

RIL group	Class	RIL	2014 (1m plots)		2015 (1m plots)		2015 (6m plots)		2016 (6m plots)		Overall	
			Length (mm)	Class	Length (mm)	Class						
CB Gr1.1	CB	HR-CB9	7.055	CB	6.805	CB	-	-	6.764	B	6.817	B
		HR-CB10	7.043	CB	6.832	B	-	-	6.702	CB	6.813	CB
		HR-CB26	7.030	C	6.739	CB	-	-	6.741	B	6.769	CB
		HR-CB14	6.993	C	6.720	CB	-	-	6.729	B	6.757	CB
		HR-CB46	6.992	C	6.762	CB	-	-	6.685	C	6.755	CB
		HR-CB56	6.936	C	6.771	CB	-	-	-	-	6.736	C
		HR-CB11	7.067	B	6.644	C	-	-	6.725	B	6.712	C
		HR-CB23	6.899	C	6.720	CB	-	-	6.552	C	6.666	C
CB Gr1.2	CB	HR-CB13	7.094	B	6.831	CB	6.633	CB	6.790	B	6.832	B
		HR-CB54	7.065	CB	6.857	CB	6.638	CB	6.744	B	6.828	B
		HR-CB2	6.988	C	6.845	CB	6.575	CB	6.670	C	6.771	CB
		HR-CB16	7.019	C	6.710	C	6.641	CB	6.725	B	6.762	CB
CB Gr1.3	B	HR-CB24	7.195	B	6.912	B	6.706	B	6.839	B	6.904	B
CB Gr1.4	B	HR-CB43	7.201	B	7.002	B	6.824	B	-	-	6.985	B
		HR-CB28	7.180	B	6.866	B	6.754	B	-	-	6.901	B
		HR-CB35	7.191	B	6.995	B	6.760	B	-	-	6.956	B
		HR-CB20	7.189	B	-	-	-	-	-	-	6.937	B
CB Gr1.5	B	HR-CB53	7.259	B	-	-	-	-	-	-	6.995	B
		HR-CB12	7.221	B	-	-	-	-	-	-	6.969	B
		HR-CB17	7.176	B	-	-	-	-	-	-	6.924	B
		HR-CB57	7.144	B	-	-	-	-	-	-	6.892	B
CB Gr1.6	B	HR-CB50	7.142	B	-	-	-	-	-	-	6.890	B

RIL group	Class	RIL	2014 (1m plots)		2015 (1m plots)		2015 (6m plots)		2016 (6m plots)		Overall	
			Length (mm)	Class	Length (mm)	Class						
CB Gr1.7	C	HR-CB21	6.934	C	-	-	-	-	-	-	6.682	C
		HR-CB34	6.925	C	-	-	-	-	-	-	6.673	C
		HR-CB31	6.898	C	-	-	-	-	-	-	6.634	C
		HR-CB1	6.865	C	-	-	-	-	-	-	6.613	C
CB Gr1.8	C	HR-CB33	7.004	C	-	-	-	-	-	-	6.752	CB
		HR-CB45	6.985	C	-	-	-	-	-	-	6.733	CB
		HR-CB39	6.986	C	-	-	-	-	-	-	6.721	CB
		HR-CB40	6.970	C	-	-	-	-	-	-	6.718	CB
		HR-CB37	6.960	C	-	-	-	-	-	-	6.708	CB
		HR-CB47	6.959	C	-	-	-	-	-	-	6.707	CB
		HR-CB41	6.956	C	-	-	-	-	-	-	6.704	C
		HR-CB36	6.955	C	-	-	-	-	-	-	6.703	C
		HR-CB42	6.938	C	-	-	-	-	-	-	6.686	C
HR-CB32	6.938	C	-	-	-	-	-	-	6.686	C		
CB Gr1.9	C	HR-CB19	7.026	CB	-	-	-	-	-	-	6.774	CB
		HR-CB3	6.872	C	-	-	-	-	-	-	6.620	C
		HR-CB60	6.883	C	-	-	-	-	-	-	6.618	C
		HR-CB59	6.857	C	-	-	-	-	-	-	6.605	C
		HR-CB52	6.819	C	-	-	-	-	-	-	6.567	C
CB Gr1.10	CB	HR-CB27	6.977	C	6.806	CB	6.582	CB	-	-	6.765	CB
		HR-CB55	7.017	C	6.722	CB	6.590	CB	-	-	6.743	CB
		HR-CB8	6.964	C	-	-	-	-	-	-	6.711	CB
		HR-CB15	7.010	C	6.704	C	6.518	C	-	-	6.711	C
		HR-CB22	6.981	C	6.642	C	6.593	CB	-	-	6.703	C
		HR-CB6	6.955	C	-	-	-	-	-	-	6.690	CB

RIL group	Class	RIL	2014 (1m plots)		2015 (1m plots)		2015 (6m plots)		2016 (6m plots)		Overall	
			Length (mm)	Class	Length (mm)	Class						
CB Gr1.11	CB	HR-CB58	7.031	CB	6.876	B	-	-	6.744	B	6.847	B
		HR-CB25	7.043	CB	6.766	CB	6.575	CB	6.647	C	6.758	CB
CB Gr1.12	CB	HR-CB5	7.138	B	6.953	B	6.681	B	6.746	B	6.876	B
		HR-CB30	7.028	CB	6.788	CB	6.588	CB	6.711	B	6.782	CB
		HR-CB29	7.050	CB	6.676	C	6.518	C	-	-	6.709	C
CB Gr1.13	CB	HR-CB7	7.150	B	7.016	B	-	-	6.744	B	6.912	B
		HR-CB44	7.077	B	6.943	B	-	-	6.823	B	6.898	B
		HR-CB38	7.118	B	6.932	B	-	-	6.807	B	6.897	B
		HR-CB18	7.012	C	6.721	CB	-	-	6.683	C	6.739	CB
CB Gr1.14	CB	HR-CB4	7.079	B	-	-	-	-	-	-	6.827	CB
CB Gr1.15	B	HR-CB48	7.128	B	-	-	-	-	-	-	6.864	CB
		HR-CB51	7.178	B	-	-	-	-	-	-	6.926	B
C-control (5A-)	C	HR-CB37-C	6.996	C	6.745	C	6.612	C	-	-	6.759	C
		HR-CB7-C	6.746	C	6.684	C	6.430	C	-	-	6.616	C
		BC4-5 (NIL)	-		6.551	C	6.419	C	6.597	C	6.603	C
		BC4-17 (NIL)	6.895	C	6.559	C	6.434	C	6.564	C	6.607	C
B-control (5A+)	B	HR-CB9-C	7.242	B	6.981	B	6.848	B	-	-	6.996	B
		HR-CB38-C	7.202	B	6.892	B	6.780	B	-	-	6.924	B
		BC4-6 (NIL)	-	-	6.873	B	6.746	B	6.819	B	6.893	B
		BC4-19 (NIL)	7.213	B	6.824	B	6.713	B	6.814	B	6.880	B

Length (mm) are the ANOVA adjusted means of grain length in each trial (or overall in the final column) each incorporating at least five replicates. Classifications were assigned using Dunnett's test to compare each line to a control (C-Control (5A-; short grains) and B-Control (5A+; long grains)): C = significantly different from the B-Control and not significantly different from the C-Control; B = significantly different from the C-Control and no significantly different from the B-Control; CB = not significantly different from the C-Control or the B-Control.



**Figure 4.12: Fine-mapping of the 5A grain length interval using individual  $BC_4$  RILs**

(a) Genetic map of the 5A QTL mapping interval based on the original set of  $BC_4$  recombinant inbred lines (RILs). Markers highlighted in green are the flanks for the fine-mapped grain length interval defined by mapping with RIL groups. Markers highlighted in blue are the flanks for the grain length interval defined using the individual RILs in (b). (b) Graphical genotypes of RILs showing the allele at each marker (Charger-like (grey; 5A-) or Badger-like (purple; 5A+)). (c) ANOVA adjusted mean grain lengths for each RIL across four field trials (1m plots 2014-2015 and 6m plots 2015-2016). Bars are coloured according to classification as C-control (5A-; grey) or B-control (5A+; purple) like according to a Dunnett's test.

#### **4.2.2.2 Haplotype analysis of the 5A grain length interval**

##### **4.2.2.2.1 The 5A grain length interval is not fixed in UK germplasm**

Defining the physical positions of markers across the interval allowed a haplotype analysis to be conducted to understand how the 5A grain length interval(s) behave in other UK wheat varieties. Exome capture data from 20 UK wheat varieties was used to identify SNPs with respect to IWGSC Chinese Spring chromosome survey sequence (IWGSC, 2014) (Figure 4.13; varieties with circles). The position of each SNP in RefSeq v1.0 was identified using BLAST. SNPs located between 300-400 Mbp on chromosome 5A were selected as this region encompassed the fine-mapped grain length interval. A total of 205 SNPs with respect to Chinese Spring were identified in this 100 Mbp region, however, 122 of these SNPs were monomorphic in all 20 varieties and therefore, not informative for this analysis. The 83 remaining SNPs were summarised into 22 groups of SNPs that showed the same pattern across the 20 varieties and a representative SNP was selected for each group (JBHap001-022; Figure 4.13). The 20 varieties were assigned to 12 distinct haplotype groups based on their genotypes across the 22 SNPs, with over half of the varieties contained within two groups (Group 2: four varieties, Group 4: seven varieties). To determine which haplotype groups the parental varieties of the 5A QTL (Charger (5A-) and Badger (5A+)) belonged to, markers were designed for each of the 22 SNPs. Both the parental varieties and a pair of 5A NILs were genotyped with the 22 haplotype markers. Using this genotyping Charger/5A- was assigned to Group 4, whilst Badger/5A+ was assigned to Group 12 (Figure 4.13; grey and purple highlighted varieties). The fact that Charger and Badger fall into different haplotype groups suggests that the positive 5A grain length allele(s) are not yet fixed in UK germplasm. Additionally, Charger belonged to the largest haplotype Group (4) whilst Group 12 (containing Badger) was small and quite different from the other haplotype groups. This could suggest that the Charger allele (i.e. the negative 5A allele) is more prevalent within UK breeding programmes and so the selection of the Badger (positive) allele could offer improvements in grain size. An alternative explanation could be that selection for the grain length effect has eroded the long range haplotype of Group 12, and hence, the positive allele is present in many varieties but not visible in this analysis. Although there are no clear recombination breakpoints to suggest this alternative explanation in this data, further analysis with additional lines and better defined intervals would be required to establish exactly how this QTL has been selected during the breeding process.

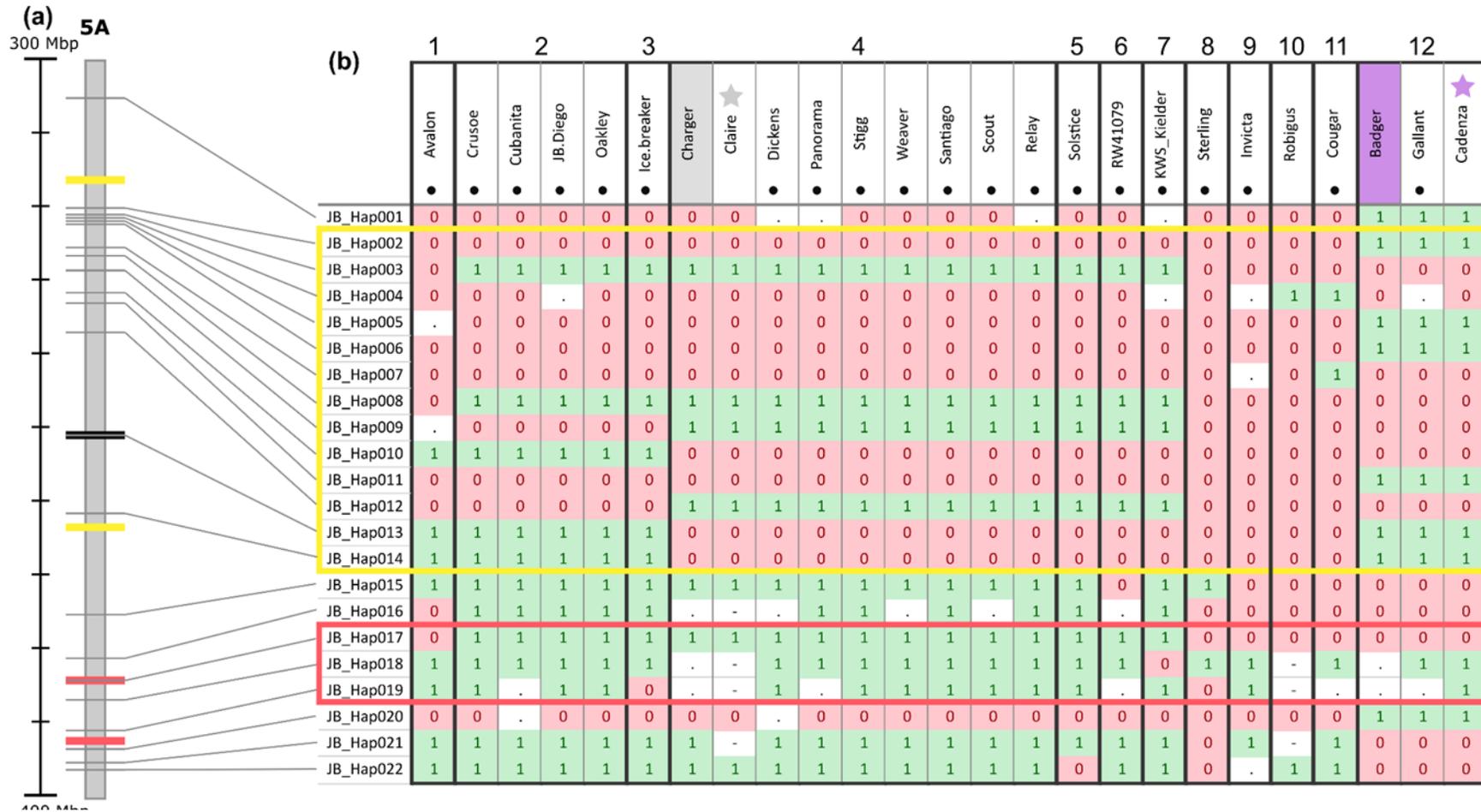


Figure 4.13: Haplotype analysis across the 5A grain length interval

(a) Grey lines are position of SNPs JB\_Hap001-022 on chromosome 5A according to RefSeq v1.0. Yellow lines are the flanking markers of Region 1 and pink lines are the flanks of Region 2. (b) Genotype of each of the SNPs JB\_Hap001-022 in different wheat varieties. '1' indicates a SNP with respect to the Chinese Spring reference, '0' indicates the Chinese Spring allele and '.' indicates a missing data point. Numbers are groups of varieties with the same genotype across all 22 SNPs. Varieties with circles indicate the 20 UK varieties for which exome capture was available.

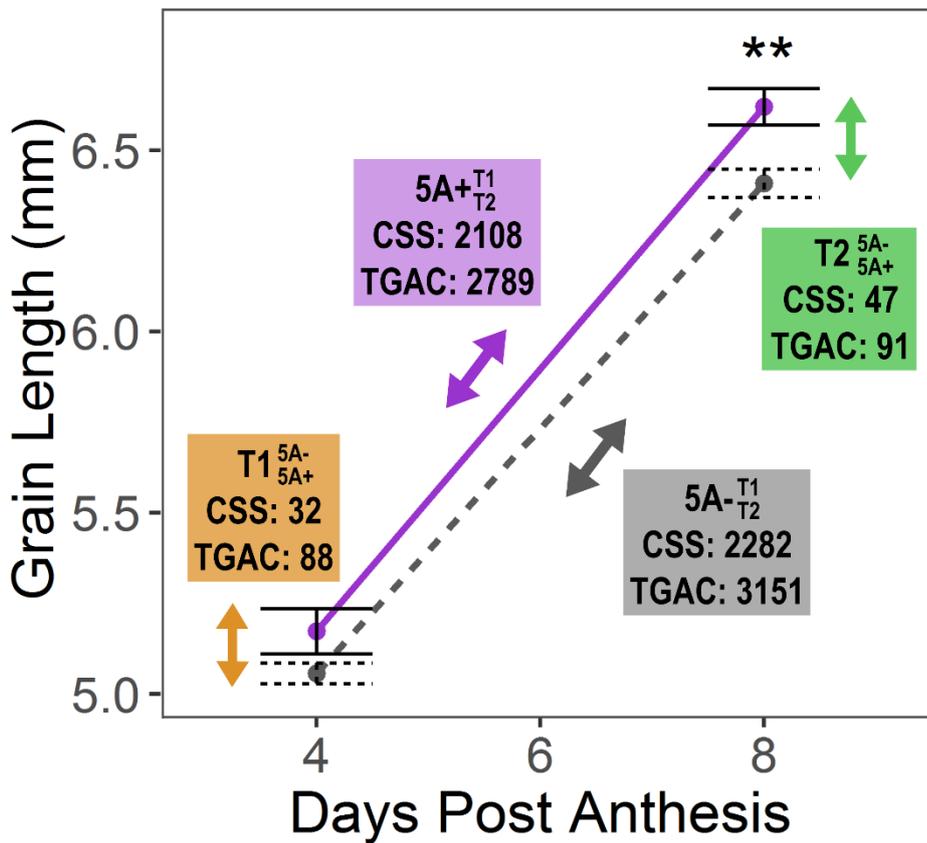
#### **4.2.2.2.2** Charger and Badger have the same haplotypes as sequenced varieties

Three additional varieties were also characterised with the 22 haplotype markers: Claire, Cadenza and Robigus. These three varieties were selected as they have been sequenced by the Earlham Institute. Claire shared a haplotype with Charger (Group 4; Figure 4.13, grey star) whilst Cadenza had the same haplotype as Badger (Group 12; Figure 4.13, purple star). Robigus was not identical to any of the haplotype groups and was allocated to its own group (10) although it was highly similar to Groups 9 and 11. The fact that Charger and Badger have the same haplotypes as sequenced varieties means that the genome sequences of Claire and Cadenza can be used as proxies for the parental varieties of the 5A grain length QTL.

### 4.3 Comparative transcriptomics of 5A NILs

#### 4.3.1 RNA-sequencing of 5A near isogenic lines

RNA-seq was performed on whole grains from two of the 5A grain length NILs (Brinton *et al.*, 2017). The time point when NILs showed the first significant differences in grain length (8 days post anthesis (dpa); T2) and the preceding time point (4 dpa; T1) were selected to capture differences in gene expression occurring during this period (Figure 4.14). We hypothesised that, although there was no significant difference in the grain length phenotype at T1, phenotypic differences were beginning to emerge and gene expression changes influencing this may already be occurring. Over 362 M reads across all 12 samples were obtained (two time points, two NILs, three biological replicates), with individual samples ranging from 15.0 M to 53.6 M reads and an average of 30.2 M reads (standard error  $\pm$  3.5 M reads) per sample (Table 4.12). Reads were aligned to two different transcriptome sequences from the reference wheat variety, Chinese Spring: the IWGSC Chromosome Survey Sequence (CSS) (IWGSC, 2014) and TGACv1 (TGAC) (Clavijo *et al.*, 2017b) reference. On average, across samples,  $69.8 \pm 0.3$  % of reads aligned to the CSS reference, whilst  $84.4 \pm 0.2$  % of reads aligned to the TGAC reference.



**Figure 4.14: Differentially expressed genes between 5A NILs across time**

RNA-seq was carried out on whole grain RNA samples taken in 4 different conditions: 5A- (short grains) and 5A+ (long grains) NILs at 4 days post anthesis (dpa; T1) and 8 dpa (T2). These were selected as the time point when the first significant difference ( $P < 0.01$ , asterisks) in grain length was observed between 5A- (grey, dashed line, short grains) and 5A+ (purple, solid line, long grains) and the preceding time point. Differentially expressed (DE) transcripts were identified for four comparisons ( $q$ -value  $< 0.05$ ). Coloured boxes indicate the numbers of DE transcripts identified for each comparison using alignments to either the IWGSC Chinese Spring Survey Sequence (CSS) or the TGACv1 (TGAC) Chinese Spring reference transcriptomes. Two ‘across time’ comparisons: 5A- $_{T1}^{T2}$  (grey box; comparing T1 and T2 samples of the 5A- NIL) and 5A+ $_{T1}^{T2}$  (purple box; comparing T1 and T2 samples of the 5A+ NIL), and two ‘between NIL’ comparisons: T1 $_{5A-}^{5A+}$  (orange box; comparing 5A- and 5A+ NILs at T1) and T2 $_{5A-}^{5A+}$  (green box; comparing 5A- and 5A+ NILs at T2).

*Table 4.12: Mapping summary of RNA-seq samples*

Genotype	Time point	Replicate	Reads	CSS gene models		TGAC gene models	
				Reads pseudoaligned	% reads pseudoaligned	Reads pseudoaligned	% reads pseudoaligned
5A -	1	1	24,443,658	17,072,939	69.85	20,549,681	84.07
5A -	1	2	34,441,799	23,349,288	67.79	28,483,090	82.70
5A -	1	3	23,462,705	16,220,597	69.13	19,664,859	83.81
5A -	2	1	21,333,672	14,839,724	69.56	18,052,324	84.62
5A -	2	2	14,967,302	10,632,519	71.04	12,803,552	85.54
5A -	2	3	35,522,754	25,491,523	71.76	30,297,336	85.29
5A +	1	1	19,267,564	13,520,181	70.17	16,317,352	84.69
5A +	1	2	22,299,102	15,479,234	69.42	18,780,525	84.22
5A +	1	3	30,531,539	20,789,582	68.09	25,436,453	83.31
5A +	2	1	51,637,607	36,192,489	70.09	43,739,451	84.70
5A +	2	2	53,575,232	37,956,887	70.85	45,497,914	84.92
5A +	2	3	30,553,421	21,604,895	70.71	25,984,674	85.05
		Total	362,036,355	253,149,858	-	305,607,211	-
		Mean	30,169,696	21,095,822	69.87	25,467,268	84.41

### 4.3.2 Comparison between Chinese Spring reference transcriptomes

A transcript was defined as expressed if it had an average abundance of > 0.5 transcripts per million (tpm) in at least one of the four conditions (2 NILs x 2 time points). This resulted in 62.5 % (64,020) and 37.1% (101,652) of the transcripts being expressed in the CSS and TGAC transcriptomes, respectively. Differentially expressed transcripts (q value < 0.05) were defined using sleuth (Pimentel *et al.*, 2017) and four pairwise comparisons were performed: two 'across time' and two 'between NIL' comparisons. The 'across time' analyses consisted of a comparison between T1 and T2 samples of the 5A- NIL (hereafter symbolised as  $5A-_{T2}^{T1}$ ; Figure 4.14, grey) and the corresponding comparison for the 5A+ NIL samples (hereafter  $5A+_{T2}^{T1}$ ; Figure 4.14, purple). In both cases, the T1 sample was used as the control condition, so transcripts were considered as upregulated or downregulated with respect to T1. The 'between NIL' analyses consisted of a comparison between the 5A- and 5A+ NILs at T1 (hereafter  $T1_{5A+}^{5A-}$ ; Figure 4.14, orange), and a comparison between the 5A- and 5A+ NILs at T2 (hereafter  $T2_{5A+}^{5A-}$ ; Figure 4.14, green). In both cases, the recurrent parent 5A- NIL was used as the control genotype. In all cases, more DE transcripts were identified in the TGAC compared with the CSS transcriptome, and similar trends were observed for both references across the four comparisons (Figure 4.14).

The comparison with the fewest DE transcripts ( $T1_{5A+}^{5A-}$ ; 32 and 88 DE transcripts for CSS and TGAC, respectively) was selected to conduct a more in depth analysis of the alignments and references. For all DE transcripts from each alignment the equivalent transcript/gene model was identified in the other reference sequence using *Ensembl* plants release 35 and the gene models were compared (Table 4.13).

**Table 4.13: Comparison between TGAC and CSS gene models**

TGAC transcript ID	CSS transcript ID	DE in TGAC?	TGAC q value	DE in CSS?	CSS q value	Comparison class
TRIAE_CS42_5AL_TGACv1_378188_AA1251790.1	Traes_5AL_CA424FE08.2	Y	2.92E-02	N	1.00E+00	CSS 3' truncation
TRIAE_CS42_2DL_TGACv1_157942_AA0502830.1	Traes_2DL_E1640BFDC.1	Y	4.21E-02	N	1.00E+00	CSS 5' and 3' truncation
TRIAE_CS42_5AL_TGACv1_374078_AA1189690.1	Traes_5AL_8BF894427.2	Y	6.35E-07	N	1.00E+00	CSS 5' and 3' truncation
TRIAE_CS42_5AL_TGACv1_374446_AA1200420.1	Traes_5AL_0573B44BE.1	Y	2.81E-02	N	1.00E+00	CSS 5' and 3' truncation
TRIAE_CS42_5AL_TGACv1_374560_AA1203240.1	Traes_5AL_32B5C730F.1	Y	3.22E-02	N	1.00E+00	CSS 5' and 3' truncation
TRIAE_CS42_5AL_TGACv1_377520_AA1247660.1	Traes_5AL_55BB0BEFC.1	Y	4.71E-02	N	1.00E+00	CSS 5' and 3' truncation
TRIAE_CS42_5AL_TGACv1_377986_AA1250630.1	Traes_5AL_999D96884.1	Y	6.14E-08	Y	1.62E-09	CSS 5' and 3' truncation
TRIAE_CS42_5AL_TGACv1_378334_AA1252720.1	Traes_5AL_1639C7AB0.1	Y	1.59E-10	N	NA	CSS 5' and 3' truncation
TRIAE_CS42_5AS_TGACv1_393493_AA1273190.4	Traes_5AS_9D5B8EA01.1	Y	7.60E-10	Y	1.69E-12	CSS 5' and 3' truncation
TRIAE_CS42_5DL_TGACv1_433728_AA1420750.1	Traes_5DL_531A38273.1	Y	1.63E-02	N	NA	CSS 5' and 3' truncation
TRIAE_CS42_7DS_TGACv1_622195_AA2034920.1	Traes_7DS_E14CFC6F2.2	Y	1.48E-04	Y	3.77E-05	CSS 5' and 3' truncation
TRIAE_CS42_5AS_TGACv1_392838_AA1265240.1	Traes_5BL_6C1EFA808.1	Y	1.60E-07	Y	4.03E-02	CSS 5' and 3' truncation + chromosome
TRIAE_CS42_5AS_TGACv1_392838_AA1265240.2	Traes_5BL_6C1EFA808.1	Y	2.92E-02	Y	4.03E-02	CSS 5' and 3' truncation + chromosome
TRIAE_CS42_5BS_TGACv1_427448_AA1393420.1	Traes_1AS_2B7CD7B59.1	Y	7.13E-07	N	1.00E+00	CSS 5' and 3' truncation + chromosome
TRIAE_CS42_5AL_TGACv1_373986_AA1186560.2	Traes_5AL_3AA4476D6.1	Y	2.24E-05	Y	6.41E-51	CSS 5' truncation
TRIAE_CS42_5AL_TGACv1_373986_AA1186560.3	Traes_5AL_3AA4476D6.1	Y	4.64E-07	Y	6.41E-51	CSS 5' truncation
TRIAE_CS42_5AL_TGACv1_373995_AA1186970.1	Traes_5AL_D57725ABD.1	N	1.00E+00	Y	2.03E-05	CSS 5' truncation
TRIAE_CS42_5AL_TGACv1_374080_AA1189800.1	Traes_5AL_F1F202C88.1	Y	1.65E-36	Y	4.13E-29	CSS 5' truncation
TRIAE_CS42_5AL_TGACv1_374080_AA1189810.1	Traes_5AL_5DE16F8EA.2	Y	6.35E-07	Y	1.20E-10	CSS 5' truncation
TRIAE_CS42_5AL_TGACv1_374097_AA1190260.1	Traes_5AL_1F7681FE3.1	Y	4.64E-07	NA	NA	CSS 5' truncation
TRIAE_CS42_5AL_TGACv1_374319_AA1196780.1	Traes_5AL_FCDD18A4D.1	Y	1.92E-03	Y	2.36E-03	CSS 5' truncation
TRIAE_CS42_5AL_TGACv1_374542_AA1202810.4	Traes_5AL_00CC4E7C6.1	Y	3.61E-07	Y	3.69E-11	CSS 5' truncation

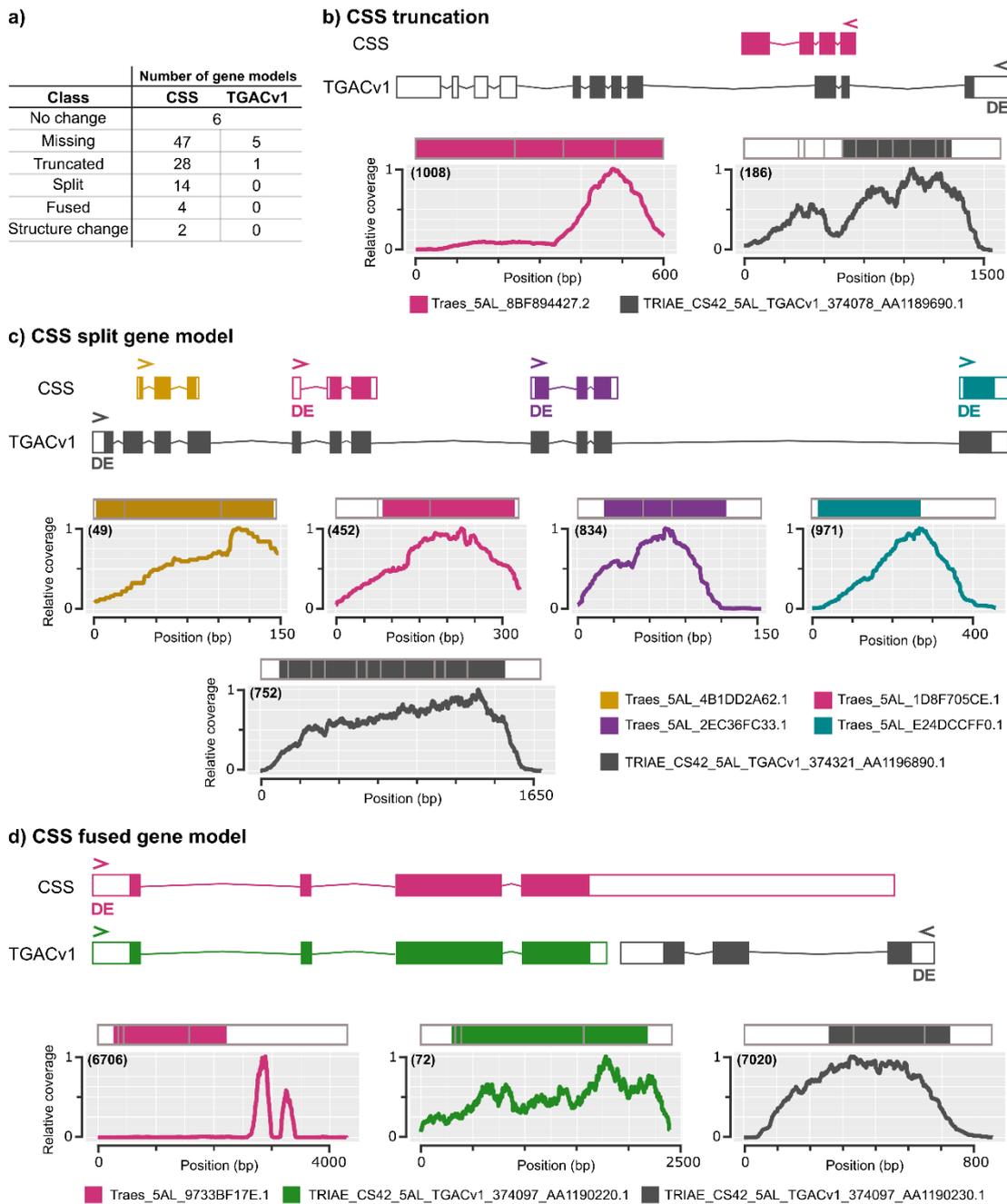
TGAC transcript ID	CSS transcript ID	DE in TGAC?	TGAC q value	DE in CSS?	CSS q value	Comparison class
TRIAE_CS42_5AL_TGACv1_375361_AA1220430.2	Traes_5AL_385883702.1	Y	4.21E-02	Y	3.33E-04	CSS 5' truncation
TRIAE_CS42_5AL_TGACv1_375845_AA1227980.1	Traes_5AL_2EDDF65BE.2	Y	1.00E-03	Y	7.99E-06	CSS 5' truncation
TRIAE_CS42_5AL_TGACv1_375845_AA1227990.1	Traes_5AL_AC299D3FF.1	Y	4.83E-15	Y	7.37E-11	CSS 5' truncation
TRIAE_CS42_5AL_TGACv1_376402_AA1236390.1	Traes_5AL_DD1665D87.2	Y	7.07E-04	Y	2.34E-02	CSS 5' truncation
TRIAE_CS42_5AL_TGACv1_376619_AA1239170.1	Traes_5AL_B8B668113.1	N	1.00E+00	Y	4.52E-02	CSS 5' truncation
TRIAE_CS42_5AL_TGACv1_376953_AA1242690.2	Traes_5AL_F09A6AECA.2	Y	1.53E-13	Y	9.28E-38	CSS 5' truncation
TRIAE_CS42_1AS_TGACv1_019083_AA0060340.1	NC	Y	1.42E-05	NC	NC	CSS missing
TRIAE_CS42_2AL_TGACv1_095650_AA0313320.1	NC	Y	4.87E-04	NC	NC	CSS missing
TRIAE_CS42_2AL_TGACv1_095650_AA0313330.1	NC	Y	4.94E-03	NC	NC	CSS missing
TRIAE_CS42_2AL_TGACv1_095956_AA0316040.1	NC	Y	5.16E-09	NC	NC	CSS missing
TRIAE_CS42_2AL_TGACv1_095956_AA0316050.1	NC	Y	1.11E-10	NC	NC	CSS missing
TRIAE_CS42_2DS_TGACv1_177196_AA0568430.1	NC	Y	4.57E-06	NC	NC	CSS missing
TRIAE_CS42_2DS_TGACv1_177488_AA0578600.1	NC	Y	4.65E-02	NC	NC	CSS missing
TRIAE_CS42_3AS_TGACv1_211411_AA0689940.1	NC	Y	1.29E-02	NC	NC	CSS missing
TRIAE_CS42_3DL_TGACv1_250330_AA0866270.1	NC	Y	2.65E-02	NC	NC	CSS missing
TRIAE_CS42_3DL_TGACv1_250633_AA0871340.1	NC	Y	9.45E-04	NC	NC	CSS missing
TRIAE_CS42_4BL_TGACv1_321219_AA1057420.1	NC	Y	2.23E-02	NC	NC	CSS missing
TRIAE_CS42_5AL_TGACv1_374025_AA1188070.1	NC	Y	8.44E-04	NC	NC	CSS missing
TRIAE_CS42_5AL_TGACv1_374097_AA1190230.1	NC	Y	2.32E-05	NC	NC	CSS missing
TRIAE_CS42_5AL_TGACv1_374249_AA1195190.1	NC	Y	8.01E-26	NC	NC	CSS missing
TRIAE_CS42_5AL_TGACv1_374374_AA1198360.1	NC	Y	8.15E-04	NC	NC	CSS missing
TRIAE_CS42_5AL_TGACv1_374446_AA1200410.1	NC	Y	4.06E-10	NC	NC	CSS missing

TGAC transcript ID	CSS transcript ID	DE in TGAC?	TGAC q value	DE in CSS?	CSS q value	Comparison class
TRIAE_CS42_5AL_TGACv1_374657_AA1205780.1	NC	Y	4.65E-02	NC	NC	CSS missing
TRIAE_CS42_5AL_TGACv1_374675_AA1206250.1	NC	Y	1.63E-02	NC	NC	CSS missing
TRIAE_CS42_5AL_TGACv1_375085_AA1215790.1	NC	Y	5.97E-04	NC	NC	CSS missing
TRIAE_CS42_5AL_TGACv1_375493_AA1222690.2	NC	Y	1.83E-06	NC	NC	CSS missing
TRIAE_CS42_5AL_TGACv1_375721_AA1226170.1	NC	Y	3.46E-04	NC	NC	CSS missing
TRIAE_CS42_5AL_TGACv1_375857_AA1228100.1	NC	Y	3.43E-03	NC	NC	CSS missing
TRIAE_CS42_5AL_TGACv1_376076_AA1231790.1	NC	Y	1.08E-03	NC	NC	CSS missing
TRIAE_CS42_5AL_TGACv1_376877_AA1241920.1	NC	Y	2.33E-02	NC	NC	CSS missing
TRIAE_CS42_5AL_TGACv1_376877_AA1241930.1	NC	Y	2.47E-30	NC	NC	CSS missing
TRIAE_CS42_5AL_TGACv1_376953_AA1242680.1	NC	Y	1.62E-17	NC	NC	CSS missing
TRIAE_CS42_5AL_TGACv1_376953_AA1242690.1	NC	Y	1.59E-13	NC	NC	CSS missing
TRIAE_CS42_5AS_TGACv1_393235_AA1270150.1	NC	Y	3.09E-05	NC	NC	CSS missing
TRIAE_CS42_5AS_TGACv1_393473_AA1272910.1	NC	Y	1.80E-05	NC	NC	CSS missing
TRIAE_CS42_5AS_TGACv1_393577_AA1274040.1	NC	Y	1.13E-26	NC	NC	CSS missing
TRIAE_CS42_5AS_TGACv1_393580_AA1274130.1	NC	Y	1.17E-18	NC	NC	CSS missing
TRIAE_CS42_5AS_TGACv1_393696_AA1275280.2	NC	Y	1.67E-35	NC	NC	CSS missing
TRIAE_CS42_5AS_TGACv1_393726_AA1275550.1	NC	Y	3.48E-10	NC	NC	CSS missing
TRIAE_CS42_5AS_TGACv1_393783_AA1275990.1	NC	Y	4.55E-22	NC	NC	CSS missing
TRIAE_CS42_5AS_TGACv1_393897_AA1277010.1	NC	Y	7.13E-07	NC	NC	CSS missing
TRIAE_CS42_5AS_TGACv1_394352_AA1279770.1	NC	Y	1.79E-04	NC	NC	CSS missing
TRIAE_CS42_5AS_TGACv1_394531_AA1280840.1	NC	Y	6.00E-03	NC	NC	CSS missing
TRIAE_CS42_5AS_TGACv1_395074_AA1282530.1	NC	Y	6.69E-18	NC	NC	CSS missing

TGAC transcript ID	CSS transcript ID	DE in TGAC?	TGAC q value	DE in CSS?	CSS q value	Comparison class
TRIAE_CS42_5AS_TGACv1_395084_AA1282570.1	NC	Y	7.37E-03	NC	NC	CSS missing
TRIAE_CS42_5AS_TGACv1_402215_AA1284070.1	NC	Y	2.32E-05	NC	NC	CSS missing
TRIAE_CS42_6AL_TGACv1_471516_AA1510240.1	NC	Y	3.11E-03	NC	NC	CSS missing
TRIAE_CS42_6BL_TGACv1_503194_AA1627460.1	NC	Y	1.12E-05	NC	NC	CSS missing
TRIAE_CS42_7AS_TGACv1_571015_AA1843630.1	NC	Y	1.55E-02	NC	NC	CSS missing
TRIAE_CS42_7BL_TGACv1_577371_AA1873630.1	NC	Y	5.06E-05	NC	NC	CSS missing
TRIAE_CS42_7BS_TGACv1_592547_AA1940160.1	NC	Y	2.39E-04	NC	NC	CSS missing
TRIAE_CS42_7DS_TGACv1_621701_AA2023630.1	NC	Y	6.64E-05	NC	NC	CSS missing
TRIAE_CS42_U_TGACv1_641674_AA2101090.1	NC	Y	6.67E-05	NC	NC	CSS missing
TRIAE_CS42_2BS_TGACv1_148693_AA0494610.1	Traes_2BS_009718F07.2	Y	2.65E-02	N	1.00E+00	CSS split
TRIAE_CS42_2BS_TGACv1_148693_AA0494610.1	Traes_2BS_2272AAEE2.2	Y		N	1.00E+00	CSS split
TRIAE_CS42_5AL_TGACv1_374231_AA1194360.2	Traes_5AL_C1E3FCB4F.1	Y	4.60E-03	N	1.00E+00	CSS split
TRIAE_CS42_5AL_TGACv1_374231_AA1194360.2	Traes_5AL_CD19FF15F.1	Y		N	1.00E+00	CSS split
TRIAE_CS42_5AL_TGACv1_374321_AA1196890.1	Traes_5AL_4B1DD2A62.1	Y	3.09E-12	N	NA	CSS split
TRIAE_CS42_5AL_TGACv1_374321_AA1196890.1	Traes_5AL_1D8F705CE.1	Y		Y	2.27E-05	CSS split
TRIAE_CS42_5AL_TGACv1_374321_AA1196890.1	Traes_5AL_2EC36FC33.1	Y		Y	1.89E-11	CSS split
TRIAE_CS42_5AL_TGACv1_374321_AA1196890.1	Traes_5AL_E24DCCFF0.1	Y		N	1.10E-01	CSS split
TRIAE_CS42_5AL_TGACv1_375394_AA1220930.4	Traes_5AL_67878B82B.1	Y	0.00E+00	N	1.00E+00	CSS split
TRIAE_CS42_5AL_TGACv1_375394_AA1220930.4	Traes_5AL_0C2D144B0.1	Y		N	1.00E+00	CSS split
TRIAE_CS42_5AS_TGACv1_393119_AA1268700.1	Traes_5AS_AF0876292.1	Y	3.11E-03	N	1.00E+00	CSS split
TRIAE_CS42_5AS_TGACv1_393119_AA1268700.1	Traes_5AS_25E2451D6.1	Y		N	1.00E+00	CSS split
TRIAE_CS42_7AL_TGACv1_556075_AA1754700.1	Traes_7AL_A29227860.2	Y	1.25E-02	N	1.00E+00	CSS split

TGAC transcript ID	CSS transcript ID	DE in TGAC?	TGAC q value	DE in CSS?	CSS q value	Comparison class
TRIAE_CS42_7AL_TGACv1_556075_AA1754700.1	Traes_7AL_773C8EC8C.1	Y		N	NA	CSS split
TRIAE_CS42_5AL_TGACv1_374233_AA1194500.1	Traes_5AL_158704A70.1	N	NA	Y	5.32E-13	CSS structure change
TRIAE_CS42_5AL_TGACv1_374413_AA1199590.1	Traes_5AL_A9CF39101.1	Y	3.27E-02	N	1.00E+00	CSS structure change
TRIAE_CS42_1BS_TGACv1_049354_AA0149980.1	Traes_1BS_C59E4945B.2	Y	4.32E-02	Y	3.00E-02	No change
TRIAE_CS42_5AL_TGACv1_375845_AA1227940.1	Traes_5AL_1C4AB8F62.1	N	1.00E+00	Y	2.40E-03	No change
TRIAE_CS42_5AL_TGACv1_376107_AA1232210.1	Traes_5AL_70C442FE1.2	N	NA	Y	2.27E-03	No change
TRIAE_CS42_5AS_TGACv1_392558_AA1260860.1	Traes_5AS_34C5341E4.1	Y	3.90E-47	Y	3.57E-51	No change
TRIAE_CS42_5AS_TGACv1_394776_AA1281770.1	Traes_5AS_78CA97493.2	Y	1.43E-06	Y	6.58E-07	No change
TRIAE_CS42_5AS_TGACv1_394776_AA1281770.3	Traes_5AS_78CA97493.2	Y	4.11E-02	Y	6.58E-07	No change
TRIAE_CS42_5AS_TGACv1_393572_AA1273920.1	Traes_5AS_BD279FFF4.2	Y	2.41E-13	Y	6.05E-16	TGAC 5' truncation
NC	Traes_5AL_78644A5C4.1	NC	NC	Y	3.20E-08	TGAC missing
NC	Traes_5AL_BAB11D9B4.3	NC	NC	Y	2.10E-02	TGAC missing
NC	Traes_5BS_3B409615C.1	NC	NC	Y	1.95E-05	TGAC missing
NC	Traes_2AS_8F1446457.2	NC	NC	Y	1.19E-04	TGAC missing
NC	TRAES3BF002600020CFD_t1	NC	NC	Y	4.44E-02	TGAC missing
TRIAE_CS42_5AL_TGACv1_374727_AA1207530.1	Traes_5AL_F8182F2FB.1	Y	2.43E-03	Y	2.15E-02	CSS fused
TRIAE_CS42_5AL_TGACv1_374727_AA1207540.1	Traes_5AL_F8182F2FB.1	N	1.00E+00	Y		CSS fused
TRIAE_CS42_5AL_TGACv1_376411_AA1236550.1	Traes_5AL_58BA759B9.5	Y	1.06E-10	Y	1.82E-64	CSS fused
TRIAE_CS42_5AL_TGACv1_376411_AA1236560.1	Traes_5AL_58BA759B9.5	Y	1.20E-03	Y		CSS fused

For 64 of the TGAC DE transcripts no equivalent CSS DE transcript was identified, either because there was no corresponding CSS gene model (47 transcripts) or the expression change between NILs was non-significant for the CSS transcript. Analogously, eleven CSS DE transcripts did not have an equivalent TGAC gene model DE, five of which were due to there being no corresponding TGAC gene model annotated. Combining both sets identified 42 groups of equivalent gene models, 26 of which were differentially expressed in both alignments. Comparing these 42 groups and taking into account fused and split gene models within each dataset, there were 97 gene models in both datasets (50 CSS + 47 TGAC) (Figure 4.15a, Table 4.13). Of these, only six were identical between the CSS and TGAC references. All other discrepant gene models fell under categories included truncations in either reference, gene models that were split/fused in one reference sequence, and gene models that differed drastically in their overall structure.



**Figure 4.15: Comparison between CSS and TGACv1 gene models**

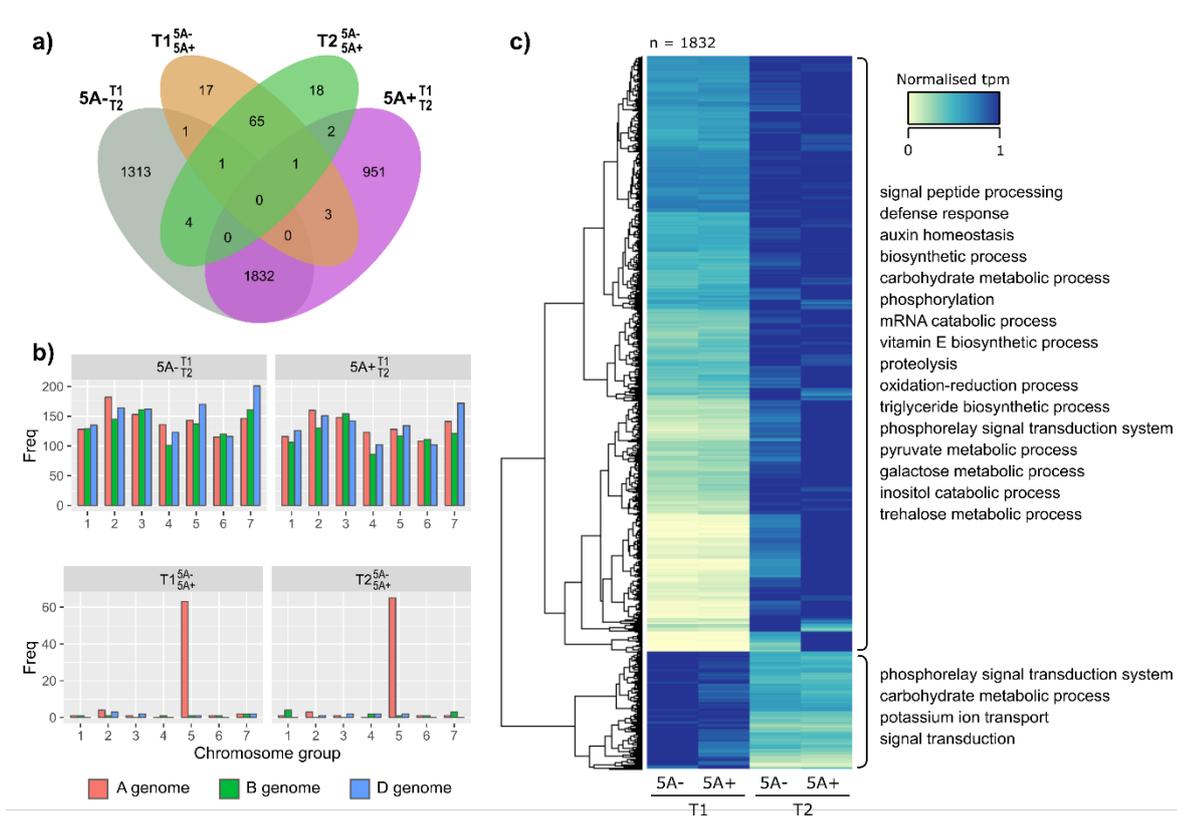
a) Discrepancies identified between gene models in the CSS and TGAC reference sequences and the number of gene models falling into categories. Panels b), c) and d) show specific examples of discrepancies. In each panel, a representation of the unspliced gene model is shown with exons as coloured boxes, untranslated regions as white boxes, and introns as thin lines. Graphs show the relative read coverage across the spliced transcript with the structure represented diagrammatically directly above each graph. The number in brackets shows the maximum absolute read depth for each gene model. > and < in the gene structures indicate the direction of transcription and a 'DE' indicates that the gene model was differentially expressed in T1<sub>5A-</sub><sub>5A+</sub> ( $q$  value < 0.05). For each panel transcript names are shown in the coloured legends.

For all discrepant gene models, transcriptome read mapping and an interspecies comparison was used to determine which gene model seemed most plausible. Figure 4.15b shows an example of the most commonly identified discrepancy where a gene model was truncated in the CSS reference (pink) relative to the TGAC reference (grey). The DE TGAC gene model was supported by this transcriptome data as read coverage was observed across the whole gene model whilst the coverage across the CSS gene model dropped at the position where an intron is predicted in the TGAC model. Another common discrepancy was a single gene model in one reference being split into multiple gene models in the other reference. Figure 4.15c shows an instance where a single DE TGAC gene model comprised four separate CSS gene models. In this case, all five gene models had coverage across the entire gene body, however, the single TGAC gene model was more similar to proteins from other species, suggesting that this single gene model was most likely correct. The final example (Figure 4.15d) shows two TGAC gene models that were fused into a single CSS gene model. The coverage across the CSS gene model was inconsistent, with most reads concentrated in the 3' untranslated region (UTR). The two TGAC gene models had more consistent coverage across the entire gene models and were both supported by protein alignments with other species. Interestingly, only the shorter TGAC gene model was DE (Figure 4.15d, grey), suggesting that differential expression of the CSS gene model was driven by the reads mapping to the putative 3' UTR rather than the coding regions of the transcript (Figure 4.15d, pink). Taking together the fact that a higher percentage of reads mapped to the TGAC gene models and that many more of the examined TGAC gene models were supported by interspecies comparison and expression data than the CSS gene models, all further analysis used the alignments to the TGAC gene models only.

### 4.3.3 Many DE transcripts during early grain development are shared between NILs

3,151 and 2,789 DE transcripts were identified across early grain development in 5A- $T_2^1$  and 5A+ $T_2^1$ , respectively (Figure 4.14, Figure 4.16a). The DE transcripts were evenly distributed across the 21 chromosomes, showing no overall bias towards any chromosome group or subgenome (Figure 4.16b). Approximately 60% (1,832) of the DE transcripts were shared between 5A- $T_2^1$  and 5A+ $T_2^1$  (Figure 3a) and 84% (1,532) of the shared transcripts were upregulated across time (Figure 4.16c). 41 significantly enriched gene ontology (GO) terms were identified in the upregulated transcripts (Table 4.14). Sixteen of the GO terms were associated with biological process and could be grouped under three parent GO terms: metabolic process (GO:0008152), defence response (GO:0006952) and biological regulation (GO:0065007) (Table 4.14; Figure 4.16b). Within metabolic process, we found terms associated with carbohydrate (GO:0005975) and pyruvate metabolism (GO:0006090), vitamin E (GO:0010189) and triglyceride biosynthesis (GO:0019432), mRNA catabolism (GO:0006402), proteolysis (GO:0006508) and phosphorylation (GO:0016310). Downregulated transcripts (300) were enriched for seven GO terms, four of which were associated with biological process: potassium ion transport (GO:0006813), signal transduction (GO:0007165),

phosphorelay signal transduction (GO:0000160) and carbohydrate metabolism (GO:0005975) (Figure 4.16c, Table 4.15). The overlap between enriched GO terms in the upregulated and downregulated transcripts (e.g. carbohydrate metabolism) suggests that different aspects of these processes are being differentially regulated during this early grain development stage. Many transcripts identified were only DE across early grain development in one of the two genotypes (i.e. unique to either the 5A- $\frac{T1}{T2}$  or 5A+ $\frac{T1}{T2}$  comparisons). However, many of these transcripts were borderline non-significant in the opposite genotype comparison illustrated by the fact that the distributions of q-values were skewed towards significance (Figure 4.17). Additionally, the uniquely DE transcripts were enriched for GO terms similar to the shared transcripts (Table 4.16 Table 4.17). Some GO terms, however, were only enriched in the uniquely DE transcripts, for example, cell wall organisation or biosynthesis (GO:0071554) and response to abiotic stimulus (GO:0009628). Overall, these results suggest that although there were some differences between genotypes, broadly similar biological processes were taking place in the grains of both the 5A NILs at the early stages of grain development.



**Figure 4.16: Overview of differentially expressed transcripts**

a) Venn diagram of differentially expressed (DE) transcripts ( $q < 0.05$ ) identified in 4 pairwise comparisons:  $T1^{5A-}_{5A+}$  (orange),  $T2^{5A-}_{5A+}$  (green),  $5A-^{T1}_{T2}$  (grey) and  $5A+^{T1}_{T2}$  (purple). b) Number of DE transcripts located on each chromosome for all comparisons. The  $5A-^{T1}_{T2}$  and  $5A+^{T1}_{T2}$  DE transcripts (top graphs) are evenly distributed across all 21 chromosomes, whereas  $T1^{5A-}_{5A+}$  and  $T2^{5A-}_{5A+}$  DE transcripts (bottom graphs) are concentrated on chromosome 5A. c) Heatmap of normalised tpm (transcripts per million) of common DE transcripts in  $5A-^{T1}_{T2}$  and  $5A+^{T1}_{T2}$  ( $n = 1,832$ ). Hierarchical clustering separated these into transcripts that were upregulated ( $n = 1,532$ ) and downregulated ( $n = 300$ ) across time. Significantly enriched GO terms (biological function only) for each group are shown on the right of the heatmap.

**Table 4.14: Enriched gene ontology (GO) terms in common upregulated transcripts differentially expressed (DE) in the 5A-  $T_2^1$  and 5A+  $T_2^1$  comparisons (n = 1532)**

GO term	DE transcripts	Total transcripts	Term ID	Ontology	Adjusted p-value
GO:0006952	72	162	defense response	BP	3.07E-83
GO:0000160	32	292	phosphorelay signal transduction system <small>Biological regulation</small>	BP	1.21E-15
GO:0006508	64	1455	Proteolysis <small>Metabolic process</small>	BP	8.87E-12
GO:0006465	8	27	signal peptide processing <small>Metabolic process</small>	BP	1.02E-06
GO:0055114	90	3409	oxidation-reduction process <small>Metabolic process</small>	BP	1.24E-05
GO:0005975	46	1455	carbohydrate metabolic process <small>Metabolic process</small>	BP	2.52E-04
GO:0006402	6	27	mRNA catabolic process <small>Metabolic process</small>	BP	2.52E-04
GO:0010189	3	3	vitamin E biosynthetic process <small>Metabolic process</small>	BP	3.22E-04
GO:0010252	3	5	auxin homeostasis <small>Biological regulation</small>	BP	2.62E-03
GO:0009058	17	372	biosynthetic process <small>Metabolic process</small>	BP	4.22E-03
GO:0016310	7	66	phosphorylation <small>Metabolic process</small>	BP	4.22E-03
GO:0019432	3	6	triglyceride biosynthetic process <small>Metabolic process</small>	BP	4.63E-03
GO:0006090	4	20	pyruvate metabolic process <small>Metabolic process</small>	BP	1.27E-02
GO:0006012	5	43	galactose metabolic process <small>Metabolic process</small>	BP	2.61E-02
GO:0005991	2	3	trehalose metabolic process <small>Metabolic process</small>	BP	3.60E-02
GO:0019310	2	3	inositol catabolic process <small>Metabolic process</small>	BP	3.60E-02
GO:0030014	6	16	CCR4-NOT complex	CC	1.22E-05
GO:0005787	5	14	signal peptidase complex	CC	1.55E-04
GO:0030904	3	8	retromer complex	CC	1.15E-02
GO:0004857	56	164	enzyme inhibitor activity	MF	9.44E-57
GO:0004869	23	67	cysteine-type endopeptidase inhibitor activity	MF	1.77E-22
GO:0004190	36	308	aspartic-type endopeptidase activity	MF	1.39E-18

*Table 4.14 continued on next page*

**Table 4.14 continued from previous page**

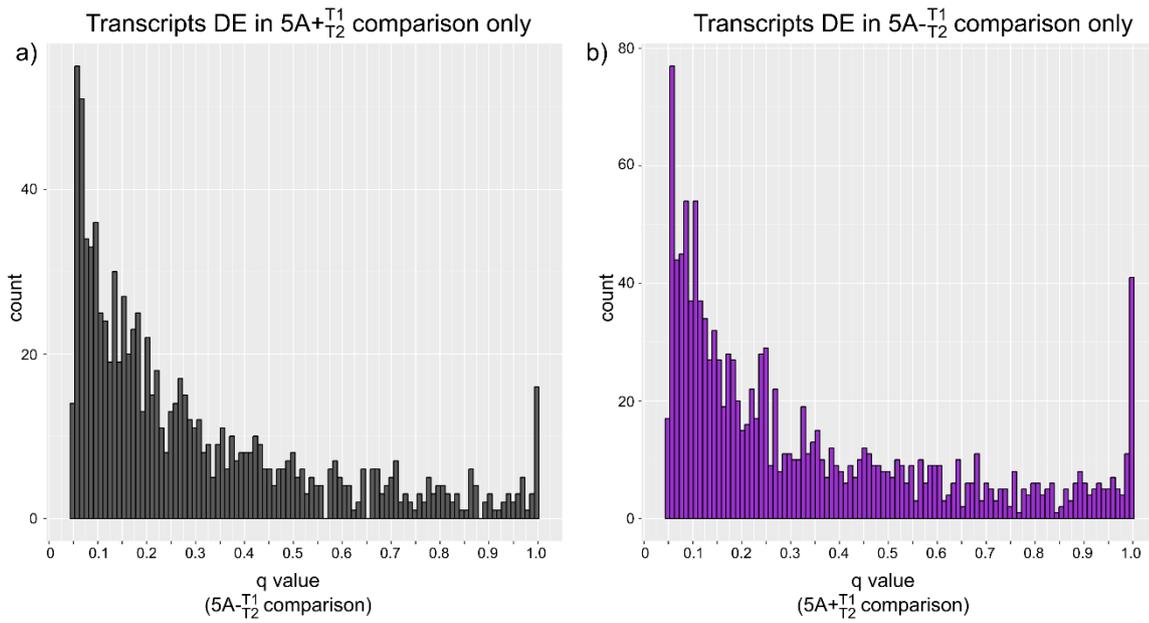
GO term	DE transcripts	Total transcripts	Term ID	Ontology	Adjusted p-value
GO:0004871	15	104	signal transducer activity	MF	1.42E-08
GO:0016813	6	10	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amidines	MF	5.91E-07
GO:0045735	9	40	nutrient reservoir activity	MF	1.34E-06
GO:0008233	12	96	peptidase activity	MF	4.38E-06
GO:0004867	7	24	serine-type endopeptidase inhibitor activity	MF	7.52E-06
GO:0020037	33	767	heme binding	MF	1.29E-05
GO:0030170	18	273	pyridoxal phosphate binding	MF	2.82E-05
GO:0016705	25	523	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	MF	6.06E-05
GO:0051741	3	3	2-methyl-6-phytyl-1,4-benzoquinone methyltransferase activity	MF	3.22E-04
GO:0050242	4	11	pyruvate, phosphate dikinase activity	MF	1.32E-03
GO:0005506	27	738	iron ion binding	MF	2.20E-03
GO:0008483	8	84	transaminase activity	MF	3.04E-03
GO:0008237	7	66	metallopeptidase activity	MF	4.22E-03
GO:0008234	11	222	cysteine-type peptidase activity	MF	3.45E-02
GO:0004555	2	3	alpha,alpha-trehalase activity	MF	3.60E-02
GO:0050113	2	3	inositol oxygenase activity	MF	3.60E-02
GO:0016772	7	99	transferase activity, transferring phosphorus-containing groups	MF	3.87E-02
GO:0003978	4	29	UDP-glucose 4-epimerase activity	MF	4.23E-02
GO:0004553	26	881	hydrolase activity, hydrolyzing O-glycosyl compounds	MF	4.84E-02

DE transcripts =DE transcripts associated with the GO term, Total transcripts = all transcripts associated with the GO term. BP = Biological Process, CC = Cellular Component, MF = Molecular Function. Superscripts are a common parent GO term. Adjusted p-values were calculated using the Benjamini Hochberg procedure.

**Table 4.15: Enriched gene ontology (GO) terms in common downregulated transcripts differentially expressed (DE) in the 5A-  $T_1^I$  and 5A+  $T_2^I$  comparisons (n = 300)**

GO term	DE transcripts	Total transcripts	Term ID	Ontology	Adjusted p-value
GO:0005975	17	1455	carbohydrate metabolic process <sup>Metabolic process</sup>	BP	1.61E-03
GO:0000160	8	292	phosphorelay signal transduction system <sup>Biological regulation</sup>	BP	1.61E-03
GO:0007165	8	404	signal transduction <sup>Biological regulation</sup>	BP	1.18E-02
GO:0006813	3	33	potassium ion transport <sup>Cation transport</sup>	BP	4.03E-02
GO:0004553	15	881	hydrolase activity, hydrolyzing O-glycosyl compounds	MF	1.84E-04
GO:0043531	18	1396	ADP binding	MF	2.97E-04
GO:0005249	3	21	voltage-gated potassium channel activity	MF	1.18E-02

DE transcripts =DE transcripts associated with the GO term, Total transcripts = all transcripts associated with the GO term. BP = Biological Process, MF = Molecular Function. Superscripts are a common parent GO term. Adjusted p-values were calculated using the Benjamini Hochberg procedure.



**Figure 4.17: Distributions of  $q$ -values of uniquely differentially expressed transcripts in the 5A-  $T_2^{T1}$  and 5A+  $T_2^{T1}$  comparisons**

a) Distribution of 5A-  $T_2^{T1}$   $q$  values for transcripts that were differentially expressed (DE) only in the 5A+  $T_2^{T1}$  comparison (and not the 5A-  $T_2^{T1}$  comparison). B) Distribution of 5A-  $T_2^{T1}$   $q$  values for DE transcripts across time in the 5A-  $T_2^{T1}$  comparison only. The fact that both distributions are skewed towards lower  $q$  values shows suggests that many of the DE genes within a single comparison were borderline non-significant in the opposite comparison.

**Table 4.16: Enriched gene ontology (GO) terms in transcripts uniquely differentially expressed (DE) in the 5A- $\frac{T1}{T2}$  comparison (n = 1319)**

GO term	DE transcripts	Total transcripts	Term ID	Ontology	Adjusted p-value
GO:0005975	53	1455	carbohydrate metabolic process <small>Metabolic process</small>	BP	8.71E-08
GO:0042546	10	77	cell wall biogenesis	BP	3.97E-05
GO:0010411	9	64	xyloglucan metabolic process <small>Metabolic process</small>	BP	5.56E-05
GO:0006073	9	71	cellular glucan metabolic process <small>Metabolic process</small>	BP	9.07E-05
GO:0000160	17	292	phosphorelay signal transduction system <small>Biological regulation</small>	BP	9.07E-05
GO:0019538	8	64	protein metabolic process <small>Metabolic process</small>	BP	3.50E-04
GO:0009664	5	22	plant-type cell wall organization	BP	1.41E-03
GO:0009765	8	91	photosynthesis, light harvesting <small>Metabolic process</small>	BP	4.06E-03
GO:0009688	3	6	abscisic acid biosynthetic process <small>Metabolic process</small>	BP	5.06E-03
GO:0009415	3	6	response to water	BP	5.06E-03
GO:0006952	9	162	defense response	BP	3.22E-02
GO:0034551	2	3	mitochondrial respiratory chain complex III assembly	BP	4.37E-02
GO:0009638	2	3	phototropism	BP	4.37E-02
GO:0055114	68	3409	oxidation-reduction process <small>Metabolic process</small>	BP	4.88E-02
GO:0005576	13	132	extracellular region	CC	1.77E-05
GO:0005618	12	130	cell wall	CC	5.56E-05
GO:0048046	10	84	apoplast	CC	5.56E-05
GO:0004553	37	881	hydrolase activity, hydrolyzing O-glycosyl compounds	MF	9.15E-07
GO:0004857	13	164	enzyme inhibitor activity	MF	8.72E-05
GO:0016762	9	72	xyloglucan:xyloglucosyl transferase activity	MF	9.07E-05
GO:0004556	5	23	alpha-amylase activity	MF	1.59E-03
GO:0009540	3	6	zeaxanthin epoxidase [overall] activity	MF	5.06E-03

GO term	DE transcripts	Total transcripts	Term ID	Ontology	Adjusted p-value
GO:0020037	25	767	heme binding	MF	5.06E-03
GO:0004871	8	104	signal transducer activity	MF	7.69E-03
GO:0005506	22	738	iron ion binding	MF	8.71E-08

DE transcripts =DE transcripts associated with the GO term, Total transcripts = all transcripts associated with the GO term. BP = Biological Process, MF = Molecular Function. Superscripts are a common parent GO term. Adjusted p-values were calculated using the Benjamini Hochberg procedure.

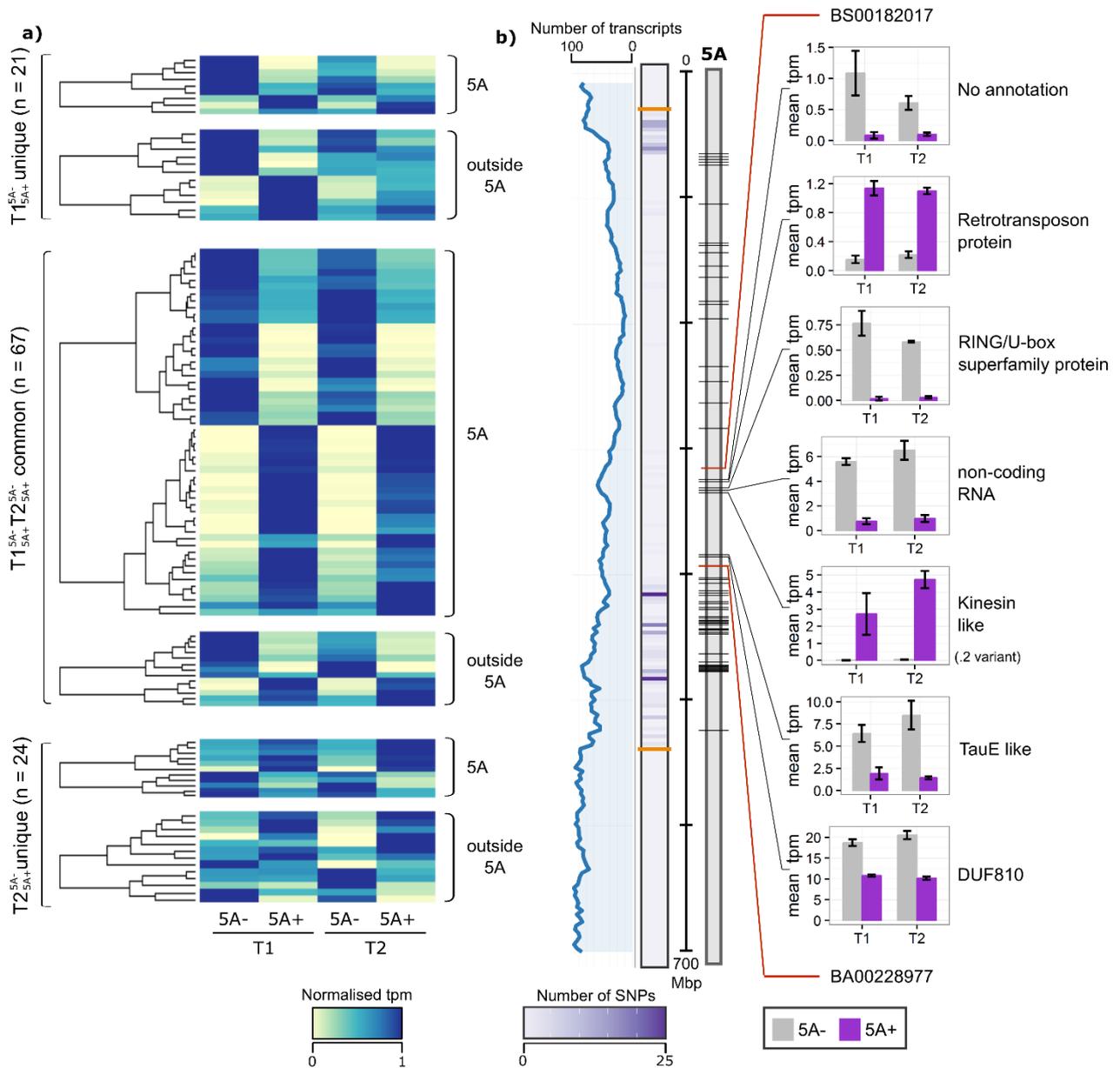
**Table 4.17: Enriched gene ontology (GO) terms in transcripts uniquely differentially expressed (DE) in the 5A+  $\frac{T1}{T2}$  comparison (n = 957)**

GO term	DE transcripts	Total transcripts	Term ID	Ontology	Adjusted p-value
GO:0005978	7	45	glycogen biosynthetic process <sup>Metabolic process</sup>	BP	2.16E-04
GO:0005975	35	1455	carbohydrate metabolic process <sup>Metabolic process</sup>	BP	1.41E-03
GO:0055114	58	3409	oxidation-reduction process <sup>Metabolic process</sup>	BP	3.29E-02
GO:0010155	3	10	regulation of proton transport <sup>Cation transport</sup>	BP	3.37E-02
GO:0004553	30	881	hydrolase activity, hydrolyzing O-glycosyl compounds	MF	2.09E-05
GO:0008878	6	28	glucose-1-phosphate adenylyltransferase activity	MF	2.16E-04
GO:0016491	39	2012	oxidoreductase activity	MF	3.29E-02

DE transcripts =DE transcripts associated with the GO term, Total transcripts = all transcripts associated with the GO term. BP = Biological Process, MF = Molecular Function. Superscripts are a common parent GO term. Adjusted p-values were calculated using the Benjamini Hochberg procedure.

#### 4.3.4 DE transcripts between NILs are concentrated on chromosome 5A

88 and 91 DE transcripts were identified between the NILs in T1  $^{5A-}_{5A+}$  and T2  $^{5A-}_{5A+}$ , respectively, many fewer than identified in  $5A-_{T2}$  or  $5A+_{T2}$ . This was expected as the NILs are genetically very similar and therefore, the difference in developmental stage between the T1 and T2 time points results in greater changes in gene expression. Of these 179 DE transcripts, 67 were common between T1  $^{5A-}_{5A+}$  and T2  $^{5A-}_{5A+}$ , whereas 45 DE transcripts between genotypes were unique and identified only at a single time point (resulting in 112 DE transcripts between NILs at any time point). No GO terms were significantly enriched in these groups. Of the 67 common DE transcripts, 54 (80%) were located on chromosome 5A, whilst in both the T1 and T2 unique groups less than 50% were located on chromosome 5A (Figure 4.18a). Similar numbers of DE transcripts were more highly expressed in either genotype, with no distinct patterns observed between the unique or common groups. Of the 74 DE transcripts located on chromosome 5A all were located within the 491 Mbp introgressed region of the NILs (Figure 4.18b). Higher numbers of DE transcripts were identified in regions of increased SNP density between the 5A NILs.



**Figure 4.18: Differentially expressed transcripts between 5A NILs at T1 and T2**

a) Heatmap of normalised tpm (transcripts per million) of DE (differentially expressed) transcripts between NILs (T1  $5A^-$ / $5A^+$  and T2  $5A^-$ / $5A^+$  comparisons). Transcripts are first grouped based on whether they were differentially expressed at both time points (T1  $5A^-$ / $5A^+$  and T2  $5A^-$ / $5A^+$  common) or at only T1 or T2 (T1  $5A^-$ / $5A^+$  unique and T2  $5A^-$ / $5A^+$  unique, respectively), and then whether they are located on chromosome 5A or not. b) Location of DE transcripts on chromosome 5A (black lines on grey rectangle). Line graph (blue) shows rolling mean of the number of transcripts located in 3 Mbp bins across chromosome 5A, alongside heatmap which shows the number of 90k iSelect SNPs between the 5A- and 5A+ NILs in similar sized bins. Orange lines on the SNP heatmap define the 491 Mbp introgression which differs between the NILs. Bar charts show the mean tpm values at T1 and T2 of DE transcripts located in the fine mapped region (5A- NILs in grey, 5A+ NILs in purple). Only one transcript variant (.2) of the kinesin-like gene is shown. Error bars are standard error of the three biological replicates.

#### 4.3.5 DE transcripts outside of chromosome 5A are enriched in specific transcription factor binding sites

As all the DE transcripts on chromosome 5A were located within the 491 Mbp introgressed region, it is possible that the differential expression was a direct consequence of sequence variation between the NILs e.g. in the promoter regions. However, the 38 DE transcripts located outside of chromosome 5A have the same nucleotide sequence as they are identical by descent (BC<sub>4</sub> NILs confirmed with 90k iSelect SNP array data (Brinton *et al.*, 2017)). It was hypothesised that these transcripts are downstream targets of DE genes, such as transcription factors (TFs), located within the 5A introgression.

To assess this, transcription factor binding sites (TFBS) were identified in the promoter regions of these 38 DE transcripts. The TFBS identified in this group of transcripts were associated with 91 distinct TF families (Table 4.18), five of which were enriched relative to all expressed transcripts (Table 4.18; adjusted P < 0.05). The enriched TFBS families were C2H2, Myb/SANT, AT-Hook, YABBY and MADF/Trihelix.

**Table 4.18: Enriched transcription factor binding sites in the promoters of differentially expressed located outside of 5A**

TF family	Outside 5A DE transcripts (n=38)	All expressed transcripts (n=101,653)	Adjusted p-value
C2H2	36	77987	0.021
Myb/SANT	38	88575	0.021
AT-Hook	38	90203	0.028
YABBY	15	19447	0.034
MADF;Trihelix	13	16632	0.042

Values are the number of transcripts in which binding sites associated with the specified transcription factor (TF) family are present. Adjusted p-values were calculated using the Benjamini Hochberg procedure.

To determine potential candidates for upstream regulators, all annotated TFs located within the introgressed region on chromosome 5A were identified (Borrill *et al.*, 2017). A total of 200 annotated TFs were identified, belonging to 35 TF families (Table 4.19). Of these, four families (across 29 genes) overlapped with enriched TFBS families. Four of the 29 TFs were located within the fine-mapped grain length interval on chromosome 5A, including C2H2, MYB and MYB\_related TFs (Table 4.19).

**Table 4.19: Transcription factors identified in the 5A NIL introgression**

<b>TF family</b>	<b>Introgression</b>	<b>Overall grain length interval</b>
<b>AP2</b>	2	-
<b>B3</b>	14	1
<b>bHLH</b>	18	-
<b>bZIP</b>	13	3
<b>C2H2*</b>	10	2
<b>C3H</b>	4	-
<b>CO-like</b>	1	1
<b>CPP</b>	2	-
<b>DBB</b>	1	-
<b>Dof</b>	3	1
<b>E2F/DP</b>	1	-
<b>EIL</b>	1	-
<b>ERF</b>	30	3
<b>FAR1</b>	20	2
<b>G2-like</b>	6	3
<b>GATA</b>	1	-
<b>GeBP</b>	1	-
<b>GRAS</b>	3	-
<b>HB-other</b>	1	-
<b>HD-ZIP</b>	4	-
<b>HSF</b>	3	-
<b>LBD</b>	3	1
<b>MIKC</b>	2	-
<b>MYB*</b>	10	1
<b>MYB_related*</b>	8	1
<b>NAC</b>	13	2
<b>NF-YC</b>	1	-
<b>SBP</b>	1	-
<b>SRS</b>	1	-
<b>TALE</b>	1	-
<b>TCP</b>	3	-
<b>Trihelix*</b>	1	-
<b>WOX</b>	2	1
<b>WRKY</b>	11	6
<b>ZF-HD</b>	4	-

A \* indicates that transcription factor (TF) binding sites associated with the TF family were significantly enriched in the promoters of transcripts that were differentially expressed between NILs and located outside of chromosome 5A

### 4.3.6 Functional annotation of DE transcripts

Having analysed DE transcripts between NILs based on chromosome location, the 112 DE transcripts were examined based on their functional annotations. Multiple categories of annotations were identified including transcripts associated with ubiquitin-mediated protein degradation, cell cycle, metabolism, transport, transposons and non-coding RNAs (Table 4.20; full annotations in Table 4.22). Few categories were exclusively located on/outside 5A or had exclusively higher expression in the either the 5A- or 5A+ NIL.

**Table 4.20: Categories of DE transcripts between NILs based on predicted function**

Category	number of transcripts	Adjusted p-value	5A/not 5A	NIL with higher expression: 5A-/5A+
non-coding RNA	15	0.141	10/5	6/9
transposon-associated	14	0.008	4/10	5/9
ubiquitin	12**	0.008	10/2	8/4
cell cycle	5	-	5/0	2/3
histone-related	5	-	3/2	3/2
heat shock	5	-	3/2	2/3
protease	4	-	3/1	3/1
transport	4	-	3/1	2/2
metabolism	5	-	5/0	4/1
homeobox	4	0.001	3/1	1/3
cell wall	3	-	2/1	2/1
transcription	3	-	2/1	0/3
non-translating	2	-	0/2	1/1
peroxisome	2	-	0/2	0/2
other*	20	-	14/6	11/9
No annotation	8	-	4/4	5/3

Adjusted p-values displayed are based on an enrichment test of the functional categories relative to all expressed transcripts. - indicates that an enrichment test was not performed as categories were based on bespoke annotations. \* includes transcripts with annotations that could not be grouped by function with other transcripts. \*\* only the 7 transcripts that were annotated as ubiquitin-related in the TGAC annotation were used in the enrichment test (see methods).

The category with the most DE transcripts was non-coding RNA (ncRNA, 15 transcripts), although this was not enriched relative to all expressed transcripts. All ncRNA transcripts were classed as long non-coding RNAs (>200bp, (Guttman & Rinn, 2012)). Four of the ncRNAs overlapped with coding transcripts (two in the antisense direction) and one ncRNA was a putative miRNA precursor (Ta-miR132-3p, 5'-3' mature sequence: TATAAACTTGGTCAAAGTTTG; (Sun *et al.*, 2014)). 13 transcripts (belonging to nine genes) were identified as putative targets of Ta-miR132-3p in the TGAC reference but none of these target transcripts were differentially expressed in this dataset (Table 4.21). The second largest transcript category was transposon-associated (14 transcripts; adjusted P = 0.008), whereas the third largest category was DE transcripts related to ubiquitin and the proteasome (12 transcripts; P = 0.008). DE transcripts annotated as homeobox were also enriched (4 transcripts; adjusted P = 0.001). Interestingly, homeodomain TFBS were identified in the promoters of 27 of the 38 outside 5A DE transcripts although this was not significantly enriched (adjusted P = 0.166).

**Table 4.21: Putative targets of Ta-miR132-3p**

<b>Transcript</b>	<b>TGAC Annotation</b>
TRIAE_CS42_2AL_TGACv1_093400_AA0279320.1	Protein phosphatase 2C containing protein
TRIAE_CS42_2BL_TGACv1_134090_AA0443680.1	Germin-like protein 4-1, Uncharacterized protein
TRIAE_CS42_2BL_TGACv1_134090_AA0443680.2	
TRIAE_CS42_2BS_TGACv1_146052_AA0454150.1	Glycosyltransferase
TRIAE_CS42_6BS_TGACv1_516121_AA1673900.3	Mitochondrial import inner membrane translocase subunit tim22, Uncharacterized protein
TRIAE_CS42_7AS_TGACv1_569800_AA1824270.1	ABC transporter C family member 10, Uncharacterized protein
TRIAE_CS42_7BL_TGACv1_577198_AA1868480.1	Isoleucyl-tRNA synthetase, cytoplasmic, Uncharacterized protein
TRIAE_CS42_7DL_TGACv1_603074_AA1975250.1	Isoleucyl-tRNA synthetase, cytoplasmic, Uncharacterized protein
TRIAE_CS42_7DS_TGACv1_622139_AA2033500.1	Shikimate kinase, Uncharacterized protein
TRIAE_CS42_7DS_TGACv1_622139_AA2033500.2	
TRIAE_CS42_7DS_TGACv1_622139_AA2033500.3	
TRIAE_CS42_7DS_TGACv1_622139_AA2033500.4	
TRIAE_CS42_U_TGACv1_641065_AA2084010.1	Glycosyltransferase

**Table 4.22: Functional annotation of differentially expressed transcripts in the T1 <sup>5A-</sup>/<sub>5A+</sub> and T2 <sup>5A-</sup>/<sub>5A+</sub> comparisons**

Transcript ID	Chr	Position	Time point DE	NIL with higher expression	Category	Annotation	Source
TRIAE_CS42_5AS_TGACv1_393645_AA1274860.1	1A	428,214,494	T2	5A+	non-coding RNA	ncRNA	TGAC
TRIAE_CS42_1BL_TGACv1_032057_AA0124830.1	1B	261,775,404	T2	5A+	non-coding RNA	ncRNA	TGAC
TRIAE_CS42_1BL_TGACv1_030315_AA0086470.1	1B	290,813,728	T2	5A+	non-coding RNA	ncRNA	TGAC
TRIAE_CS42_2AS_TGACv1_112274_AA0334670.1	2A	313,287,504	T2	5A+	non-coding RNA	ncRNA	TGAC
TRIAE_CS42_3DL_TGACv1_250633_AA0871340.1	3D	279,882,343	T1	5A-	non-coding RNA	ncRNA	TGAC
TRIAE_CS42_3DL_TGACv1_250330_AA0866270.1	3D	608,417,264	T1 + T2	5A-	non-coding RNA	ncRNA	TGAC
TRIAE_CS42_5AS_TGACv1_393897_AA1277010.1	5A	85,149,732	T1 + T2	5A-	non-coding RNA	ncRNA	TGAC
TRIAE_CS42_5AS_TGACv1_393783_AA1275990.1	5A	104,011,459	T1 + T2	5A+	non-coding RNA	ncRNA	TGAC
TRIAE_CS42_5AS_TGACv1_394352_AA1279770.1	5A	139,518,885	T1 + T2	5A+	non-coding RNA	ncRNA	TGAC
TRIAE_CS42_5AS_TGACv1_393726_AA1275550.1	5A	162,029,855	T1 + T2	5A+	non-coding RNA	ncRNA; repeat associated	TGAC
TRIAE_CS42_5AL_TGACv1_375857_AA1228100.1*	5A	334,343,515	T1 + T2	5A-	non-coding RNA	ncRNA	TGAC
TRIAE_CS42_5AL_TGACv1_376953_AA1242680.1	5A	427,317,205	T1 + T2	5A-	non-coding RNA	ncRNA	TGAC
TRIAE_CS42_5AL_TGACv1_374498_AA1201570.1	5A	434,793,971	T2	5A+	non-coding RNA	ncRNA	TGAC
TRIAE_CS42_5AL_TGACv1_376877_AA1241920.1	5A	447,314,890	T1	5A-	non-coding RNA	ncRNA	TGAC
TRIAE_CS42_5AL_TGACv1_374413_AA1199590.1	5A	475,323,308	T1 + T2	5A+	non-coding RNA	ncRNA	TGAC
TRIAE_CS42_1AS_TGACv1_019083_AA0060340.1	1A	27,390,992	T1 + T2	5A-	transposon-associated	Repeat associated	TGAC
TRIAE_CS42_1BS_TGACv1_049354_AA0149990.1	1B	185,585,375	T2	5A+	transposon-associated	Retrotransposon-like	Manual
TRIAE_CS42_2AL_TGACv1_095650_AA0313320.1	2A	398,328,221	T1 + T2	5A-	transposon-associated	AT hook motif-containing protein, putative, Putative helicase; retrotransposon-like	TGAC; Manual
TRIAE_CS42_2DS_TGACv1_177196_AA0568430.1	2D	174,418,517	T1 + T2	5A+	transposon-associated	Repeat associated	TGAC
TRIAE_CS42_3AS_TGACv1_211411_AA0689940.1	3A	547,984,779	T1	5A+	transposon-associated	Transposon protein, putative, mutator sub-class	TGAC

*Table 4.22 continued on next page*

*Table 4.22 continued from previous page*

Transcript ID	Chr	Position	Time point DE	NIL with higher expression	Category	Annotation	Source
TRIAE_CS42_4BL_TGACv1_321219_AA1057420.1	4B	527,725,294	T1 + T2	5A-	transposon-associated	Replication factor-A carboxy-terminal domain protein	TGAC
TRIAE_CS42_5AS_TGACv1_393577_AA1274040.1	5A	71,474,627	T1 + T2	5A+	transposon-associated	Retrotransposon protein; Ty3-gypsy subclass	TGAC
TRIAE_CS42_5AS_TGACv1_394531_AA1280840.1	5A	199,145,177	T1	5A-	transposon-associated	zinc ion binding, nucleic acid binding; putative retrotransposon	TGAC; Manual
TRIAE_CS42_5AL_TGACv1_374025_AA1188070.1*	5A	328,818,968	T1	5A+	transposon-associated	Retrotransposon, putative, Ty1-copia subclass	TGAC
TRIAE_CS42_5AL_TGACv1_374249_AA1195190.1	5A	436,603,964	T1 + T2	5A+	transposon-associated	Transposon protein, CACTA, En/Spm sub-class	TGAC
TRIAE_CS42_5DL_TGACv1_433728_AA1420750.1	5D	443,143,427	T1	5A+	transposon-associated	Transposon-related, RICESLEEPER 2-like	Manual
TRIAE_CS42_5DL_TGACv1_435337_AA1449220.1	5D	458,092,027	T2	5A-	transposon-associated	Transposon-related	Manual
TRIAE_CS42_6AL_TGACv1_471516_AA1510240.1	6A	555,225,210	T1 + T2	5A+	transposon-associated	Retrotransposon protein-like	Manual
TRIAE_CS42_7DS_TGACv1_621701_AA2023630.1	7D	28,807,835	T1	5A+	transposon-associated	Transposon protein, Mutator sub-class	TGAC
TRIAE_CS42_1BS_TGACv1_049354_AA0149980.1	1B	185,548,016	T1 + T2	5A-	ubiquitin	UBCc domain; E2 ubiquitin conjugating enzyme	Manual
TRIAE_CS42_5AL_TGACv1_375493_AA1222690.2	5A	265,539,274	T1 + T2	5A-	ubiquitin	BTB/POZ domain-containing protein	TGAC
TRIAE_CS42_5AL_TGACv1_374542_AA1202810.4*	5A	333,439,847	T1 + T2	5A-	ubiquitin	RING/U-box superfamily protein; putative E3 ligase	TGAC; Manual
TRIAE_CS42_5AL_TGACv1_375361_AA1220430.2	5A	413,416,970	T1 + T2	5A-	ubiquitin	RAD23, ubiquitin receptor, proteasome associated	Manual
TRIAE_CS42_5AL_TGACv1_374374_AA1198360.1	5A	421,856,090	T1 + T2	5A-	ubiquitin	F-box protein-like	Manual
TRIAE_CS42_5AL_TGACv1_374097_AA1190230.1	5A	439,551,515	T1 + T2	5A+	ubiquitin	Ubiquitin, Polyubiquitin 14; NEDD8-like protein RUB1	TGAC; Manual
TRIAE_CS42_5AL_TGACv1_374675_AA1206250.1	5A	439,853,084	T1	5A-	ubiquitin	RING/U-box superfamily protein, Zinc finger protein-like protein	TGAC
TRIAE_CS42_5AL_TGACv1_378415_AA1253190.1	5A	470,075,745	T2	5A+	ubiquitin	F-box protein	TGAC

*Table 4.22 continued on next page*

Table 4.22 continued from previous page

Transcript ID	Chr	Position	Time point DE	NIL with higher expression	Category	Annotation	Source
TRIAE_CS42_5AL_TGACv1_377520_AA1247660.1	5A	475,464,797	T1 + T2	5A+	ubiquitin	eIF3 n terminal; PCI/PINT associated module	Manual
TRIAE_CS42_5DL_TGACv1_434064_AA1428250.1	5D	207,168,354	T2	5A+	ubiquitin	E3 ubiquitin-protein ligase SDIR1	TGAC
TRIAE_CS42_5AS_TGACv1_402215_AA1284070.1	NA	NA	T1 + T2	5A-	ubiquitin	Ubiquitin, Polyubiquitin 4	TGAC
TRIAE_CS42_5AS_TGACv1_393696_AA1275280.2	NA	NA	T1 + T2	5A-	ubiquitin	Polyubiquitin, Ubiquitin	TGAC
TRIAE_CS42_5AS_TGACv1_393572_AA1273920.1	5A	73,805,941	T1 + T2	5A-	cell cycle	Kinesin-like protein	TGAC
TRIAE_CS42_5AL_TGACv1_373986_AA1186560.2*	5A	336,456,148	T1 + T2	5A+	cell cycle	Kinesin-like protein	TGAC
TRIAE_CS42_5AL_TGACv1_373986_AA1186560.3*	5A	336,456,148	T1 + T2	5A+	cell cycle	Kinesin-like protein	TGAC
TRIAE_CS42_5AL_TGACv1_374322_AA1196910.1	5A	408,190,618	T2	5A-	cell cycle	IMP dehydrogenase/GMP reductase; HAUS augmin-like complex subunit 5	TGAC; Manual
TRIAE_CS42_5AL_TGACv1_376411_AA1236560.1	5A	473,423,951	T1 + T2	5A+	cell cycle	SHAGGY-like kinase	Manual
TRIAE_CS42_3AL_TGACv1_195567_AA0651320.1	3A	746,292,914	T2	5A+	histone-related	Histone H2A	Manual
TRIAE_CS42_5AS_TGACv1_394776_AA1281770.1	5A	248,528,260	T1 + T2	5A-	histone-related	Histone deacetylase 14 isoform	Manual
TRIAE_CS42_5AS_TGACv1_394776_AA1281770.3	5A	248,528,260	T1 + T2	5A-	histone-related	Histone deacetylase 14 isoform	Manual
TRIAE_CS42_5AL_TGACv1_374195_AA1193180.2	5A	477,295,158	T2	5A+	histone-related	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein; Lysine-specific demethylase JMJ30	TGAC; Manual
TRIAE_CS42_U_TGACv1_643349_AA2131010.1	Un	103,526,863	T2	5A-	histone-related	Histone superfamily protein; Histone H4 superfamily	TGAC; Manual
TRIAE_CS42_5AL_TGACv1_376851_AA1241550.4	5A	414,772,190	T2	5A+	heatshock	TPR repeat-containing thioredoxin TDX; heatshock related	TGAC; Manual
TRIAE_CS42_5AL_TGACv1_375394_AA1220930.4	5A	474,541,233	T1 + T2	5A+	heatshock	HSP90 superfamily	Manual
TRIAE_CS42_5AL_TGACv1_375394_AA1220930.5	5A	474,541,233	T2	5A+	heatshock	HSP90 superfamily	Manual
TRIAE_CS42_4DL_TGACv1_343485_AA1135140.1	Un	100,323,297	T2	5A-	heatshock	HSP70, DnaK	Manual
TRIAE_CS42_7BL_TGACv1_578302_AA1892720.1	NA	NA	T2	5A-	heatshock	Retrotransposon putative; HEAT-STRESS-ASSOCIATED 32	TGAC; Manual
TRIAE_CS42_2DS_TGACv1_177488_AA0578600.1	2D	13,913,395	T1	5A-	protease	Protease domain; ankryin repeats	Manual

Table 4.22 continued on next page

**Table 4.22 continued from previous page**

Transcript ID	Chr	Position	Time point DE	NIL with higher expression	Category	Annotation	Source
TRIAE_CS42_5AL_TGACv1_374321_AA1196890.1	5A	422,062,336	T1 + T2	5A-	protease	Aspartyl aminopeptidase; Zinc peptidase-like superfamily	TGAC; Manual
TRIAE_CS42_5AL_TGACv1_375845_AA1227990.1	5A	444,535,471	T1 + T2	5A-	protease	Abi superfamily, CAAX protease	Manual
TRIAE_CS42_5AL_TGACv1_374078_AA1189690.1	5A	476,435,013	T1 + T2	5A+	protease	Serine carboxypeptidase-like protein 9	TGAC
TRIAE_CS42_5AS_TGACv1_393493_AA1273190.4	5A	77,084,934	T1 + T2	5A+	transport	Calcium transporting ATPase	TGAC
TRIAE_CS42_5AL_TGACv1_375949_AA1229270.2*	5A	387,399,431	T2	5A-	transport	TauE superfamily	Manual
TRIAE_CS42_5AL_TGACv1_374155_AA1191930.1	5A	448,605,465	T2	5A-	transport	Potassium transporter	TGAC
TRIAE_CS42_7AL_TGACv1_556075_AA1754700.1	7A	567,001,946	T1	5A+	transport	ABC transporter G family	Manual
TRIAE_CS42_5AS_TGACv1_395074_AA1282530.1	5A	143,730,436	T1 + T2	5A+	metabolism	Glyoxylate reductase	TGAC
TRIAE_CS42_5AL_TGACv1_376402_AA1236390.1	5A	404,531,058	T1 + T2	5A-	metabolism	Acetate--CoA ligase	Manual
TRIAE_CS42_5AL_TGACv1_374560_AA1203240.1	5A	417,045,347	T1	5A-	metabolism	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex	TGAC
TRIAE_CS42_5AL_TGACv1_374231_AA1194360.2	5A	438,141,289	T1 + T2	5A-	metabolism	Monogalactosyldiacylglycerol synthase	Manual
TRIAE_CS42_5AS_TGACv1_393119_AA1268700.1	5D	180,466,588	T1 + T2	5A-	metabolism	quinolinate synthase	TGAC
TRIAE_CS42_5AS_TGACv1_392838_AA1265240.1	5A	181,399,775	T1 + T2	5A+	homeobox	Homeobox protein knotted-1-like 2	TGAC
TRIAE_CS42_5AS_TGACv1_392838_AA1265240.2	5A	181,399,775	T1 + T2	5A+	homeobox	Homeobox protein knotted-1-like 2	TGAC
TRIAE_CS42_5AL_TGACv1_374741_AA1207970.1	5A	463,451,614	T2	5A-	homeobox	homeobox-leucine zipper protein	TGAC
TRIAE_CS42_7BL_TGACv1_577371_AA1873630.1	7B	692,600,853	T1 + T2	5A+	homeobox	Homeobox protein knotted-1-like 2	TGAC
TRIAE_CS42_5AL_TGACv1_375085_AA1215790.1	5A	475,524,596	T1 + T2	5A+	cell wall	Fascilin-like arabinogalactan protein	Manual
TRIAE_CS42_5AL_TGACv1_374319_AA1196780.1	5A	476,667,345	T1 + T2	5A-	cell wall	Pectin lyase-like superfamily protein; Polygalacturonase-like	TGAC; Manual
TRIAE_CS42_7DS_TGACv1_622195_AA2034920.1	7D	106,414,593	T1	5A-	cell wall	Callose synthase	Manual
TRIAE_CS42_5AL_TGACv1_378188_AA1251790.1	2B	667,276,330	T1 + T2	5A+	transcription	Far1-related sequence 5-like protein	TGAC
TRIAE_CS42_4DS_TGACv1_361025_AA1159110.4	4D	79,211,678	T2	5A+	transcription	LIM-domain binding protein, SEUSS orthologue	Manual

**Table 4.22 continued on next page**

*Table 4.22 continued from previous page*

Transcript ID	Chr	Position	Time point DE	NIL with higher expression	Category	Annotation	Source
TRIAE_CS42_5AS_TGACv1_392558_AA1260860.1	5A	185,647,816	T1 + T2	5A+	transcription	SSXT protein; GRF1-interacting-factor	TGAC; Manual
TRIAE_CS42_2DL_TGACv1_157942_AA0502830.1	2D	614,383,115	T1	5A+	non-translating	Non-translating	TGAC
TRIAE_CS42_7AS_TGACv1_571015_AA1843630.1	7A	32,210,288	T1 + T2	5A-	non-translating	Non-translating	TGAC
TRIAE_CS42_5AL_TGACv1_374446_AA1200410.1	5A	473,092,584	T1 + T2	5A+	peroxisome	Putative peroxisomal targeting signal 1 receptor	TGAC
TRIAE_CS42_5AL_TGACv1_374446_AA1200420.1	5A	473,195,032	T1	5A+	peroxisome	Putative peroxisomal targeting signal 1 receptor	TGAC
TRIAE_CS42_2AL_TGACv1_095650_AA0313330.1	2A	398,347,440	T1	5A-	other	homolog of yeast autophagy 18 (ATG18) G	TGAC
TRIAE_CS42_2AL_TGACv1_095956_AA0316040.1	2A	398,971,137	T1	5A-	other	DEA(D/H)-box RNA helicase family protein	TGAC
TRIAE_CS42_2AL_TGACv1_095956_AA0316050.1	2A	398,979,949	T1 + T2	5A-	other	DNA binding protein-like	TGAC
TRIAE_CS42_2BS_TGACv1_148693_AA0494610.1	2B	58,987,101	T1	5A-	other	transducin family protein, WD-40 repeat family	Manual
TRIAE_CS42_3DL_TGACv1_254173_AA0896070.1	3D	276,574,630	T2	5A+	other	IQM1; Calmodulin binding	Manual
TRIAE_CS42_5AS_TGACv1_393235_AA1270150.1	5A	138,079,992	T1 + T2	5A+	other	Hydroxyproline-rich glycoprotein family protein	TGAC
TRIAE_CS42_5AS_TGACv1_393580_AA1274130.1	5A	155,209,029	T1 + T2	5A+	other	Endoplasmic reticulum, stress-associated Ramp4	TGAC
TRIAE_CS42_5AL_TGACv1_374163_AA1192120.1	5A	284,690,697	T2	5A-	other	Rho GTPase-activating protein gacA	TGAC
TRIAE_CS42_5AL_TGACv1_374727_AA1207530.1*	5A	388,637,659	T1 + T2	5A-	other	DUF810 family protein	TGAC
TRIAE_CS42_5AL_TGACv1_376076_AA1231790.1	5A	403,283,773	T1	5A+	other	tyrosine--tRNA ligase 1	Manual
TRIAE_CS42_5AL_TGACv1_376953_AA1242690.1	5A	427,319,739	T1 + T2	5A-	other	Hypersensitive induced response protein 3	TGAC
TRIAE_CS42_5AL_TGACv1_376953_AA1242690.2	5A	427,320,059	T1 + T2	5A-	other	Hypersensitive induced response protein 3	TGAC
TRIAE_CS42_5AL_TGACv1_374097_AA1190260.1	5A	439,569,122	T1 + T2	5A+	other	DNA-3-methyladenine glycosylase	TGAC
TRIAE_CS42_5AL_TGACv1_375721_AA1226170.1	5A	439,767,852	T1 + T2	5A+	other	Bet1-like SNARE 1-1	TGAC
TRIAE_CS42_5AL_TGACv1_378334_AA1252720.1	5A	444,849,607	T1 + T2	5A-	other	Vacuolar processing enzyme 4	TGAC
TRIAE_CS42_5AL_TGACv1_376411_AA1236550.1	5A	473,427,021	T1 + T2	5A+	other	ribonucleoside--diphosphate reductase large subunit partial match	Manual
TRIAE_CS42_5AL_TGACv1_374080_AA1189800.1	5A	473,625,152	T1 + T2	5A+	other	Vacuolar protein sorting-associated protein; Sec3 superfamily	TGAC; Manual

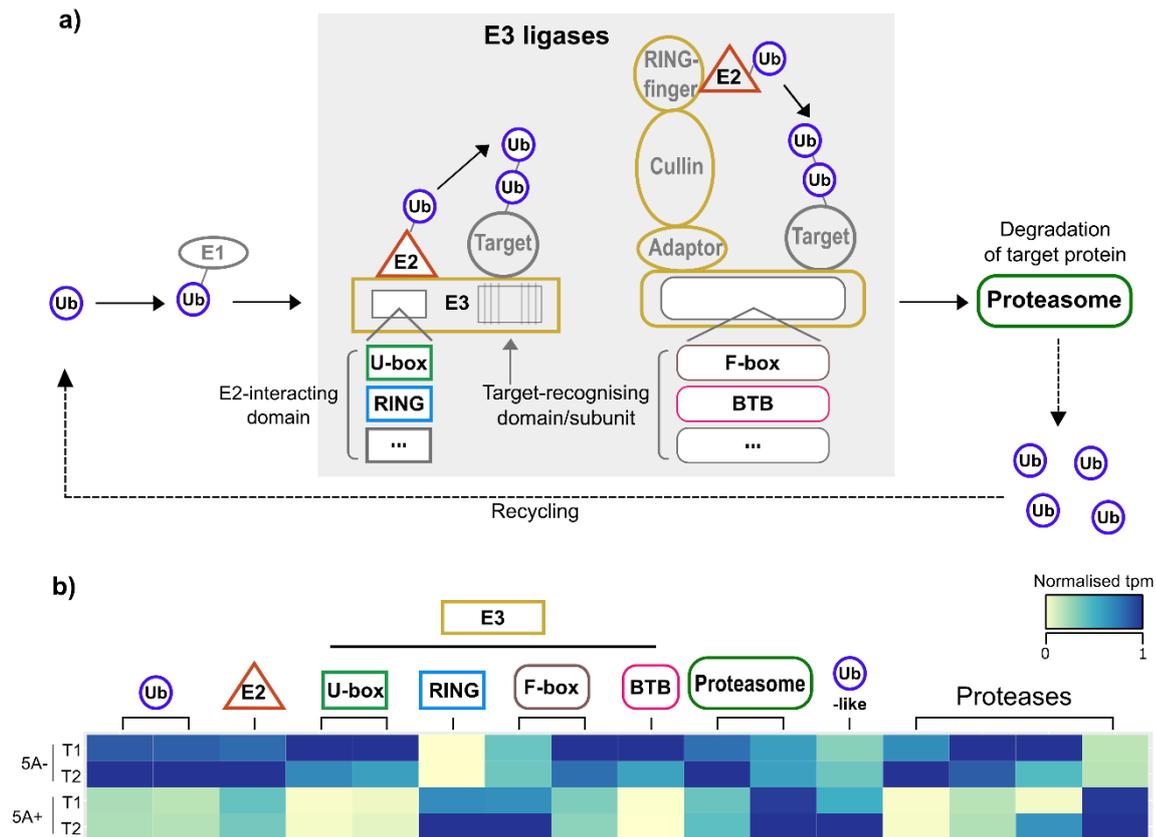
*Table 4.22 continued on next page*

Table 4.22 continued from previous page

Transcript ID	Chr	Position	Time point DE	NIL with higher expression	Category	Annotation	Source
TRIAE_CS42_5AL_TGACv1_374080_AA1189810.1	5A	473,637,539	T1 + T2	5A+	other	Glycosyltransferase protein 2-like	TGAC
TRIAE_CS42_5AL_TGACv1_377986_AA1250630.1	5A	524,318,116	T1 + T2	5A-	other	Generative cell specific-1; Hapless 2	TGAC; Manual
TRIAE_CS42_6BL_TGACv1_503194_AA1627460.1	6B	623,718,695	T1 + T2	5A-	other	Allene oxide cyclase 4	TGAC
TRIAE_CS42_4BS_TGACv1_327817_AA1075010.1	4B	95,073,178	T2	5A-	No annotation	NA	NA
TRIAE_CS42_5AS_TGACv1_395084_AA1282570.1	5A	70,454,003	T1 + T2	5A-	No annotation	NA	NA
TRIAE_CS42_5AS_TGACv1_393473_AA1272910.1	5A	237,953,207	T1 + T2	5A-	No annotation	NA	NA
TRIAE_CS42_5AL_TGACv1_374657_AA1205780.1*	5A	328,545,606	T1	5A-	No annotation	NA	NA
TRIAE_CS42_5AL_TGACv1_375845_AA1227980.1	5A	444,539,760	T1	5A-	No annotation	NA	NA
TRIAE_CS42_5AL_TGACv1_376877_AA1241930.1	5A	447,310,958	T1 + T2	5A-	No annotation	NA	NA
TRIAE_CS42_5BS_TGACv1_427448_AA1393420.1	5B	52,914,927	T1 + T2	5A+	No annotation	NA	NA
TRIAE_CS42_U_TGACv1_641674_AA2101090.1	6B	719,335,147	T1	5A+	No annotation	NA	NA
TRIAE_CS42_7BS_TGACv1_592547_AA1940160.1	7B	198,610,669	T1 + T2	5A+	No annotation	NA	NA

\* indicates that the transcript is located in the fine-mapped interval for grain length. Chromosome (Chr) and position of the transcripts are based on an *in silico* mapping of TGACv1 gene models to IWGSC RefSeq v1.0 (NA indicates that no position could be assigned using this method). NA in the annotation columns indicates that no annotation could be obtained.

The DE transcripts related to ubiquitin were of particular interest as ubiquitin-mediated protein turnover has previously been associated with the control of seed/grain size in wheat (Simmonds *et al.*, 2016) and other species including rice and arabidopsis (Disch *et al.*, 2006; Song, X-J *et al.*, 2007; Xia, Tian *et al.*, 2013). The pathway acts through the sequential action of a cascade of enzymes (see Figure 4.19a legend) to add multiple copies of the protein ubiquitin (ub) to a substrate protein that is then targeted for degradation by the proteasome. Differentially expressed transcripts were identified at almost all steps of this pathway (excluding E1): two ubiquitin proteins and one ubiquitin-like protein, one E2 conjugase, six potential E3 ligase components and two putative components of the proteasome (Figure 4.19). In addition to these, we also identified four DE transcripts annotated as proteases (Figure 4.19), which are known substrates regulated by this pathway (Huang *et al.*; Du *et al.*, 2014; Dong *et al.*, 2017) and that influence organ size through the regulation of cell proliferation. Most of the components of the ubiquitin pathway that were differentially expressed were more highly expressed in the 5A- NIL (11/16, including proteases) (Figure 4.19b).



**Figure 4.19: Differential regulation of the ubiquitin pathway in 5A NILs**

a) Differentially expressed (DE) transcripts with functional annotations related to ubiquitin-mediated protein turnover were enriched relative to the whole genome (a). This pathway acts to add multiple copies of the protein Ubiquitin (Ub) to a substrate protein through the sequential action of a cascade of three enzymes: E1 (Ub-activating enzymes), E2 (Ub-conjugating enzymes) and E3 (Ub ligases). The tagged substrate is then targeted for degradation by the 26S proteasome and the Ub proteins are recycled. The E3 ligases are the most diverse of the three enzymes and both single subunit proteins and multi-subunit complexes exist. A subset of these classes is shown in the grey box in (a), selected based on the annotations of DE transcripts. Single subunit E3 ligases have an E2-interacting domain (e.g. U-box, RING, etc. (...)) and a substrate-recognising domain. Multi-subunit complexes also have E2-interacting complexes and substrate-recognising subunits (e.g. F-box, BTB, etc. (...)). In the context of organ size control, some proteases have been identified as downstream targets of this pathway (e.g. DA1, UBP15 (Du *et al.*, 2014; Dong *et al.*, 2017)). b) Heatmap of normalised tpm of DE transcripts associated with ubiquitin, the proteasome and proteases.

## 5 Discussion

The overall aim of this thesis was to understand the mechanisms that control grain length and width in hexaploid wheat through the characterisation of two distinct grain weight QTL located on chromosomes 5A and 6A. Specifically, this PhD combined phenotypic characterisation, genetic mapping and transcriptomics to answer the following questions:

- Do the 5A and 6A QTL increase grain weight via the same or different mechanisms?
- What are the genes/pathways underlying the 5A and 6A QTL?
- Is *TaGW2\_A* the gene underlying the 6A QTL?

### 5.1 Mechanisms and genes underlying the 6A and 5A QTL

As determined in Chapter 4.1, the 6A and 5A QTL act to increase grain weight through different mechanisms, consistent with previous reports that these grain size parameters are under independent genetic control (Gegas *et al.*, 2010). The 5A QTL acts primarily to increase grain length during early grain development, but post-fertilisation, through increased pericarp cell size. The 5A QTL also has a pleiotropic effect on grain width during late grain development, which is smaller than the effect on length. This late-season width effect is potentially more sensitive to environmental variation and determines the magnitude of the final grain weight increase. On the other hand, the 6A QTL acts during very early grain development, perhaps before fertilisation (i.e. carpel development), and specifically increases grain width, with no differences observed in final grain length. Although no cell size/number data was obtained for the 6A QTL, we hypothesised that this is likely to be an effect on cell number due to the timing of initial grain size differences, although this has not yet been tested experimentally.

#### 5.1.1 Genes and pathways underlying the 6A QTL

Speculating on the identity of candidates for the causal genes underlying the 6A QTL remains challenging, as the high confidence fine-mapped interval contains > 2,000 genes and even the tentative narrower interval contains > 400 genes. The carpel/grain developmental time-courses showed that the QTL acts during the very early stages of carpel/grain development, highlighting the importance of this early stage in determining final grain weight (discussed further in section 5.1.4). However, it was not possible to define the exact time during development when these differences are first established and this meant that we could not select time points for RNA-Seq studies in the same way as for the 5A QTL. This meant that we have limited information about genes that might be regulated differently between NILs and therefore, related to the final grain weight phenotype.

Based on previous studies that show that predominantly cell proliferation is occurring at the very early stages of grain development (Drea *et al.*, 2005; Radchuk *et al.*, 2011), we hypothesise that this QTL acts to influence cell number. Studies in species including rice and *Arabidopsis* have

identified genes that influence seed/grain weight through the control of cell proliferation. These genes have a diverse range of functions, including transcription factors, G-protein signalling, phytohormone signalling, cell cycle components, cytochromes and proteases (Mizukami & Fischer, 2000; Schruff *et al.*, 2006; Adamski *et al.*, 2009; Huang *et al.*, 2009; Qi *et al.*, 2012; Xia. *et al.*, 2013; Du *et al.*, 2014). For example, in rice a *SQUAMOSA PROMOTER-BINDING LIKE (SPL)* transcription factor, *OsSPL16*, was found to influence grain size through positive regulation of cell proliferation by modulating the expression of certain components of the cell cycle machinery (Wang *et al.*, 2012). Negative regulators of cell proliferation have also been identified as important for the control of grain size, such as the E3 ubiquitin ligase, *GW2* (Song, X-J *et al.*, 2007), the A genome wheat orthologue of which (*TaGW2-A*) was considered as a potential candidate gene for the 6A QTL.

#### **5.1.1.1 Is *TaGW2-A* the causal gene underlying the 6A QTL?**

One of the aims of this thesis was to determine whether *TaGW2-A* is the causal gene underlying the 6A QTL as it mapped within the original 6A grain weight QTL interval (Simmonds *et al.*, 2014). This hypothesis was assessed phenotypically and genetically in Chapters 2 and 3. Taking together all the evidence, we concluded that *TaGW2-A* is unlikely to be the gene underlying the 6A QTL and that they act through different mechanisms (Chapter 3). The main lines of evidence leading to this conclusion were that:

- *TaGW2-A* maps outside of the 0.28 cM fine-mapped 6A grain width interval
- *TaGW2-A* has no coding region polymorphisms in the parental varieties, Spark and Rialto (Simmonds *et al.*, 2014)
- *TaGW2-A* NILs have significantly different final grain width and length (two experiments) whilst differences in final grain length were not observed in 6A NILs (three experiments)
- *TaGW2-A* NILs have differences in carpel/grain length throughout carpel/grain development but differences in carpel/grain length were rarely observed between 6A NILs
- *TaGW2-A* NILs have clear differences in carpel width and length at heading, whereas no significant differences in carpel size or weight were observed between 6A NILs at heading

However, this conclusion is subject to confirmation of the 0.28 cM fine-mapped interval using additional phenotypic data from the larger 6A RIL population from the 2017 field trials. This confirmation is particularly critical because the high confidence 4.6 cM fine-mapped interval does include *TaGW2-A* and it cannot be excluded as the causal gene based on phenotypic differences alone, as discussed in Chapter 2. One of the main arguments for *TaGW2-A* as the causal gene underlying the 6A QTL, aside from its effect on grain weight in general, is the presence of an A/G promoter SNP at the -593 bp position between the parental varieties, Spark and Rialto (Simmonds

*et al.*, 2014). This SNP has previously been associated with grain width and TGW in Chinese germplasm; however, studies have generated contradictory results (Su *et al.*, 2011; Zhang, X *et al.*, 2013). These studies used association analysis in similar panels of Chinese germplasm to identify a putative effect of the *TaGW2-A* promoter SNP on TGW. Su *et al.* found that the A allele at the -593 position was associated with increased grain weight, whilst Zhang *et al.* found that the G allele was associated with increased grain weight. This could be explained by the extended linkage disequilibrium that exists in the *TaGW2-A* region given its proximal position on chromosome 6A as observed in the 6A RIL populations in Chapter 3. This would determine extended haplotypes that could encompass hundreds of genes in addition to *TaGW2-A*, any of which could underlie the observed variation in grain weight.

Direct manipulation of *TaGW2-A* through induced (Simmonds *et al.*, 2016; Chapter 2) and natural missense mutations (Yang *et al.*, 2012) has clearly established a role of *TaGW2-A* on grain size in wheat. Molecular studies have also shown that the ubiquitination activity of rice *GW2* is conserved in *TaGW2-A* (Bednarek *et al.*, 2012). However, it is still an open question as to whether the association effects in the two contradictory studies are due to allelic differences in *TaGW2-A* itself or in a linked gene across the haplotype block. Indeed, the same logic can be applied to the 4.6 cM fine-mapped grain width interval on chromosome 6A in this thesis: in a region that contains > 2000 genes it is not possible to say whether *TaGW2-A* is the causal gene or not regardless of the presence of the promoter SNP. This is reminiscent of the cloning of the pre-harvest sprouting QTL (*Phs-A1*) where *PM19-A1* was incorrectly identified as the causal gene due to the presence of a promoter deletion and a demonstrated effect of the gene on grain dormancy through direct manipulation (Barrero *et al.*, 2015). However, it was subsequently shown that *PM19-A1* was in fact linked to the true causal gene, *TaMKK3*, in the germplasm studied resulting in a spurious association with the QTL phenotype and *PM19* promoter deletion (Shorinola *et al.*, 2017b). If data from the 2017 field trials confirm that *TaGW2-A* maps separately from the 6A QTL, this will open up some interesting new avenues for potential further studies. An important question to ask will be precisely which aspects of grain development the 6A gene and *TaGW2-A* affect. Given that the two pairs of NILs seem to have similar phenotypic differences, it is possible that the two genes may influence the same processes. Characterisation of the NILs on a cellular level during carpel/grain development will provide insights into this and these studies are currently underway. Additionally, it would be interesting to understand how the two genes interact and whether beneficial alleles of both genes can be combined to give additive or synergistic increases in TGW. If so, then this could have implications in breeding as well as providing mechanistic insight. Currently, breeders are selecting for a large physical region on chromosome 6A, encompassing both the 6A grain weight effect and *TaGW2-A*. The separation of these two loci could allow for novel combinations of alleles to be deployed, although this would still be limited by the low rates of recombination observed across this region on chromosome 6A.

### **5.1.1.2 Future steps to identify genes and pathways underlying the 6A QTL**

The fact that genes with so many diverse functions can influence cell proliferation makes it premature to speculate on the identity of a candidate gene from 488 genes solely based on predicted function. In addition to the additional phenotypic data for the larger RIL population, the marker density across the interval will also be increased. No gene based SNP calling has yet been conducted on the 6A NILs/parental varieties and this will be performed using exome capture data. This will be useful for identifying additional markers and also will identify genes with potentially deleterious mutations, which could assist in prioritising candidate genes for further study. A promoter capture array will also be employed to access variation in the 2 kb upstream of all genes in this region.

### **5.1.2 Genes and pathways underlying the 5A QTL**

More insight was gained into the potential genes and mechanisms underlying the 5A QTL during this PhD. We found that the 5A QTL acts primarily to increase grain length and this was associated with increased cell length in the pericarp. The first differences in grain length were observed at around 12 dpa (8 – 15 dpa across years, ~ 6.5 mm), which is consistent with a role of the QTL in cell expansion as cell proliferation in the pericarp decreases shortly after fertilisation (Drea *et al.*, 2005; Radchuk *et al.*, 2011). Similar to the 6A and *TaGW2-A* data, the results from the 5A NILs emphasise the importance of the early stage of grain development in determining the final grain size.

Overall, these results suggest that the gene(s) underlying the 5A QTL either directly or indirectly regulate cell expansion in the pericarp, a mechanism that is known to be a key determinant of grain/seed size in several species. Some genes, such as expansins and XTH (xyloglucan endotransglucosylase/hydrolases), affect cell expansion directly by physically modifying or “loosening” the cell wall (reviewed in Cosgrove, 2005), and the expression of these enzymes has been associated with pericarp cell expansion in wheat and barley (Lizana *et al.*, 2010; Radchuk *et al.*, 2011; Munoz & Calderini, 2015). The properties of the cell wall can also be modified, for example accumulation of certain tannins in the cell wall can change its competence for elongation. The Arabidopsis WRKY transcription factor, *TTG2*, regulates some steps of the tannin biosynthesis pathway. *ttg2* mutants have smaller seeds due to smaller cells in the seed coat, likely due to a reduced capacity of the cell wall for elongation (Johnson *et al.*, 2002; Garcia *et al.*, 2005). In rice, *SRS3*, a kinesin 13 protein, was shown to regulate grain length through cell size likely through the regulation of microtubule dynamics (Kitagawa *et al.*, 2010). Other genes regulate pericarp/seed coat cell size through more indirect mechanisms, for example through the regulation of sugar metabolism and subsequent accumulation in the vacuole (Ohto *et al.*, 2005; Ohto *et al.*, 2009) and endoreduplication (Chevalier *et al.*, 2014). Many of the genes identified within the fine-mapped region(s) for grain length have functional annotations similar to these genes, but as with the 6A QTL, the intervals remain too large to speculate on a causal gene based on function alone.

As discussed briefly in Chapter 4, seed/grain development requires the coordination of processes across the pericarp/seed coat, endosperm and embryo. It has been proposed in multiple species, that the size of the maternal pericarp/seed coat exerts its influence on final grain size by physically restricting endosperm growth (Calderini *et al.*, 1999; Adamski *et al.*, 2009; Hasan *et al.*, 2011). Grain size in rice is limited by the size of the spikelet hull in an analogous way (Song *et al.*, 2005). In wheat, both pericarp width (Gegas *et al.*, 2010; Simmonds *et al.*, 2016) and length (Lizana *et al.*, 2010; Hasan *et al.*, 2011) have been proposed as key determinants of final grain. The results from this thesis support this idea and we hypothesise that the 5A cell expansion effect increases the physical space available for endosperm growth during the middle and late stages of grain development. This increased physical capacity could then lead to the increase in grain width that is only established at the later stages of grain development, consistent with the time during grain development associated with grain filling and endosperm growth (Olsen, 2001; Shewry *et al.*, 2012). It is not clear whether the increased capacity for grain filling is utilised by increasing the rate or duration of grain filling in 5A+ NILs. A more detailed time course of grain development with more frequent time points and continuing until the final grain weight had been achieved would be required to determine this. Additionally, time courses would ideally be measured in degree days rather than absolute days to properly calculate grain filling rates, especially in order to compare across years whilst accounting for environmental variation in temperature. Unfortunately, uninterrupted weather data from the weather station at Church farm across the entire time course was not available in any year.

It has been shown that the cross-talk between the endosperm and pericarp/seed coat extends beyond purely mechanical constraints and increased cell size in the seed coat/pericarp can be achieved as an indirect effect of increased endosperm growth. For example, the *HAIKU (IKU)* genes act to promote endosperm growth in Arabidopsis. *iku* mutants have smaller seeds due to reduced endosperm growth and indirectly reduce cell elongation in the integument/seed coat (Garcia *et al.*, 2003). The indirect effect on cell size in the seed coat (a maternal tissue) was determined by demonstrating that *iku* double mutants pollinated with WT pollen had WT-like seeds, therefore showing that the *iku* mutations do not have a direct effect on the maternal integument. Already this could suggest a level of communication between the two tissues (Garcia *et al.*, 2003). The *IKU* genes interact on a genetic basis with *TTG2* (described above) and *iku ttg2* double mutants have seeds even smaller than *iku* mutants, due to the *ttg2* mutation compromising the elongation capacity of integument cell walls hence restricting endosperm growth further. This is in accordance with the size of the pericarp imposing a physical constraint on endosperm. However, combining the *iku* mutations with lines that have reduced cell proliferation in the integuments (due to overexpression of *KIP RELATED PROTEIN2*) did not show an additive effect on seed size and instead the reduction in cell number in the integument was compensated for by increased cell elongation (Garcia *et al.*, 2005). This suggests that in some cases the size of the pericarp/seed coat can be adjusted to accommodate the growth of the endosperm, providing additional evidence

there must be communication/signalling between the tissues. An example of communication from seed coat to endosperm/embryo in cereals can be seen in the control of seed dormancy. A group of three genes, known as the R genes, are responsible for determining grain colour specifically by controlling pigmentation in the seed coat. It is proposed that a pleiotropic effect of the seed coat pigmentation is to regulate grain dormancy (Flintham, 2000). The exact nature of the communication between tissues is not fully understood. Whilst much progress has been made in species such as *Arabidopsis*, with roles demonstrated for phytohormones, epigenetic factors and sugars (amongst others; reviewed in Nowack *et al.*, 2010; Locascio *et al.*, 2014; Radchuk & Borisjuk, 2014), still relatively little is understood about the molecular basis of this signalling in cereals. Caution should be exercised when translating insight gained from *Arabidopsis* into cereals as it is possible that not all these processes and mechanisms are conserved, particularly as there are fundamental differences in the final composition of the seed/grain. For example, the *Arabidopsis* endosperm consists of a single cell type whilst the endosperm of mature wheat grains contains four major cell types (Olsen, 2001).

From the results presented in this thesis, it is therefore, not possible to say conclusively whether the increased pericarp cell size in 5A+ NILs is due to a direct effect on cell expansion in the pericarp or an indirect effect of increased endosperm growth. As discussed previously, the early stage at which the grain length phenotype appears would suggest a direct effect on pericarp cell size, but this will need to be confirmed genetically. This could be tested through the assessment of pericarp cell size in F<sub>1</sub> hybrids from reciprocal crosses between 5A- and 5A+ NILs. As the pericarp is an exclusively maternal tissue, if the 5A QTL directly affects cell size in the pericarp then F<sub>1</sub> grains resulting from a 5A+ NIL pollinated by a 5A- NIL would have the 5A+ large pericarp cell size phenotype, whilst the reciprocal cross would not. Conversely, if the 5A QTL affects pericarp cell size as an indirect effect of endosperm growth then only F<sub>1</sub> grains from the 5A- NIL pollinated by the 5A+ NIL would have the large pericarp cell size phenotype. These experiments are currently being conducted. Usually, studies of this nature are challenging in wheat as the subtle phenotypic differences associated with QTL in polyploids can be masked by the phenotypic variation observed between individual F<sub>1</sub> grains (e.g. ~ 5% difference in grain size components in the case of the 5A and 6A QTL). However, the robust effect on pericarp cell size in the 5A NILs that is independent of absolute grain length could overcome this and opens up new opportunities for parent of origin studies in wheat.

It would also be interesting to assess how the development of the endosperm is affected by the 5A QTL. Whilst the work in this PhD has provided insights into the mechanisms underlying both QTL by breaking down overall grain yield into its constituent parts in the form of specific grain size components, the understanding remains mostly on a whole grain level, both from the phenotype and transcriptome points of view. The next steps to take would be to dissect this down even further to look at the individual tissues within the grain such as the endosperm, embryo and pericarp. Indeed, even breaking the grain down into the three main tissues remains quite a simplistic view as

each tissue is composed of several different layers and cell types. It would be very interesting to examine these tissues microscopically during carpel/grain development to understand the effects of the QTL in more mechanistic detail and this could be complemented by tissue specific expression studies.

### 5.1.3 Maternal control of grain size

All three pairs of NILs assessed (5A, 6A and *TaGW2-A*) point towards the maternal control of final grain size. Differences in carpel size were observed between *TaGW2-A* NILs before heading suggesting that *TaGW2-A* acts on maternal tissue. Borderline non-significant differences in carpel size were observed between 6A NILs, suggesting that this QTL could also act on maternal tissue before fertilisation. Although differences in grain length were established after fertilisation in 5A NILs, the QTL is associated with larger cells in the pericarp, a maternal tissue. Although as discussed above, this may or may not be as a result of a direct effect on pericarp cell expansion. The maternal control of seed/grain size has been demonstrated both genetically and phenotypically in many species including Arabidopsis, rice, wheat and maize (Hasan *et al.*, 2011; Li & Li, 2015; Zhang *et al.*, 2016). Studies have shown that the maternal control of seed/grain size can be exerted through a range of mechanisms, affecting cells in maternal tissues both pre- or post-fertilisation. For example, the Arabidopsis gene, *KLUH* acts maternally to increase seed size through the positive regulation of cell proliferation in the integument (Adamski *et al.*, 2009) and studies suggest that this function is conserved in the wheat orthologue, *TaCYP78A* (Ma *et al.*, 2016). Conversely, the Arabidopsis *ARF2* gene acts as a negative regulator of seed size with a loss-of-function mutant producing 20-40% heavier seeds. The increase in seed weight was associated with increased numbers of cells in the seed coat as a result of increased cell proliferation in the integument/ovule before fertilisation (Schruff *et al.*, 2006). Similarly, *GW2* in rice and its orthologue in Arabidopsis (*DA2*) influence grain/seed size through restriction of cell proliferation in the maternal tissue (Song, X-J *et al.*, 2007; Xia, T. *et al.*, 2013). This is consistent with the results from this thesis and Simmonds *et al.*, 2016 that *TaGW2-A* acts on maternal tissue although the effect on cell size and number has not yet been determined. *DA1*, a target of *DA2* in Arabidopsis, also acts synergistically with *DA2* to limit cell proliferation in the integument (Xia, T. *et al.*, 2013; Dong *et al.*, 2017). Genes have also been identified that act maternally to influence cell expansion in maternal tissue, for example *TTG2* and *AP2* (discussed above; Garcia *et al.*, 2005; Ohto *et al.*, 2005).

Programmed cell death (PCD) in the pericarp tissue has also been shown to be important for the maternal control of grain size. It has been proposed that PCD is an important step for enlargement of the pericarp to accommodate endosperm growth. Downregulation of *VACUOLAR-PROCESSING ENZYME 4* (*VPE4*) by RNAi in barley resulted in delayed PCD in the pericarp and consequently smaller grains (Radchuk *et al.*, 2017). One of the DE genes between 5A NILs was annotated as *VPE4* (Table 4.22) and taking this together with the differential regulation observed of proteolytic components, this could suggest a role of PCD in regulation of grain size in the 5A NILs.

However, the most extensive PCD occurs during the later stages of grain development and so this could be a downstream effect. Differences in the progression of PCD in 5A NILs could be identified by a histological analysis of developing grains as in Radchuk *et al.*, 2017. However, this might be challenging due to the number of samples that may be required to detect subtle differences between NILs.

The assignment of the 5A, 6A and *TaGW2-A* effects to the maternal parent will need to be confirmed genetically and this could be determined with the F<sub>1</sub> experiments described above. The maternal parent may contribute to final seed size through other mechanisms in addition to the presence of the pericarp/seed coat and the mechanical constraints it imposes. The mother plant plays many important roles in the development of the grain/seed including provisioning of nutrients to the developing grain, responses to the environment during grain development and the imprinting of genes after fertilisation all of which have been shown to influence final grain size (discussed in Zhang *et al.*, 2016). If the effects of either the 6A or 5A QTL can be assigned to the maternal parent then it will be interesting to understand exactly how the maternal parent contributes to the final phenotype. Identifying genes that act maternally to influence grain size could have advantages in a breeding context, particularly with respect to hybrid seed generation.

#### **5.1.4 Importance of early grain development**

Regardless of the putative direct maternal effects of the 5A and 6A QTL, all three pairs of NILs highlight the importance of early carpel/grain development in determining final grain size, consistent with previous studies in wheat and other cereals. However, despite the importance of these early stages, relatively little is known about the mechanisms underlying early grain development. Studies have characterised these stages phenotypically to a certain extent but most characterisation has focussed on the later stages of grain development, mainly on endosperm development. The same is especially true in terms of characterisation on the transcriptional and molecular level, as discussed briefly in Chapter 4. Although numerous wheat grain RNA-Seq studies have been performed, very few have focussed on stages of grain development as early as those described in Chapter 4. This is evidenced by the fact that of 148 grain RNA seq samples in the wheat expVIP database only six were taken at stages earlier than 8 dpa, four of which formed part of the same study. Additionally, there are no RNA-Seq samples in the expVIP database from ovules i.e. pre-anthesis. The results from this thesis strongly suggest that understanding the mechanisms underlying these early stages will be critical to identify ways to manipulate final grain size. Based on the ability of grains to compensate for early events in grain development, it is tempting to speculate that manipulating genes and pathways that affect these early stages could provide grain size increases that are more robust to environmental variation.

## **5.2 Potential consequences of increasing grain size and pleiotropic effects of the 5A and 6A QTL**

Increases in grain weight are often associated with pleiotropic effects either on the grain itself or on other plant organs. When considering the pleiotropic effects of QTL, the effects could either be due to other genes within the QTL interval. Alternatively, they could be due to the gene(s) controlling grain size themselves either as an indirect results of increasing grain size or as a direct effect of the gene in another part of the plant.

### **5.2.1 Pleiotropic effects on yield components**

Despite grain weight being more stably inherited than overall yield itself, increases in grain weight have previously been associated with negative pleiotropic effects on other yield components such as grain number and spike number (Kuchel *et al.*, 2007). However, results from this PhD and other studies have shown that these components are not inherently linked and can be genetically separated (Griffiths *et al.*, 2015). Indeed, the fact that alterations in spike and grain number have downstream effects on grain size, likely due to competition to resources, this does not necessarily mean that changes in grain size will affect grain and spike number. During this PhD, these components were assessed in the 5A and 6A NILs. In the 6A NILs there were no consistent negative effects across years on either grain number or spike (tiller) number, although there were some negative effects in individual years (Table 4.2). This was consistent with the previous studies of these NILs (Simmonds *et al.*, 2014). In terms of the 5A NILs, 5A+ NILs had significantly reduced grain and tiller number across years although these effects were both driven by particularly strong effects in a single year (Table 4.7). We hypothesised that the combination of these smaller negative effects could explain why neither the 6A or 5A NILs had consistent differences in final grain yield, despite consistent increases in TGW. Alternatively, our evaluation of yield components based on a ten-spike sampling might not be robust enough to allow us to detect differences. This could be due to the fact that we usually select ten spikes, corresponding to the main tiller in most cases. For these spikes, we observe increase in spike yield (Table 4.2, Table 4.7). However, by using this sampling strategy we could be missing pleiotropic effects on spikes further behind in development (e.g. third or fourth spike), which could arise from compensation effects from the larger grains in the main spikes of the 6A+ and 5A+ NILs. The negative effects on other yield components could either be as a result of additional genes in the introgressed regions of the NILs or as an effect of the causal genes themselves. For example, a minor QTL for tiller number was identified in the 6A introgression, but this mapped distal to the QTL for TGW (Simmonds *et al.*, 2014). This suggests that in the 6A NILs the pleiotropic effect on tiller number is due to another gene in the interval rather than an effect of the 6A causal gene itself.

### 5.2.2 Pleiotropic developmental effects

We also observed developmental differences between 6A NILs, including differences in flowering time and senescence, with 6A+ NILs flowering earlier and senescing later. This could suggest that the 6A QTL is associated with an extended grain filling period. However, this was not assessed directly and it has previously been shown that an increase in the time between flowering and senescence (green canopy duration; GCD) does not always result in an increased duration of grain filling (Borrill *et al.*, 2015b). Similar to the tillering effect, a QTL for GCD was identified in the 6A NIL introgression but did not show any correlation with TGW in the original QTL analysis and so the two traits are likely to be under independent genetic control (Simmonds *et al.*, 2014). That said, some genes that influence grain/seed size in other species have also been shown to affect other developmental traits such as senescence and flowering time. For example, in *Arabidopsis*, *DA1* acts as a negative regulator of seed size but also promotes senescence with mutants having larger seeds and delayed senescence (Li *et al.*, 2008; Vanhaeren *et al.*, 2016). Additionally, *DA1* also affects the size of other organs in addition to the seed, for example, it acts as a negative regulator of petal and leaf size suggesting that it is a general regulator of plant growth rather than specifically seed size.

It is not known whether the 5A and 6A genes have grain specific effects or whether they could be general regulators of organ size, for example, these genes could influence leaf and root size. This could have implications both in positive and negative ways. For example, a non-grain specific effect could be seen as wasteful with resources going into non-grain biomass production. Alternatively, plants with larger leaves could have increased photosynthetic capacity through increased area for light interception and a larger root system could also be beneficial. If these genes do have similar effects on the development of grains and leaves, e.g. the 5A gene(s) could increase cell expansion in both tissues, then this could be a useful tool for determining the mechanism by which the genes act. The leaf could act as a more tractable system for performing experiments than the grain, and this has proven to be a useful tool for understanding the function of genes that control seed size in *Arabidopsis* such as *DA1*, *DA2* and *BIG BROTHER*. It would be useful to investigate this possibility in subsequent studies and determine if putative effects are sufficiently strong and robust to be properly quantified.

### 5.2.3 Pleiotropic effects on grain nutrient composition

An avenue that was not explored in this PhD was the effect of these QTL on the composition of the grain itself, aside from increasing the overall size and weight. For example, we did not examine the effect that manipulating the grain size has on the micronutrient, protein or starch content of the grain. Negative correlations between grain weight and grain protein content have been documented (Simmonds, 1995), proposed to be a dilution effect of increased starch in the grain. It is, therefore, possible that the increases in grain weight associated with the 6A and 5A QTL could be associated with a decrease in nutritional value and quality. Based on the fact that both QTL act

during very early grain development, before grain filling and starch accumulation has begun (Drea *et al.*, 2005; Shewry *et al.*, 2012), I would hypothesise that these QTL act to enhance grain filling capacity rather than a particular aspect of grain filling itself. Therefore, I would expect the grain filling process to proceed in the same way, with the relative proportions of protein, starch and micronutrients etc. remaining roughly the same.

#### **5.2.4 Understanding the causes of pleiotropic effects**

These pleiotropic effects could be assessed in the 5A and 6A NILs, but this could be challenging for a number of reasons. It will not be possible using the NILs to separate effects that are due to the causal gene(s) themselves from those effects that are a result of other genes within the introgressed intervals. Additionally, as the phenotypic differences between NILs are very subtle, it may be difficult to separate truly biological effects from random biological variation. Identifying the causal gene(s) will allow larger variation to be explored and open new opportunities for studying the function of the gene in more detail. Techniques such as RNAi, CRISPR or TILLING (reviewed in Uauy, 2017b) will provide routes to explore a wide range of variation in the underlying genes ranging from understanding the effects of knock-out mutations to exploring how more subtle allelic variations can affect gene function.

### **5.3 Combining beneficial alleles**

Understanding the specific biological mechanisms and genes underlying the 5A and 6A QTL allows hypotheses about combining beneficial alleles of genes to be generated and tested in an informed and targeted way. This can occur on many different levels.

#### **5.3.1 Combining homoeologues**

Identifying the causal genes of the 5A and 6A QTL will allow the B and D homoeologues to be identified. This will be important due to the subtle effects of grain weight QTL in hexaploid wheat compared to grain weight QTL in diploid species (Borrill *et al.*, 2015a; Uauy, 2017b).

Simultaneously modulating the function of all three homoeologues has the potential to expand the range of phenotypic variation and achieve effects comparable to those in diploids, for example *NAM-B1* (Uauy *et al.*, 2006; Avni *et al.*, 2014). The increased phenotypic range will be important both for understanding gene function and also for providing breeders with novel allelic combinations as simultaneous beneficial mutations in all three homoeologues are unlikely to occur naturally. Alternatively, the three homoeologues may not have completely redundant functions as certain copies may have diverged in function and/or regulation. Although the causal genes underlying the 5A and 6A QTL have not yet been identified, this concept is being explored with the 5A candidate genes discussed above. Additionally, studies in the lab are currently investigating the effects of combining *Tagw2-A* with TILLING knock out mutations in the B and D homoeologues (*TaGW2-B* and *-D*). Other groups have also generated lines with mutations in all three *TaGW2*

homoeologues using CRISPR, which provide an alternative and complementary method to understand the function and interaction between the homoeologues.

### **5.3.2 Combining components of pathways involved in grain size regulation**

Different components of the same pathway could be combined to give additive effects in grain size. As discussed above, in *Arabidopsis*, *DA1* is a target of the E3 ubiquitin ligase *DA2*. Individually, both act to negatively regulate organ size through the suppression of cell proliferation and *da1 da2* double mutants have a synergistic effect on organ size. Combining components of the same pathway may not always provide additive/synergistic effects and mutations could be epistatic (i.e. the phenotypic effect of one gene is dependent on the presence of the second modifier gene). Regardless, identifying specific pathways will allow the function of different components to be investigated and manipulated to fine-tune the final grain size. The ubiquitin-related differentially expressed genes from the 5A RNA-Seq study represent good candidates for initial investigation in this way.

Additionally, genes affecting different pathways and grain size components could be combined to give additive/synergistic effects. For example, combining genes that regulate cell expansion with genes regulating cell proliferation or genes that increase grain length with genes that increase grain width. In the lab, NILs have been generated that combine the 5A QTL (grain length; cell expansion) and the 6A QTL (grain width; possibly cell proliferation) in a common genetic background. Initial results suggest that combining the two QTL does have an additive effect on grain weight through increased grain width and grain length. In the 2017 field trials we also assessed the cell size phenotype of these NILs, the results of which are currently being analysed. We hypothesise that NILs with both the 5A and 6A QTL will have increased cell number and cell size. Interestingly, combining the two QTL seems to have a 'stabilising' effect on the final grain weight. In 2016, the 6A grain width effect did not perform well alone and did not have significantly increased grain weight (Table 4.1). However, combining the 5A QTL and 6A QTL still had significantly higher grain weight than the 5A QTL alone, suggesting that there could be some interaction between the two QTL. This was also seen when analysing historical data from the UK public Avalon x Cadenza population (Simmonds, unpublished results). Identifying the genes underlying these QTL will allow the exact nature of this interaction to be investigated further.

### **5.3.3 Combining grain size genes with other aspects of plant development**

Combining genes that affect different yield components could provide a solution to overcome the negative pleiotropic effects associated with increasing individual yield components. For example, increases in grain number are often associated with decreases in grain size and consequently no increase in overall grain yield is achieved. Combining a gene that increases grain number and a gene that influences grain size could act to increase grain number while maintaining or enhancing the grain size. Similar approaches could be taken to maintaining or increasing the nutritional value of the grain.

Lastly, while this thesis has focused on the genetic mechanisms underlying grain size and yield, the agronomic aspects should not be ignored. Breeders select to maximise yield under specific planting densities and agronomy conditions. When developing NILs, we modify a single region of the genome which is extremely useful to study the trait in question (grain size in this thesis), but could alter the overall balance of the canopy that was selected to maximise yield. Therefore, it is likely that changes in agronomy practices may be required to maximise the chance that the positive NILs for grain size translate into yield. This is currently being tested for the 2017-2018 field season by modifying seeding rates and fertilisation regimes to better understand the interactions between genetics, environment and agronomy management.

## 5.4 Concluding statement

Overall, this thesis has provided new insights into the mechanisms controlling grain size in wheat through the characterisation of two distinct grain size QTL in multiple different ways. The results presented here highlight the importance of early grain development in determining final grain size in wheat, and provide direct genetic evidence for the importance of the pericarp tissue. Fine-mapping of the two QTL revealed complex underlying genetic architectures. Although the causal genes were not identified, the intervals were reduced and the new flanking markers have been shared with breeders to facilitate more efficient selection of the beneficial regions. The 5A transcriptomic study identified differentially expressed genes and pathways that could be involved in the control of grain size, a subset of which are now being functionally characterised. Ultimately, identifying the genes and pathways that control grain size and understanding how they interact will allow breeders to manipulate and fine-tune final grain yield in wheat in novel ways.

## 6 References

- Adamski NM, Anastasiou E, Eriksson S, O'Neill CM, Lenhard M. 2009. Local maternal control of seed size by *KLUH/CYP78A5*-dependent growth signaling. *Proceedings of the National Academy of Sciences* **106**(47): 20115-20120.
- Akhunov ED, Goodyear AW, Geng S, Qi LL, Echalié B, Gill BS, Miftahudin, Gustafson JP, Lazo G, Chao S, et al. 2003. The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. *Genome Res* **13**(5): 753-763.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**(3): 403-410.
- Andrews S 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Avni R, Zhao RR, Pearce S, Jun Y, Uauy C, Tabbita F, Fahima T, Slade A, Dubcovsky J, Distelfeld A. 2014. Functional characterization of *GPC-1* genes in hexaploid wheat. *Planta* **239**(2): 313-324.
- Barrero JM, Cavanagh C, Verbyla KL, Tibbits JFG, Verbyla AP, Huang BE, Rosewarne GM, Stephen S, Wang P, Whan A, et al. 2015. Transcriptomic analysis of wheat near-isogenic lines identifies *PM19-A1* and *A2* as candidates for a major dormancy QTL. *Genome Biology* **16**(1): 93.

- Bednarek JP, Boulafloous A, Girousse C, Ravel C, Tassy C, Barret P, Mouzeyar MFB, Said. 2012.** Down-regulation of the TaGW2 gene by RNA interference results in decreased grain size and weight in wheat. *Journal of Experimental Botany* **63**(16): 5945-5955.
- Borrill P, Adamski N, Uauy C. 2015a.** Genomics as the key to unlocking the polyploid potential of wheat. *New Phytologist* **208**(4): 1008-1022.
- Borrill P, Fahy B, Smith AM, Uauy C. 2015b.** Wheat Grain Filling Is Limited by Grain Filling Capacity rather than the Duration of Flag Leaf Photosynthesis: A Case Study Using NAM RNAi Plants. *PLoS one* **10**(8): e0134947.
- Borrill P, Harrington SA, Uauy C. 2017.** Genome-Wide Sequence and Expression Analysis of the NAC Transcription Factor Family in Polyploid Wheat. *G3: Genes/Genomes/Genetics*. **7**:30149
- Borrill P, Ramirez-Gonzalez R, Uauy C. 2016.** expVIP: a Customizable RNA-seq Data Analysis and Visualization Platform. *Plant physiology* **170**(4): 2172-2186.
- Botella JR. 2012.** Can heterotrimeric G proteins help to feed the world? *Trends in Plant Science* **17**(10): 563-568.
- Bray NL, Pimentel H, Melsted P, Pachter L. 2016.** Near-optimal probabilistic RNA-seq quantification. *Nat Biotech* **34**(5): 525-527.
- Breseghele F, Sorrells ME. 2007.** QTL analysis of kernel size and shape in two hexaploid wheat mapping populations. *Field Crops Research* **101**(2): 172-179.
- Brinton J, Simmonds J, Minter F, Leverington-Waite M, Snape J, Uauy C. 2017.** Increased pericarp cell length underlies a major quantitative trait locus for grain weight in hexaploid wheat. *New Phytologist* **215**(3): 1026-1038.
- Calderini D, Abeledo L, Savin R, Slafer GA. 1999.** Effect of temperature and carpel size during pre-anthesis on potential grain weight in wheat. *The Journal of Agricultural Science* **132**(04): 453-459.
- Chapman JA, Mascher M, Buluç A, Barry K, Georganas E, Session A, Strnadova V, Jenkins J, Sehgal S, Olliker L, et al. 2015.** A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biology* **16**(1): 26.
- Cheng Y, Qin G, Dai X, Zhao Y. 2007.** NPY1, a BTB-NPH3-like protein, plays a critical role in auxin-regulated organogenesis in *Arabidopsis*. *Proceedings of the National Academy of Sciences* **104**(47): 18825-18829.
- Chevalier C, Bourdon M, Pirrello J, Cheniclet C, Gévaudant F, Frangne N. 2014.** Endoreduplication and fruit growth in tomato: evidence in favour of the karyoplasmic ratio theory. *Journal of Experimental Botany* **65**(10): 2731-2746.
- Chow C-N, Zheng H-Q, Wu N-Y, Chien C-H, Huang H-D, Lee T-Y, Chiang-Hsieh Y-F, Hou P-F, Yang T-Y, Chang W-C. 2016.** PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic Acids Research* **44**(Database issue): D1154-D1160.
- Clavijo BJ, Garcia Accinelli G, Wright J, Heavens D, Barr K, Yanes L, Di Palma F. 2017a.** W2RAP: a pipeline for high quality, robust assemblies of large complex genomes from short read data. *bioRxiv*. <https://doi.org/10.1101/110999>
- Clavijo BJ, Venturini L, Schudoma C, Garcia Accinelli G, Kaithakottil G, Wright J, Borrill P, Kettleborough G, Heavens D, Chapman H, et al. 2017b.** An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome Research* **27**: 885-896.
- Cosgrove DJ. 2005.** Growth of the plant cell wall. *Nature Reviews Molecular Cell Biology* **6**(11): 850-861.
- Disch S, Anastasiou E, Sharma VK, Laux T, Fletcher JC, Lenhard M. 2006.** The E3 ubiquitin ligase BIG BROTHER controls *Arabidopsis* organ size in a dosage-dependent manner. *Current Biology* **16**(3): 272-279.
- Distelfeld A, Li C, Dubcovsky J. 2009.** Regulation of flowering in temperate cereals. *Current Opinion In Plant Biology* **12**(2): 178-184.

- Dong H, Dumenil J, Lu FH, Na L, Vanhaeren H, Naumann C, Klecker M, Prior R, Smith C, McKenzie N, et al. 2017.** Ubiquitylation activates a peptidase that promotes cleavage and destabilization of its activating E3 ligases and diverse growth regulatory proteins to limit cell proliferation in Arabidopsis. *Genes Dev* **31**(2): 197-208.
- Drea S, Leader DJ, Arnold BC, Shaw P, Dolan L, Doonan JH. 2005.** Systematic spatial analysis of gene expression during wheat caryopsis development. *The Plant Cell* **17**(8): 2172-2185.
- Du L, Li N, Chen L, Xu Y, Li Y, Zhang Y, Li C, Li Y. 2014.** The Ubiquitin Receptor *DA1* Regulates Seed and Organ Size by Modulating the Stability of the Ubiquitin-Specific Protease *UBP15/SOD2* in Arabidopsis. *The Plant Cell* **26**(2): 665-677.
- Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, Li X, Zhang Q. 2006.** *GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theoretical and Applied Genetics* **112**(6): 1164-1171.
- Fang W, Wang Z, Cui R, Li J, Li Y. 2012.** Maternal control of seed size by *EOD3/CYP78A6* in Arabidopsis thaliana. *The Plant Journal* **70**(6): 929-939.
- FAO 2017.** Online statistical database: Food balance. FAOSTAT.
- FAO, IFAD, UNICEF, WFP, WHO. 2017.** The State of Food Security and Nutrition in the World 2017. Building resilience for peace and food security. Rome, FAO
- Farré A, Sayers L, Leverington-Waite M, Goram R, Orford S, Wingen L, Mumford C, Griffiths S. 2016.** Application of a library of near isogenic lines to understand context dependent expression of QTL for grain yield and adaptive traits in bread wheat. *BMC Plant Biology* **16**.
- Flintham JE. 2000.** Different genetic components control coat-imposed and embryo-imposed dormancy in wheat. *Seed Science Research* **10**(1): 43-50.
- Garcia D, Gerald JNF, Berger F. 2005.** Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in Arabidopsis. *The Plant Cell* **17**(1): 52-60.
- Garcia D, Saingery V, Chambrier P, Mayer U, Jürgens G, Berger F. 2003.** Arabidopsis haiku Mutants Reveal New Controls of Seed Size by Endosperm. *Plant Physiology* **131**(4): 1661-1670.
- Garrison E, Marth G. 2012.** Haplotype-based variant detection from short-read sequencing. *arXiv.1207.3907*
- Gegas VC, Nazari A, Griffiths S, Simmonds J, Fish L, Orford S, Sayers L, Doonan JH, Snape JW. 2010.** A genetic framework for grain size and shape variation in wheat. *The Plant Cell* **22**(4): 1046-1056.
- Gonzalez-Navarro OE, Griffiths S, Molero G, Reynolds MP, Slafer GA. 2016.** Variation in developmental patterns among elite wheat lines and relationships with yield, yield components and spike fertility. *Field Crops Research* **196**(Supplement C): 294-304.
- Grant CE, Bailey TL, Noble WS. 2011.** FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**(7): 1017-1018.
- Griffiths S, Wingen L, Pietragalla J, Garcia G, Hasan A, Miralles D, Calderini DF, Ankleshwaria JB, Waite ML, Simmonds J. 2015.** Genetic dissection of grain size and grain number trade-offs in CIMMYT wheat germplasm. *PloS one* **10**(3): e0118847.
- Guttman M, Rinn JL. 2012.** Modular regulatory principles of large non-coding RNAs. *Nature* **482**(7385): 339-346.
- Hasan AK, Herrera J, Lizana C, Calderini DF. 2011.** Carpel weight, grain length and stabilized grain water content are physiological drivers of grain weight determination of wheat. *Field Crops Research* **123**(3): 241-247.
- Hershko A, Ciechanover A. 1998.** The ubiquitin system. *Annu Rev Biochem* **67**: 425-479.

- Hong Y, Chen L, Du L-p, Su Z, Wang J, Ye X, Qi L, Zhang Z. 2014.** Transcript suppression of TaGW2 increased grain width and weight in bread wheat. *Functional & Integrative Genomics* **14**(2): 341-349.
- Huang K, Wang D, Duan P, Zhang B, Xu R, Li N, Li Y. 2017.** WIDE AND THICK GRAIN 1, which encodes an otubain-like protease with deubiquitination activity, influences grain size and shape in rice. *The Plant Journal* **91**(5): 849-860.
- Huang R, Jiang L, Zheng J, Wang T, Wang H, Huang Y, Hong Z. 2013.** Genetic bases of rice grain shape: so many genes, so little known. *Trends in Plant Science* **18**(4): 218-226.
- Huang X, Qian Q, Liu Z, Sun H, He S, Luo D, Xia G, Chu C, Li J, Fu X. 2009.** Natural variation at the *DEP1* locus enhances grain yield in rice. *Nature Genetics* **41**(4): 494-497.
- IWGSC. 2014.** A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **345**(6194): 1251788.
- Jiang W-B, Huang H-Y, Hu Y-W, Zhu S-W, Wang Z-Y, Lin W-H. 2013.** Brassinosteroid regulates seed size and shape in Arabidopsis. *Plant physiology* **162**(4): 1965-1977.
- Johnson CS, Kolevski B, Smyth DR. 2002.** TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of Arabidopsis, encodes a WRKY transcription factor. *The Plant Cell* **14**(6): 1359-1375.
- Kuchel H, Williams KJ, Langridge P, Eagles Ha, Jefferies SP. 2007.** Genetic dissection of grain yield in bread wheat. I. QTL analysis. *Theoretical and Applied Genetics* **115**(8): 1029-1041.
- Kumar A, Mantovani EE, Seetan R, Soltani A, Echeverry-Solarte M, Jain S, Simsek S, Doehlert D, Alamri MS, Elias EM, et al. 2016.** Dissection of Genetic Factors underlying Wheat Kernel Shape and Size in an Elite x Nonadapted Cross using a High Density SNP Linkage Map. *Plant Genome* **9**(1).
- Langmead B, Salzberg SL. 2012.** Fast gapped-read alignment with Bowtie 2. *Nature Method* **9**(4): 357-359.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009.** The sequence alignment/map format and SAMtools. *Bioinformatics* **25**.
- Li J, Jiang J, Qian Q, Xu Y, Zhang C, Xiao J, Du C, Luo W, Zou G, Chen M, et al. 2011.** Mutation of Rice *BC12/GDD1*, Which Encodes a Kinesin-Like Protein That Binds to a GA Biosynthesis Gene Promoter, Leads to Dwarfism with Impaired Cell Elongation. *The Plant Cell* **23**(2): 628-640.
- Li N, Li Y. 2014a.** Ubiquitin-mediated control of seed size in plants. *Frontiers in Plant Science* **5**: 332.
- Li N, Li Y. 2015.** Maternal control of seed size in plants. *Journal of Experimental Botany* **66**(4): 1087-1097.
- Li N, Li Y. 2016.** Signaling pathways of seed size control in plants. *Current Opinion in Plant Biology* **33**(Supplement C): 23-32.
- Li Q, Li L, Yang X, Warburton ML, Bai G, Dai J, Li J, Yan J. 2010.** Relationship, evolutionary fate and function of two maize co-orthologs of rice GW2 associated with kernel size and weight. *BMC Plant Biology* **10**(1): 143.
- Li YH, Zheng LY, Corke F, Smith C, Bevan MW. 2008.** Control of final seed and organ size by the DA1 gene family in Arabidopsis thaliana. *Genes & Development* **22**(10): 1331-1336.
- Liu W, Zhihui Wu, Yufeng Zhang, Dandan Guo, Yuzhou Xu, Weixia Chen, Haiying Zhou, Mingshan You, Li B. 2014.** Transcriptome analysis of wheat grain using RNA-Seq. *Frontiers of Agricultural Science and Engineering* **1**(3): 214-222.
- Liu Y, Wang F, Zhang H, He H, Ma L, Deng XW. 2008.** Functional characterization of the Arabidopsis ubiquitin-specific protease gene family reveals specific role and redundancy of individual members in development. *The Plant Journal* **55**(5): 844-856.
- Lizana XC, Riegel R, Gomez LD, Herrera J, Isla A, McQueen-Mason SJ, Calderini DF. 2010.** Expansins expression is associated with grain size dynamics in wheat (*Triticum aestivum* L.). *Journal of Experimental Botany* **61**(4): 1147-1157.

- Locascio A, Roig-Villanova I, Bernardi J, Varotto S. 2014.** Current perspectives on the hormonal control of seed development in Arabidopsis and maize: a focus on auxin. *Frontiers in Plant Science* **5**: 412.
- Ma M, Zhao H, Li Z, Hu S, Song W, Liu X. 2016.** *TaCYP78A5* regulates seed size in wheat (*Triticum aestivum*). *Journal of Experimental Botany* **67**(5): 1397-1410.
- Mir RR, Kumar N, Jaiswal V, Girdharwal N, Prasad M, Balyan HS, Gupta PK. 2012.** Genetic dissection of grain weight in bread wheat through quantitative trait locus interval and association mapping. *Molecular Breeding* **29**(4): 963-972.
- Mizukami Y, Fischer RL. 2000.** Plant organ size control: *AINTEGUMENTA* regulates growth and cell numbers during organogenesis. *Proceedings of the National Academy of Sciences* **97**(2): 942-947.
- Munoz M, Calderini DF. 2015.** Volume, water content, epidermal cell area, and *XTH5* expression in growing grains of wheat across ploidy levels. *Field Crops Research* **173**: 30-40.
- Nowack Moritz K, Ungru A, Bjerkan Katrine N, Grini Paul E, Schnittger A. 2010.** Reproductive cross-talk: seed development in flowering plants. *Biochemical Society Transactions* **38**(2): 604-612.
- Ohto M-a, Fischer RL, Goldberg RB, Nakamura K, Harada JJ. 2005.** Control of seed mass by *APETALA2*. *Proceedings of the National Academy of Sciences of the United States of America* **102**(8): 3123-3128.
- Ohto M-a, Floyd SK, Fischer RL, Goldberg RB, Harada JJ. 2009.** Effects of *APETALA2* on embryo, endosperm, and seed coat development determine seed size in Arabidopsis. *Sexual Plant Reproduction* **22**(4): 277-289.
- Olsen OA. 2001.** ENDOSPERM DEVELOPMENT: Cellularization and Cell Fate Specification. *Annu Rev Plant Physiol Plant Mol Biol* **52**: 233-267.
- Pallotta M, Warner P, Fox R, Kuchel H, Jefferies S, Langridge P 2003.** Marker assisted wheat breeding in the southern region of Australia. *Proceedings of the 10th international wheat genetics symposium, Paestum, Italy*: Istituto Sperimentale per la Cerealicoltura Roma, Italy. 789-791.
- Peng FY, Hu Z, Yang R-C. 2016.** Bioinformatic prediction of transcription factor binding sites at promoter regions of genes for photoperiod and vernalization responses in model and temperate cereal plants. *BMC Genomics* **17**: 573.
- Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. 2017.** Differential analysis of RNA-seq incorporating quantification uncertainty. *Nature Method* **14**(7): 687-690.
- Qi P, Lin Y-S, Song X-J, Shen J-B, Huang W, Shan J-X, Zhu M-Z, Jiang L, Gao J-P, Lin H-X. 2012.** The novel quantitative trait locus *GL3.1* controls rice grain size and yield by regulating *Cyclin-T1;3*. *Cell Res* **22**(12): 1666-1680.
- Quinlan AR, Hall IM. 2010.** BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**(6): 841-842.
- Radchuk V, Borisjuk L. 2014.** Physical, metabolic and developmental functions of the seed coat. *Frontiers in Plant Science* **5**: 510.
- Radchuk V, Tran V, Radchuk R, Diaz-Mendoza M, Weier D, Fuchs J, Riewe D, Hensel G, Kumlehn J, Munz E, et al. 2017.** Vacuolar processing enzyme 4 contributes to maternal control of grain size in barley by executing programmed cell death in the pericarp. *New Phytologist*: **218**:1127
- Radchuk V, Weier D, Radchuk R, Weschke W, Weber H. 2011.** Development of maternal seed tissue in barley is mediated by regulated cell expansion and cell disintegration and coordinated with endosperm growth. *Journal of Experimental Botany* **62**(3): 1217-1227.
- Ramirez-Gonzalez RH, Uauy C, Caccamo M. 2015.** PolyMarker: A fast polyploid primer design pipeline. *Bioinformatics* **31**(February): 2038-2039.
- Riefler M, Novak O, Strnad M, Schmülling T. 2006.** Arabidopsis cytokinin receptor mutants reveal functions in shoot growth, leaf senescence, seed size, germination, root development, and cytokinin metabolism. *The Plant Cell* **18**(1): 40-54.

- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. 2012.** Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**(7): 676-682.
- Schruff MC, Spielman M, Tiwari S, Adams S, Fenby N, Scott RJ. 2006.** The *AUXIN RESPONSE FACTOR 2* gene of *Arabidopsis* links auxin signalling, cell division, and the size of seeds and other organs. *Development* **133**(2): 251-261.
- Shewry PR, Mitchell RaC, Tosi P, Wan Y, Underwood C, Lovegrove A, Freeman J, Toole Ga, Mills ENC, Ward JL. 2012.** An integrated study of grain development of wheat (cv. Hereward). *Journal of Cereal Science* **56**(1): 21-30.
- Shorinola O, Balcárková B, Hyles J, Tibbits JFG, Hayden MJ, Holuřova K, Valárik M, Distelfeld A, Torada A, Barrero JM, et al. 2017b.** Haplotype Analysis of the Pre-harvest Sprouting Resistance Locus Phs-A1 Reveals a Causal Role of TaMKK3-A in Global Germplasm. *Frontiers in Plant Science* **8**(1555).
- Simmonds J, Scott P, Brinton J, Mestre TC, Bush M, Blanco A, Dubcovsky J, Uauy C. 2016.** A splice acceptor site mutation in *TaGW2-A1* increases thousand grain weight in tetraploid and hexaploid wheat through wider and longer grains. *Theoretical and Applied Genetics* **129**(6): 1099-1112.
- Simmonds J, Scott P, Leverington-Waite M, Turner AS, Brinton J, Korzun V, Snape J, Uauy C. 2014.** Identification and independent validation of a stable yield and thousand grain weight QTL on chromosome 6A of hexaploid wheat (*Triticum aestivum* L.). *BMC Plant Biology* **14**(1): 191.
- Simmonds NW. 1995.** The relation between yield and protein in cereal grain. *Journal of the Science of Food and Agriculture* **67**(3): 309-315.
- Slafer GA. 2003.** Genetic Basis of Yield as Viewed From a Crop Physiologist 's Perspective. *Annals of Applied Biology* **142**(2): 117-128.
- Song QJ, Shi JR, Singh S, Fickus EW, Costa JM, Lewis J, Gill BS, Ward R, Cregan PB. 2005.** Development and mapping of microsatellite (SSR) markers in wheat. *Theoretical and Applied Genetics* **110**(3): 550-560.
- Song X-J, Huang W, Shi M, Zhu M-Z, Lin H-X. 2007.** A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nature Genetics* **39**(5): 623-630.
- Su Z, Hao C, Wang L, Dong Y, Zhang X. 2011.** Identification and development of a functional marker of TaGW2 associated with grain weight in bread wheat (*Triticum aestivum* L.). *TAG. Theoretical and Applied Genetics*. **122**(1): 211-223.
- Sun F, Guo G, Du J, Guo W, Peng H, Ni Z, Sun Q, Yao Y. 2014.** Whole-genome discovery of miRNAs and their targets in wheat (*Triticum aestivum* L.). *BMC Plant Biology* **14**(1): 142.
- Trick M, Adamski NM, Mugford SG, Jiang C-C, Febrer M, Uauy C. 2012.** Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. *BMC Plant Biology* **12**(1): 14-14.
- Trusov Y, Botella JR. 2016.** Plant G-Proteins Come of Age: Breaking the Bond with Animal Models. *Frontiers in Chemistry* **4**(24).
- Uauy C. 2017a.** Wheat genomics comes of age. *Current opinion in plant biology* **36**(Supplement C): 142-148.
- Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J. 2006.** A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. *Science* **314**(5803): 1298-1301.
- Vanhaeren H, Nam Y-J, De Milde L, Chae E, Storme V, Weigel D, Gonzalez N, Inzé D. 2016.** Forever Young: The Role of Ubiquitin Receptor *DA1* and E3 Ligase *BIG BROTHER* in Controlling Leaf Growth and Development. *Plant physiology*. **173**:1269
- Wan Y, Poole RL, Huttly AK, Toscano-Underwood C, Feeney K, Welham S, Gooding MJ, Mills C, Edwards KJ, Shewry PR. 2008.** Transcriptome analysis of grain development in hexaploid wheat. *BMC genomics* **9**:121

- Wang S, Wu K, Yuan Q, Liu X, Liu Z, Lin X, Zeng R, Zhu H, Dong G, Qian Q, et al. 2012.** Control of grain size, shape and quality by *OsSPL16* in rice. *Nature genetics* **44**(8): 950-954.
- Williams K, Sorrells ME. 2014.** Three-Dimensional Seed Size and Shape QTL in Hexaploid Wheat (*Triticum aestivum* L.) Populations. *Crop Science* **54**(1): 98-110.
- Winfield MO, Allen AM, BurrIDGE AJ, Barker GLA, Benbow HR, Wilkinson PA, Coghill J, Waterfall C, Davassi A, Scopes G, et al. 2016.** High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant biotechnology journal* **14**(5): 1195-1206.
- Wu TD, Watanabe CK. 2005.** GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**(9): 1859-1875.
- Xia T, Li N, Dumenil J, Li J, Kamenski A, Bevan MW, Gao F, Li YH. 2013.** The Ubiquitin Receptor *DA1* Interacts with the E3 Ubiquitin Ligase *DA2* to Regulate Seed and Organ Size in Arabidopsis. *Plant Cell* **25**(9): 3347-3359.
- Xiao W, Brown RC, Lemmon BE, Harada JJ, Goldberg RB, Fischer RL. 2006.** Regulation of Seed Size by Hypomethylation of Maternal and Paternal Genomes. *Plant physiology* **142**(3): 1160-1168.
- Xing Y, Zhang Q. 2010.** Genetic and molecular bases of rice yield. *Annual review of plant biology* **61**: 421-442.
- Yang Z, Bai Z, Li X, Wang P, Wu Q, Yang L, Li L, Li X. 2012.** SNP identification and allelic-specific PCR markers development for *TaGW2*, a gene linked to wheat kernel weight. *Theoretical and Applied Genetics* **125**(5): 1057-1068.
- Young MD, Wakefield MJ, Smyth GK, Oshlack A. 2010.** Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* **11**(2): R14.
- Zhang K, Wang J, Zhang L, Rong C, Zhao F, Peng T, Li H, Cheng D, Liu X, Qin H, et al. 2013.** Association Analysis of Genomic Loci Important for Grain Weight Control in Elite Common Wheat Varieties Cultivated with Variable Water and Fertiliser Supply. *PloS one* **8**(3): e57853.
- Zhang X, Hirsch CN, Sekhon RS, de Leon N, Kaeppler SM. 2016.** Evidence for maternal control of seed size in maize from phenotypic and transcriptional analysis. *Journal of Experimental Botany* **67**(6): 1907-1917.