WHEAT FLORAL BIOLOGY

PROSPECTS FOR IMPROVING THE EFFICIENCY OF HYBRID SEED PRODUCTION AND ABIOTIC STRESS TOLERANCE

Carus John-Bejai

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

The University of Nottingham
School of Biosciences

Sutton Bonington Campus
Leicestershire, LE12 5RD
September 2017

Acknowledgements

I would firstly like to thank God Almighty for providing me with the opportunities that have led to the production of this thesis and for being a source of strength and guidance throughout my life.

This thesis is dedicated to the memory of my uncle Martin John, to whom I will forever be indebted to for being a positive role model and for supporting and caring for my family during my time away from home.

My eternal gratitude is owed to my parents Monica John and Jairam Bejai who have provided unconditional support, love and guidance in the pursuit of my goals and to my grandparents whose sacrifices have allowed me to pursue higher education.

I would like to extend a special thank you to Zoe Wilson, who over the past four years has served as my academic supervisor but also as my mentor and confidant.

My heart-felt gratitude must be paid to my co-supervisors Jacob Lage and Julie King, my family and friends, current and past members of the Wilson lab and the staff of KWS UK Ltd. who have supported and encouraged me throughout my studies.

This work was funded by an industrial scholarship from KWS UK Ltd.

Table of contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Description (units)
AD	Anthesis duration (days)
AE	Number of anthers extruded per floret
AL	Anther length (mm)
AM	Association mapping
ANOVA	Analysis of variance
AUFPC	Area under the flowering progress curve
AWNS	Awn presence/absence
BLUE	Best linear unbiased estimate
BLUP	Best linear unbiased prediction
CI	Confidence interval
CIM	Composite interval mapping
FS	Number of fertile shoots per m^2
GPE	Grains per ear/spike
GPS	Grains per spikelet
GS	Genomic selection
HCA	Hierarchical clustering analysis
HD	Days to heading from May 1st (days)
HT	Plant height (cm)
K	Kinship matrix
LD	Linkage disequilibrium
LM	Linkage mapping
MAS	Marker assisted selection
MLM	Mixed linear model
MQM	Multiple QTL mapping
MTA	Marker-trait association
PC	Principal component
PCA	Principal component analysis
QTL	Quantitative trait locus
REML	Restricted maximum likelihood
SIM	Single interval mapping
SNP	Single nucleotide polymorphism
SP	Number of spikelets per spike

Abstract

The modification of floral characteristics will be beneficial in improving the efficiency of hybrid seed production and the breeding of more climate resilient varieties in bread wheat (*Triticum aestivum*). Methods for phenotyping floral traits were initially tested using small genotype panels under controlled conditions and in field trials. Low-tech phenotyping methods appropriate for use by breeders and researchers were developed and demonstrated to be accurate. A panel of 111 genotypes was subsequently assessed in field trials using these methods. A high level of genotypic variation was observed for anther extrusion, anther length and anthesis duration/pattern and phenotypes were found to be stable across trials. Using this phenotypic data set, floral trait marker-trait associations (MTAs) were detected by association mapping and additional anther extrusion quantitative trait loci (QTLs) have been detected by linkage mapping in a bi-parental population. The phenotypic effects of candidate loci co-localizing with MTAs and QTLs were investigated using TILLING mutants and modifications to floral characteristics have been observed in some mutant lines. The utilization of phenotyping and genomic resources described in the present study is discussed and areas of future research have been identified.

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

1.1.1 The need for genetic improvement in hexaploid bread wheat (*Triticum aestivum*).

The challenge facing global agricultural systems is to drastically improve crop productivity in an environmentally sustainable manner to meet the demands of a rapidly expanding population. Recent estimates suggest that yield increases of 50-70% will be required by 2050 if dietary and industrial needs are to be met (Tester and Langridge, 2010, Licker et al., 2010).

Improvements to the productivity of wheat (*T. aestivum*) can make a substantial contribution towards achieving global food security; this is in part due to its capacity to adapt to a range of environmental conditions as well as the fact that our population relies on the species for approximately 20% of our calorific intake (Reynolds et al., 2012). The present rate at which yield gains in wheat are being realized, appears to be inadequate if production is to match predicted increases in demand (Reynolds et al., 2012). Coupled to this, annual reductions in the availability of arable land, increased land degradation and global climate change all emphasize the need for sustained investment towards improving the genetic potential of *T. aestivum*.

1.1.2 Potential role of floral biology in achieving yield gains

In past years, genetic improvement efforts targeting traits such as resource partitioning and disease resistance have led to significant improvements in yield (Reynolds et al., 2012). To date, comparable advances have yet to be realized in our understanding of floral traits and their role as determinants of grain yield.

An improved understanding of floral biology can aid efforts towards improving yield stability and improving the efficiency of hybrid seed production; an avenue for yield improvement that is being implemented by a number of breeding companies such as Bayer Cropscience (CPM, 2017), Syngenta (SeedWorld, 2015) and KWS (Jacob Lage, KWS UK Ltd., personal communication, August 2017).

Genotypic variation for pollen production, anther extrusion, pollen shedding and anthesis duration has been reported (Cahn, 1925, Heyne and Livers, 1967, Kherde et al., 1967, Wilson, 1968, Johnson and Schmidt, 1968, D'SouzA, 1970, Athwal and Kimber, 1970, Beri and Anand, 1971, De Vries, 1971, De Vries, 1972, De Vries, 1973, Khan et al., 1973, De Vries, 1974, Jôst et al., 1976). The effective exploitation of this diversity would necessitate that the physiological and genetic factors underlying phenotypic variation be dissected.

1.2 Wheat floral biology

1.2.1 Floral structure and development in wheat

The wheat ear or spike is an inflorescence (Figure 1.1.), composed of a number of spikelets located along the main axis (Kirby, 2002, De Vries, 1971). Each spikelet is composed of two to four fertile florets (Figure 1.2 A); each consisting of an outer (lemma) and internal (palea) glume (Figure 1.2 B). Typically one or more of the higher order (upper) florets are infertile (Kirby, 2002, De Vries, 1971).

Figure 1.1: Wheat inflorescence; *Triticum aestivum* cv. Cadenza. Horizontal bar, 5 cm.

The reproductive organs are found between the two glumes; three stamens, a single pistil and two lodicules (Figure 1.2 C and D). The lodicules are not directly involved in reproduction but facilitate the opening of the floret by swelling during anthesis (Kirby, 2002, De Vries, 1971). The stamens comprise a filament, which supports an anther at its top, in which the pollen grains are formed. The pistil is composed of an ovary with two stigmas are attached to its top; the stigmas take on a feathery appearance when receptive to pollen (Figure 1.2 D).

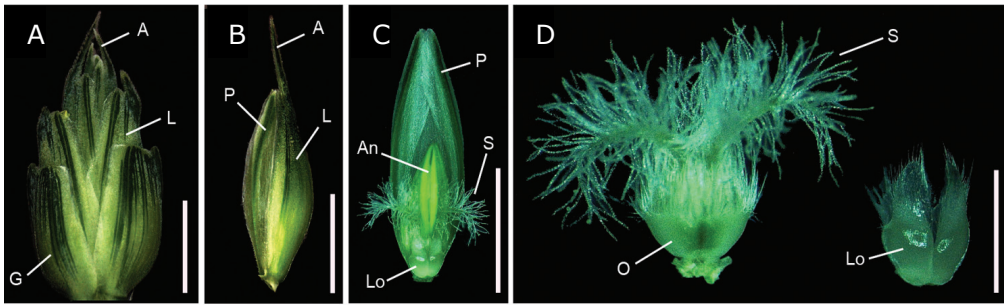

Figure 1.2: Structure of wheat flowers and spikes. **A)** Wheat spikelet; **B)** floret; **C)** palea and reproductive tissues; **D)** lodicule and reproductive tissues.
Bars, 5 mm (A-C); 2 mm (D). Abbreviations, A awn, An anther, G glume, L lemma, Lo Lodicule, O Ovary, P palea, S Stigma. Adapted from Whitford et al. (2013).

Subsequent to leaf initiation, the apex of a wheat tiller begins producing around 20 spikelet primordia, ending in a terminal spikelet (Kirby, 2002). The most developed spikelet primordia are found within the middle region of a spike (Kirby, 2002). After the double ridge stage, the structures of a spikelet are initiated; the glume primordia being first, followed by that of the florets (Kirby, 2002). Ten floret primordia are typically initiated, after which the spikelet apex begins to degenerate.

A developmental gradient therefore exists within the spikelet, with more mature florets being found at the base while the most distal florets are poorly developed. Within the spikelet, the lemma and palea are the first to form, followed by stamens and the carpel (Barnard, 1955, Williams, 1974).

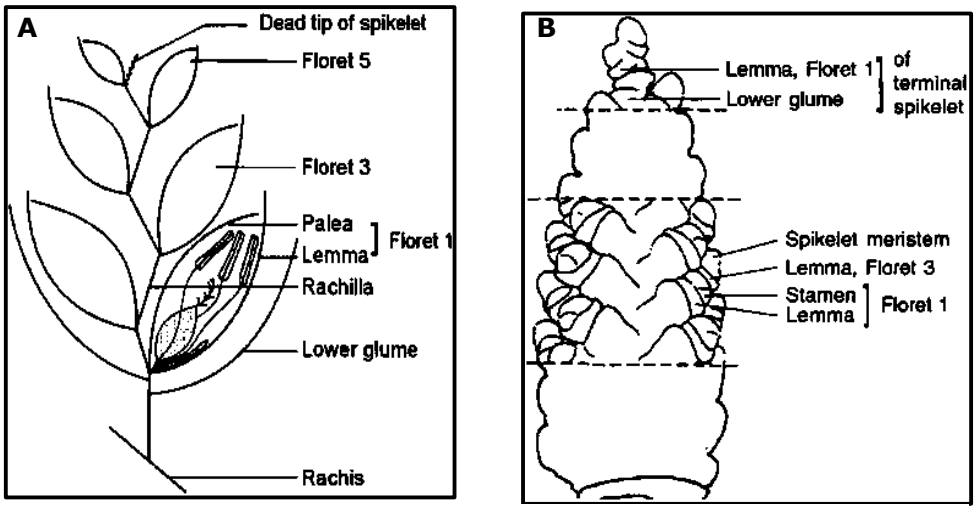

Figure 1.3: A) Diagram of a wheat spikelet. B) Shoot apex at the terminal spikelet stage.
Adapted from Kirby and Appleyard (Kirby and Appleyard, 1987)

1.2.2 The flowering process

Flowering is initiated in the primary florets within the middle third of a spike; flowering then progresses rapidly upwards, with the downward progression being notably slower (De Vries, 1971). During flowering the lodicules, swell due to an influx of water from the ovary, pushing the palea and lemma apart; with the glumes typically achieving maximum opening within five minutes (De Vries, 1971).

Filament elongation and anther extrusion typically occur simultaneous to floret opening; elongation is rapid, with three fold increases in length occurring within three minutes (Peterson, 1965). Stigmas begin to grow apart at this time, spreading out to take on a feathery appearance (De Vries, 1971). Flowering is initiated in the main tiller of a plant, with

subsequent tillers starting 3-4 days after and the entire process takes around ten days.

As emphasised by De Vries (1971), descriptions such as those provided here should only be considered as a framework. It is likely that there is genetic and environmental variation for each of these flowering processes, which ultimately will influence outcrossing potential.

1.3 Applications of floral biology research

1.3.1 Improving the efficiency of hybrid seed production

In recent years, there has been a sustained effort by the commercial sector to develop hybrid breeding programmes in wheat (CPM, 2017, SeedWorld, 2015). A number of recent studies into the extent of variation for floral characteristics and the genetic architecture of these traits support the feasibility of hybrid breeding as an avenue for yield improvement (Boeven et al., 2016, Muqaddasi et al., 2016, Whitford et al., 2013, Longin et al., 2013, Langer et al., 2014a, Muqaddasi et al., 2017). Additionally greater yield stability, which is needed in regions with a predisposition for abiotic stress events, and yield gains of as much as ten (10)

percent have been reported for hybrid lines (Longin et al., 2013).

Autogamous crops exhibit lower levels of heterosis as compared to allogamous crops such as maize and rye. The profitability of hybrid wheat breeding is therefore determined largely by the cost-efficiency of seed production (Langer et al., 2014a, Whitford et al., 2013).

Improved cost-efficiency can be achieved by reducing the number of male parental lines relative to the number of female lines. Taking into consideration the strong inbreeding characteristics of most varieties, it is likely that this would result in unpredictable and undesirable levels of seed set. Seed set is largely dependent on the extent of pollen grains shed external to the floret by pollinator lines (Wilson, 1968, Beri and Anand, 1971); the feasibility of this strategy would therefore require modification of the floral and flowering characteristics of wheat so as to promote outcrossing. Boeven et al. (2016) emphasized that the lack of suitable male parental lines is still a bottleneck in hybrid wheat breeding.

A theoretical pipeline for the development of a male and female parental pool for hybrid breeding is described in Figure 1.4. Subsequent to sourcing germplasm with adaptations to the target environment, and genetic characterization, decisions will need to be made as to which lines are potential

candidates for the male and female parental pools. One approach would be to initially identify lines possessing traits that would be anticipated to promote out-crossing potential; individuals meeting this criterion, that have a high degree of genetic similarity (Whitford et al., 2013) can become the founders of the male parental pool. Lines that are genetically distinct from these male founders can then be designated to the female parental pool, even if they have a high out-crossing potential. High throughput and accurate methods for phenotyping floral traits will undoubtedly be required at this stage to inform decision making.

Assessments of the general combining ability (GCA) of members of each pool to the other pool will be required (Akel et al., 2018); this can be achieved by the direct production of hybrids and subsequent yield testing. While the GCA of individuals in the male and female pools may be initially low, reciprocal recurrent selection (RRS) can be applied to improve GCA (Gilmore, 1964, Pandey and Gardner, 1992).

RRS entails selection based on performance of the bulked testcrosses from an individual line; this provides a measure of the combining ability of a male line, for example, with the female parental pool. With each successive cycle, application of marker-assisted selection (MAS) can ensure that out-crossing promoting traits are reinforced in the male pool (Xu

and Crouch, 2008); prerequisite for this would be the development and validation of molecular markers linked to variation in floral characteristics.

Over time, new lines found to have a good GCA with the female pool, can also be incorporated into the male pool. Application of MAS can help improve the out-crossing potential of these lines if initially low.

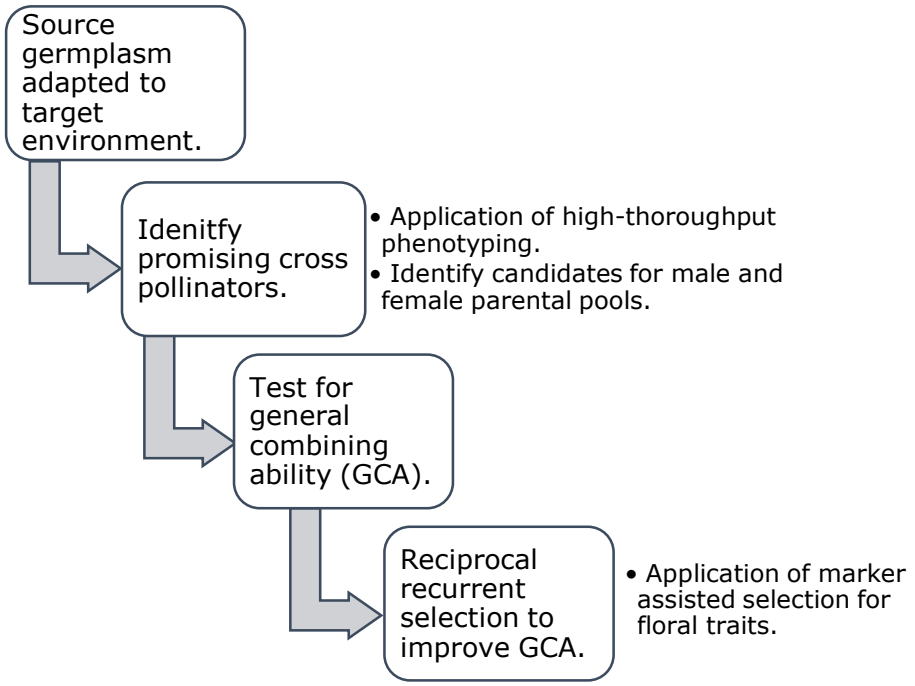

Figure 1.4: Schematic diagram showing a theoretical pipeline for the establishment of male and female parental pools for hybrid wheat breeding.

1.3.2 Improving abiotic stress tolerance: adaptation breeding

In light of predicted increases in the frequency of extreme weather events, wheat varieties with greater resilience to abiotic stresses will be required (Langer et al., 2014b, Lukac et al., 2012). In recent literature, an improved understanding of floral traits has been predominantly discussed in relation to improving hybrid seed production; however, work in other species has demonstrated that floral characteristics can mitigate against the effects of abiotic stress.

The production of high quantities of pollen has been described as a strategy to mitigate against the impact of environmental stress on seed set by some authors. In rice (*Oryza sativa* L.), Saito et al. (2001) reported longer anthers to be associated with improved cold tolerance. The authors proposed that longer anthers confer plants with the capacity to buffer against reductions in pollen viability due to cold stress, based on the assumption that longer anthers translate into a higher number of pollen grains produced per anther (Oka and Morishima, 1967). Further to this, Rang et al. (2011) demonstrated that a strong relationship exists between the number of pollen grains germinated on the stigma and pollen production under conditions of heat stress; demonstrating the value of higher pollen production.

Liu et al. (2006) reported genotypic variation for anther dehiscence and lower stigma pollen density in rice. Lower anther dehiscence associated with the onset of drought stress led to reductions in spikelet fertility of 80% in the variety IR64, while the variety Moroberekan exhibited a reduction of only 16%. The authors identified two traits in Moroberekan that they attributed to its higher stress tolerance; better development of the fibrous structures in the endothecium in the apex and base of the anther as well as a greater capacity to maintain the size of pollen grains. These floral characteristics; high pollen production, stable anther dehiscence and maintenance of pollen size all represent key targets that should be investigated in *T. aestivum*

The tailoring of flowering date to the local environment can help avoid periods of high temperature and drought stress during pollen formation and anthesis. These stages in male reproductive development have been found to be especially vulnerable to the onset of abiotic stress (Demotes-Mainard et al., 1995). For example, early flowering genotypes will be advantageous in climates prone to early summer drought (Zheng et al., 2012); while genotypes with asynchronous floret development between spikelets and tillers can mitigate against sterility induced by the onset of sudden adverse climatic events (Lukac et al., 2012, Rang et al., 2011).

1.4 Target floral traits

1.4.1 Pollen production and its relationship to anther size

With reference to the development of the male ideotype, a useful starting point would be to ascertain the level of genotypic variation for pollen production within existing germplasm. Prior investigations have reported genotypic variation in the number of pollen grains produced per anther in wheat; with Cahn (1925) reporting a range of 856-1380; Heynes and Livers (1967) 1200 to 1600 and Beri and Anand (1971),581 to 2153.

These numbers are substantially lower than that produced by cross pollinated cereal crops (Nguyen et al., 2015, Johnson and Schmidt, 1968). Nguyen et al. (2015) found the number of pollen grains per anther produced by wheat as being 15% of that produced by rye. Assuming 2500 pollen grains, per anther, three anthers per floret, two fertile florets per spikelet and 20 fertile spikelets per spike, De Vries (1971) described the pollen produced by a single wheat inflorescence as being 10% that of rye.

Manipulation of pollen production is not only of relevance to hybrid breeding, as it may provide a means of achieving more stable pollination and thereby more reliable seed set. In rice,

it has been demonstrated that cultivars producing greater quantities of pollen per anther are more tolerant of cool temperatures as they are capable of compensating for cold induced pollen sterility (Saito et al., 2001, Suzuki, 1981). Further, the value of higher pollen production in ensuring consistent grain yields across seasons has been demonstrated in commercial rye production by the POLLENPLUS® technology of KWS (KWS, 2017).

High throughput and cost effective methods for estimation of pollen production are lacking. Towards the development of the male ideotype for hybrid breeding, indirect selection for this trait using a more easily scored floral trait would be a practical approach. Past investigations have reported positive correlations between anther size and the number of pollen grains produced per anther (Nguyen et al., 2015, Beri and Anand, 1971, De Vries, 1974, Jôst et al., 1976).

Within the aforementioned studies, conclusions have been drawn from panels with a relatively low number of genotypes. It can be argued that establishing whether there exists a true genetic and/or physiological link between the two traits would require a large diverse assembly of genotypes. Alternatively, a bi-parental population generated from a variety exhibiting both high pollen production capacity and long anthers, with a

standard variety can help elucidate the extent of linkage between the two traits.

Work carried out by Nguyen et al. (2015) demonstrated that genetic factors capable of effecting phenotypic variation in anther length and pollen production in wheat, are present on chromosome four of rye (*Secale cereale*). The longest anthers within the panel of rye addition lines they investigated were observed in a line possessing chromosome four of rye (4R) and another possessing the long arm of the chromosome (4RL). The line possessing the short arm (4RS) line produced anthers marginally longer than the parental wheat variety (Nguyen et al., 2015).

Within their panel of addition lines, a strong positive correlation between anther length and the number of pollen grains per anther was observed (r=0.93). The 4RS line's anthers despite being shorter than those of the 4RL line produced a greater quantity of pollen, which was more comparable to that of the 4R line (Nguyen et al., 2015). This observation suggests that genetic factors introgressed from the long arm of rye chromosome four exert a larger influence anther length; while factors on the short arm influence the number of pollen grains produced per anther (Nguyen et al., 2015).

Within the literature, moderate to high heritability has been reported for anther length (Komaki and Tsunewaki, 1981, Jôst et al., 1976, Guo et al., 2015, Langer et al., 2014a). Across two years, Komaki and Tsunewaki (1981) observed minor intra-varietal variation in anther length, the largest being 0.27 mm, and suggested the heritability of the trait as being above 0.8. Guo et al. (2015) reported it as being above 0.68 for each of the four floret positions assessed at different developmental stages. Additionally, they reported a strong relationship between anther lengths at different floret positions within the genotypes investigated. These observations suggest that anther length is a fairly stable character under strong genetic control.

Jost and Jost (1976) reported that in Yugoslavian material, anther length was positively correlated with the length of the vegetative growth phase (r=0.57); similarly Komaki and Tsunewaki (1981) reported anther length as being lower in early heading Japanese varieties. This appears to be a physiological link as opposed to genetic; while Komaki and Tsunewaki (1981) observed a positive correlation between heading date and anther length in the F1 families investigated, this correlation was largely absent in F2 families.

1.4.2 Anther extrusion

Anther extrusion (AE) is a function of two processes mentioned previously; floret opening and filament elongation. In recent years AE has garnered significant attention due to its relevance to hybrid breeding (Langer et al., 2014a, Muqaddasi et al., 2016) and its relationship to biotic stress tolerance, with special reference being made to *Fusarium* head blight (Skinnes et al., 2010). The possibility that anther extrusion can stabilize seed set by mitigating against loss of viable pollen in earlier flowering florets via intra-spike cross fertilization has also been proposed (Lukac et al., 2012).

Wheat is considered as a cleistogamous species, with the majority of pollen being shed within the floret before or just after floret opening. Authors have suggested this characteristic may have begun to develop during the domestication process (D'SouzA, 1970) and appears to have been reinforced in past decades (Boeven et al., 2016). Boeven et al. (2016) reported a trend towards more cleistogamous varieties being released in recent years, the authors speculate this may be due to the high level of homogeneity required for variety registration.

The species' potential for outcrossing should not be dismissed; genotype level variation has previously been reported for anther extrusion capacity (Boeven et al., 2016,

Muqaddasi et al., 2017, Muqaddasi et al., 2016, Langer et al., 2014a). Additionally, genotypic variation has been described for at least two of the processes that influence the trait; filament extension (Beri and Anand, 1971) and the extent of floret opening (Johnson and Schmidt, 1968, De Vries, 1971).

Mechanical barriers imposed by spike architecture influence the extent of floret gaping achievable by some varieties. Lax headed varieties, in which spikelets are less densely packed along the spike, have been noted as exhibiting greater separation angles between the palea and lemma (Johnson and Schmidt, 1968). Additionally, the position of a floret within a spikelet influences its capacity for opening; De Vries (1971) noted that inner florets had a greater tendency to be cleistogamous due to the barriers imposed by outer florets.

With reference to filament elongation, relationships between plant height and filament length have been reported in wheat. Beri and Anand (1971) observed that WI202, possessing three dwarfing genes, exhibited an average filament length of 4.5mm; conversely C273 and C306 which possessed no dwarfing genes had an average of 11.99mm. In line with these observations, Lu et al. (2013) reported that the *Rht-B1* dwarfing allele co-localizes with an anther retention QTL. Boeven et al. (2016) and Langer et al. (2014a) both reported positive, albeit weak, correlations between plant height and

anther extrusion; with Boeven et al. (2016) reporting *Rht-D1* as showing the strongest association with anther extrusion within their association panel.

Early investigations have reported that the extent of anther extrusion is influenced by meteorological conditions; De Vries et al. (1971) reported that low temperatures and rainfall increased the number of cleistogamous florets. Despite variation attributable to environmental conditions, varietal characteristics appeared to be maintained (De Vries, 1971). A number of recent studies have reported a high trait heritability for anther extrusion; 0.84- 0.91 (Boeven et al., 2016, Muqaddasi et al., 2017, Muqaddasi et al., 2016, Langer et al., 2014a).

These observations support the feasibility of selection for higher anther extrusion within breeding pools. With reference to screening a large number of lines, harvesting of ears at the end of anthesis and storage for later determination of the numbers of retained/extruded anthers, either for all florets within a spike or a predetermined number of florets, appears to be feasible (Muqaddasi et al., 2016, Langer et al., 2014a, Muqaddasi et al., 2017). As opposed to visual anther extrusion scoring, this approach would help to avoid biases arising from environmental conditions on the day of scoring.

It should be noted that although lower than that obtained by counting anthers, Boeven et al. (2016) and Langer et al. (2014a) observed a high heritability for visually scored anther extrusion; 0.71 and 0.79 respectively. It can be argued that in breeding programs selection via visual anther extrusion may be appropriate while for detailed phenotypic and genetic studies, such as QTL mapping, counting of anthers will be required.

1.4.3 Pollen shedding capacity

With specific reference to hybrid wheat breeding, it is clear that traits should not be considered in isolation from each other. The amount of pollen shed external to the wheat inflorescence is in effect a function of each of these traits. Beri and Anand (1971) emphasize that if an efficient male parental lines is to be developed, several traits must selected for simultaneously. Despite producing the lowest number of pollen grains per anther, two varieties WL95 and S331 were capable of achieving levels of external pollen shedding that were greater than varieties producing more pollen, in part attributable to their longer filaments (Beri and Anand, 1971). Similarly, Kherde et al. (1967) reported that selecting for larger anthers within the male pollinator pool did not translate into improved seed set on male sterile lines within a hybrid breeding scheme.

Within the literature, the strength of anther filaments and the anther dehiscence have not been discussed extensively as targets for modifying the pollen shedding capacity of wheat varieties. Theoretically, if filaments capable of suspending extruded anthers for a long period are coupled with anthers that do not dehisce entirely along their length, this enhancing of cross-fertilization capacity could benefit hybrid seed production. Conversely, researchers have proposed that long anther basal dehiscence promotes self-pollination, and therein more reliable seed set, under conditions that compromise pollination such as the onset of heat stress (Zhao et al., 2016).

Apart from Langer et al. (2014a), who attempted to indirectly assess the size of the dehiscence zones through scoring anther colour, no investigations targeting these traits have been carried out in wheat. In rice however, clear genotypic variation has been observed for the length of dehiscence in the basal and apical parts of the anther thecae and the proportion of dehisced anthers (Das et al., 2014, Zhao et al., 2016). While these two dehiscence traits can be relatively easily assessed in wheat germplasm, needing only a steromicroscope, filament strength will present a greater challenge to researchers.

Assessing pollen shedding capacity under field conditions may be achieved via the use of adhesive surfaces suspended below the crop or by bagging ears and subsequently measuring the mass of pollen shed. While the adhesive surfaces may provide a more reliable estimate of pollen shedding capacity, bagging of ears may be more time efficient and less laborious and therefore more amendable to high throughput phenotyping. As the optimal conditions for pollen shedding are determined by temperature, relative humidity and solar radiation; all of which are likely to be affected by the bagging of ears (Langer et al., 2014a) further evaluation of this method is required.

Boeven et al. (2016) and Langer et al. (2014a) both reported pollen mass as having a relatively low heritability compared to other floral traits assessed; likely due to the trait assessment being more error prone. Validation of the method would require the simultaneous application of the pollen mass and adhesive surface methods on a relatively small number of genotypes.

1.4.4 Flowering time

Loci governing flowering time in wheat can facilitate 1) the fine-tuning of flowering to a particular environment to avoid suboptimal conditions and 2) synchronization of flowering between male and female parental lines in hybrid breeding. Flowering time in wheat is regulated by three signalling

pathways; the vernalization (*Vrn*), photoperiod (*Ppd*) and earliness *per se* (*Eps*) pathways (Langer et al., 2014b, Kamran et al., 2014).

Vrn genes determine winter or spring growth habit as they govern the transition from the vegetative to the reproductive phase. With respect to autumn sown spring and winter wheat varieties, the *Vrn* pathways does not exert an influence on flowering time as vernalization requirements are typically met. In such instances, flowering time is instead determined predominantly by photoperiod sensitivity/insensitivity.

Photoperiod insensitive alleles of the *Ppd* genes, present on the group two chromosomes, allow for flowering irrespective of day length (Langer et al., 2014b, Kamran et al., 2014). The relative influence of each of the homoeologus loci varies, with *Ppd-D1* being described as have the greatest influence (Kamran et al., 2014). Within European germplasm, *Ppd-D1* is described as the primary determinant of flowering time (Langer et al., 2014b, Zanke et al., 2014).

Photoperiod insensitive genotypes are capable of transitioning into reproductive development as soon as temperatures increase during spring; this would be advantageous in southern latitudes as a means to avoid the onset of drought and high temperatures in late summer. Less of an advantage would be conferred in northern regions, where associated

reductions in the duration of the vegetative growth phase can translate into lower yield potentials.

With regard to synchronizing flowering in hybrid breeding schemes, loci with smaller effects may be more applicable to fine tuning flowering time. The *Eps* pathway collectively refers to loci that influence flowering independently of the *Vrn* and *Ppd* pathways. A number of *Eps* loci have been identified in past QTL studies (Kamran et al., 2014, Langer et al., 2014b, Zanke et al., 2014, Pánková et al., 2008, Griffiths et al., 2009). These loci typically differ in flowering time by a few days as compared to weeks as is typical for variation in the photoperiod sensitivity loci (Langer et al., 2014b).

1.4.5 Pollen viability

The viability of pollen grains is inferred from the level of enzymatic activity within the grain at the point of examination; from this the fertilization capacity of the grains is inferred (Dafni and Firmage, 2000). The longevity of pollen grains, i.e. how long viability is retained, can have practical implications for hybrid seed production as well as abiotic stress tolerance.

The results of past investigations suggest that loss of pollen viability, and ultimately fertilization capacity, should not be viewed as a limiting factor to hybrid seed production (Fritz and Lukaszewski, 1989, De Vries, 1971). Fritz and

Lukaszweski (1989) reported that seed set was achievable using pollen grains stored up to 45 minutes under ambient conditions; leading the authors to assert that the longevity of wheat pollen in germplasm existing at that time was ample to ensure fertilization of female parental lines within the conditions of hybrid breeding schemes.

The emphasis on the need to reduce the number of male parental lines in relation to female parental lines, to achieve cost effectiveness, indicates that future hybrid breeding programmes will differ from those present at the time of Fritz and Lukaszweski's work; in such a scenario there will be a need for pollen grains to travel further. This coupled to reports that intra-ear fertilization can potentially mitigate against abiotic stress induced sterility in earlier flowering florets (Lukac et al., 2012) suggests that pollen viability in wheat and the extent of genotypic variation for this trait in existing germplasm should be revisited.

Approaches towards assessing pollen longevity include hand pollinations and the application of chemical assays such as the fluorescein diacetate (FDA) test; results of which have been found to be correlated with results of hand pollination assessments (Fritz and Lukaszewski, 1989, De Vries, 1971).

Pollen viability assessments are however time consuming and require some degree of technical skill. For these reasons,

assessment of pollen viability and longevity may not be practical in the initial stages of a breeding cycle, which are typically characterized by large population sizes. Efforts by researcher towards dissecting the genetic factors underlying the trait can aid the effective exploitation of any existing genotypic variation within breeding programmes.

1.5 Functional genomics and molecular breeding

The application of DNA marker technologies in plant breeding is termed marker assisted selection (MAS) which falls within the wider discipline of 'molecular breeding' (Collard and Mackill, 2008). MAS is based on the principle that DNA markers tightly linked to loci that underlie phenotypic variation can be identified and subsequently used to track the inheritance of desirable traits within breeding populations. MAS has been described as having the potential to improve the efficiency and precision of plant breeding (Collard and Mackill, 2008).

MAS will be feasible, in terms of both cost and trait improvement over time, provided that markers linked to QTLs with large phenotypic effects can be identified (Nakaya and Isobe, 2012). Further, due to the cost investment required the application of MAS will be more practical for traits that are

time consuming and impractical to assess in early breeding stages.

To date, evaluated methods for phenotyping the majority of floral traits are both time and labour demanding (Boeven et al., 2016, Langer et al., 2014a). This led Langer et al. (2014a) to propose a pyramid breeding approach, in which traits requiring the greatest investment in time and labour are screened for in the later selection stages of the breeding cycle. Application of MAS will be more cost-effective at this stage due to smaller population sizes and can help improve selection accuracy for traits with complex architectures and those that are more responsive to environmental conditions.

For some traits, such as anther extrusion, phenotyping methods described previously appear to provide a reasonable level of accuracy without requiring a significant investment in time or cost (Langer et al., 2014a). It can therefore be argued that due to cost investment required for MAS, phenotypic selection is a more practical approach towards improving a trait such as anther extrusion (Boeven et al., 2016).

With reference to hybrid breeding, genetic markers tagging anther extrusion loci may instead be of value for genomic selection (GS) assisted recurrent selection in the developmental of male parental pools (Boeven et al., 2016). The incorporation of genome-wide marker data together with

phenotypic data from a training population in GS enables breeders to predict the performance of a population for which only genotypic data is provided, the test population (Meuwissen et al., 2001). GS can provide an approach to ensure that out-crossing potential is reinforced during the recurrent selection cycles required to enhance the general combing ability of male parental pools. The work of Boeven et al. (2016) provides support for this approach; with a prediction accuracy of 0.7 being obtained by incorporating the anther extrusion QTLs detected in their study into GS.

The detection of reliable markers that are tightly linked to the loci governing traits of interest is a pre-requisite for the development of a MAS breeding programme. The identification of such markers can be achieved either through the mapping of quantitative trait loci (QTLs) by linkage or association mapping (Crossett et al., 2010, Yu and Buckler, 2006). QTL mappings studies aim to locate genetic factors governing a trait of interest to positions within the genome, while simultaneously ascertaining the relative influence of allelic variation at identified loci.

Low marker coverage and density, along with populations characterised by a relatively low number of recombination events per line, have imposed major constraints on the resolution of QTL studies in the past (Crossett et al., 2010).

Significant advances have been made towards the development of populations with higher rates of recombination; such as inter-mated recombinant inbred line (iRIL) populations in maize (Crossett et al., 2010). Concurrent to this, genetic marker density and coverage has significantly improved for the majority of crop species with the advent of SNP genotyping platforms and the development of transcript derived markers (Crossett et al., 2010). Crossett et al. (2010) emphasized that the challenge facing statisticians is to now the develop statistical methods that can accommodate these rapid gains in QTL resolution.

1.5.1 Linkage mapping

Interval mapping for single QTLs (SIM) entails performing a regression of the phenotypic difference between the homozygous classes at each marker loci (Zeng, 1994). The putative position of a QTL is moved along a chromosome, with SIM testing every two centimorgans (cM) between pairs of adjacent markers, calculating a logarithm of the odds (LOD) score during each test. LOD scores are then plotted along the chromosome being tested to give a LOD curve or profile; those exceeding a predetermined significance threshold are taken as indicative of the presence of a QTL in that region.

SIM as implemented in the Haley-Knott method however has inherent limitations.

If two QTLs are linked in coupling, SIM is likely to detect a single large 'ghost' QTL between the two. Conversely if QTL are linked in repulsion, they may remain undetected. Composite Interval Mapping (CIM) aims to minimize the influence of these effects on the detection of true QTLs through the simultaneous implementation of interval mapping and multiple regression (Jansen and Stam, 1994). The reductions in the power due to QTLs segregating elsewhere in the genome can be mitigated through introduction of covariates into interval mapping analysis. This is the principle on which CIM is based.

LOD profiles can vary considerably between CIM and SIM. The extent of these differences being influenced by the number of selected marker covariates, the minimum distance from a test site for the inclusion of covariates into the model and the threshold of statistical significance (Crossett et al., 2010). With reference to the the minimum distance for marker covariate inclusion, drastic changes in the shape of LOD curves can arise from inclusion of marker covariates that are genetically linked, i.e. in high linkage disequilibrium, to the test position. In such a case, changes in LOD profiles arise from changes in user defined parameters as opposed to actual

changes in the probability that a true QTL exists at the test site (Crossett et al., 2010).

Software packages such as QTL Cartographer and 'rqtl' provide a forward selection method for identification of marker covariates for use in CIM. A multiple QTL model is constructed by first incorporating the position with the highest LOD score from SIM, then the position with the second highest LOD score and so on. The positions that remain significant when fitted with the previously included QTLs are retained in the model.

CIM still only searches for a single QTL at a time; a two locus equivalent of interval mapping to simultaneously scan every pair of intervals. This can be implemented in 'rqtl' via stepwise selection for multiple QTLs (MQM); a forward/backward selection method is used to construct a multiple QTL model with the final model being chosen based on a penalized LOD score. The penalized LOD score is derived based on the number of QTLs and QTL interactions. MQM has been described by its developers as having the following advantages; higher power and precision in QTL detection, protection against overfitting, prevention of ghost QTL detection and detection of negating QTLs (Arends et al., 2014). In wheat, MQM has previously been demonstrated to be a powerful tool for QTL detection; facilitating the detection

of major resistance QTLs for Fusarium head blight (Löffler et al., 2009) and crown rot (Ma et al., 2010) as well as the discovery of a frost tolerance gene on chromosome 5B (Tóth et al., 2003).

1.5.2 Association mapping

Linkage mapping and causative gene identification within bi-parental populations is inherently limited due to restricted allelic diversity and low genomic resolution (Brachi et al., 2011). Association mapping (AM), a technique originally developed for use in human epidemiological studies has garnered significant attention by crop scientists in the past decade as its potential to overcome the limitations of linkage mapping has been recognized.

The adoption of AM in crop studies has been accelerated by the advent of high-density single-nucleotide polymorphism (SNP) genotyping platforms which allow for whole-genome scans to identify genetic variations associated with phenotypic variation (Brachi et al., 2011). AM has been demonstrated to be capable of detecting robust loci that exhibit effects across genetic backgrounds (Jannink, 2007, Brachi et al., 2011).

AM is based on the premise that due to recombination events within the genetic history of an association mapping panel, only markers that are tightly linked, i.e. in high linkage

disequilibrium (LD) with causative loci, will be found to be significantly associated with phenotypic variation (Platt et al., 2010).

In AM, the discovery of a statistically significant association is considered as proof of linkage between phenotype and variation at a causal site; this assumes that LD decays rapidly with genetic distance within the population being investigated (Pritchard and Rosenberg, 1999). The extent of LD within a population is however influenced by a number of factors including the level of inbreeding, selection for favourable alleles, the domestication process, as well as the presence of introgressed chromosomal regions from related species (Chao et al., 2010).

1.5.2.1 Association mapping compared to linkage mapping

AM can provide a more cost effective approach towards identifying QTLs. AM circumvents the need for the development of large experimental populations as it can be applied to a panel assembled from existing germplasm (Atwell et al., 2010). Coupled to this, many traits can be investigated simultaneously (Atwell et al., 2010) and historical phenotype data can be used; reducing genotyping and phenotyping costs respectively.

As compared to a bi-parental mapping population, AM panels are characterized by a relatively higher level of recombination events. This arises from the large number of meiotic crossovers that would have occurred during the development history of each variety. Having a wider genetic background coupled to broader genetic variation also facilitates the simultaneous evaluation of many alleles at an individual locus. These factors led Yu and Buckler (2006) to describe AM as more likely to detect markers that would be of use in MAS.

Linkage mapping does still offer some advantages over currently available AM methods. AM studies can often result in false associations due to the presence of population structure and individual relatedness (kinship) within the panel. Additionally, AM is influenced by allele frequency distribution, with reductions in power to detect true associations occurring once minor allele frequencies fall below a critical threshold (Yu et al., 2006, Brachi et al., 2011). Thus causative rare alleles may not necessarily be identified.

Within an AM panel variation is anticipated for phenological characteristics, such as plant height at maturity or flowering time. Variation in these traits can make it difficult to identify genetic markers consistently associated with variation in traits, especially so for those with a complex genetic architecture such as yield (Lopes et al., 2015, Dodig et al.,

2012). In wheat AM studies, the confounding effects of phenology are known to impose a constraint on the identification adaptive trait loci. As these traits typically interact with flowering time and plant height, loci with major effects can effectively mask minor effect genes (Reynolds et al., 2009, Rebetzke et al., 2013).

1.5.2.2 Linkage disequilibrium in wheat

Prior information on the extent of LD can inform the design of AM studies, specifically with regard to the marker density required in order to improve statistical power (Chao et al., 2010). Previous studies conducted in wheat show that LD varies not only between genomes and chromosomes but also within individual chromosomes (Würschum et al., 2013); this would suggest that as many polymorphic markers as available should be utilized in AM in order to ensure adequate coverage of the genomes.

Estimates for the rate of LD decay in wheat vary. Würschum et al. (2013) observed significant LD to decay at 5-10 cM within their panel of elite European germplasm. The authors concluded that LD decay in wheat appears to be slow, as compared to that observed for maize and suggest that this may limit the resolution of AM in wheat. Historical LD is noted

as being considerably lower for outcrossing species such as maize; the repeated selfing that occurs in inbreeding species such as wheat, results in the maintenance of homozygosity, in turn rendering recombination events ineffective at reducing LD (Würschum et al., 2013).

Chao et al. (2010) reported significant genome wide LD to to extend to 19.2 cM, similarly Bentley et al. (2014) reported significant LD as extending to 20 cM. LD typically decays more slowly in the D genome (Mackay et al., 2014, Akhunov et al., 2010, Chao et al., 2007, Chao et al., 2010). Chao et al. (2010) hypothesize this as being an artefact of the genetic bottleneck that accompanied the speciation of hexaploid wheat. Despite the relatively slow decay of LD, simulations conducted by Bentley et al. (2014) demonstrate that adequate statistical power to detect QTLs can be achieved.

1.5.2.3 False positives in association mapping

AM studies can fail through detection of non-causal loci as linked to trait variation, these are termed false positives. These false positives can arise firstly by patterns being present within a panel that are not representative of the larger population; this type of false positive can be effectively managed by proper sampling theory, adequate replication and incorporation of appropriate error terms within the statistical model being used (Larsson et al., 2013).

Correlation among loci and the factors that are responsible for trait variation can also result in false positives. These indirect associations are more difficult to control for, as increasing sampling size and marker density will only serve to increase the significance of these non-causal associations (Larsson et al., 2013).

The applicability of AM in crop species was initially assessed by Thornsbery et al. (2001); while the authors concluded that AM offered greater resolution than traditional bi-parental linkage mapping, it became evident that adequate control for false positives was prerequisite. Subsequent studies revealed that the associations between *Dwarf8* (*d8*) and flowering time in maize that were initially reported were likely false positives, with the effect of the loci being overestimated due to

inadequate correction for population structure (Larsson et al., 2013, Andersen et al., 2005).

Long range LD i.e. genome wide correlations between loci, is an artefact of population structure which in turn leads to detection of non-causative associations (Zhao et al., 2007, Platt et al., 2010). One approach to correct for population structure, that has been widely implemented, is principal component analysis (PCA) (Price et al., 2006); wherein variation observed across markers is summarized into a smaller number of component variables. Each PC can be thought of as a separate unobserved subpopulation from which individuals originate (Larsson et al., 2013); with the loadings of each individual describing how related they are to each subpopulation. PCA, along with its alternative structured association (SA), is inherently limited in that it corrects based on broad patterns of genetic variation within the population and may not adequately account for the relatedness of individuals (Larsson et al., 2013).

The unified mixed linear model (MLM), also known as the Q+K model was developed in order to address the limitations of PCA and SA based association studies (Yu et al., 2006); by correcting simultaneously for population structure and less apparent familial relatedness. Covariates accounting for population structure are included as fixed effect (Q) while

individuals within the panel are treated as random effects within the model. Additionally, a kinship matrix (K) is incorporated, an estimate of the variance-covariance between individuals. Kang et al. (2010) described kinship matrices as removing "the effect of a large number of unlinked selectively neutral factors each with uninterestingly small effects".

1.5.2.4 Population structure in wheat AM studies

The consensus is that within elite European winter wheat germplasm there is only a small extent of population structure; leading some authors to not include covariates for population structure in AM (Reif et al., 2011, Würschum et al., 2013, Bentley et al., 2014). Reif et al. (2011) attribute this to the implementation of line breeding in wheat improvement, the constant exchange of germplasm between breeding programmes and the absence of distinct genetic pools.

Reif et al. (2011) reported PC1 and PC2 as each accounting for approximately 6% of variation within the panel, while Würschum et al. (2013) reported PC1 and PC2 as accounting for 5.1% and 4.3% respectively. Bentley et al. (2014) observed a greater level of underlying population structure 8.6% and 3.9% for PC1 and PC2 respectively, in this instance related to country of origin.

1.5.2.5 Trait assessments towards MTA discovery

In principle, the higher the heritability of a trait the greater the likelihood of detecting a significant marker-trait association (MTA) in AM (Myles et al., 2009). The number of components influencing a trait, the accuracy at which the trait can be assessed and the relative importance of genotype by environment interaction all influence heritability and in turn the capacity to detect MTAs.

With complex traits, a number of associations of low statistical significance are likely to be detected with AM. Authors have demonstrated the value of dissecting phenotype into several component traits whereby fewer associations are detected but with greater statistical significance (Kloth et al., 2012).

Accuracy can be achieved through experimental replications. Multiple replications within a site are typically required in order to minimize environmental noise and measurement error. Replication across multiple environments and multiple years can additionally provide insight into the importance of the genotype by environment interaction and the stability of a phenotype. Replicated data can then be used to generate phenotype estimates, less biased by the environmental variable and measurement errors; best linear unbiased predictor (BLUPs) and best linear unbiased estimator (BLUEs).

1.6. Aims and objectives

The present study aimed to improve the understanding of the genetic and physiological factors that underlie variation in key floral traits, namely anther extrusion, anther length, pollen production per anther and anthesis duration/pattern. The findings were anticipated to provide useful starting points for the development of tools that would be valuable in breeding floral trait ideotypes.

A more detailed list of the specific objectives is listed below:

1. To evaluate the accuracy and time/labour demands of phenotyping methods for floral traits by application within small genotype panels in both controlled environment and field conditions (Chapter 2).
2. To assess the extent of variation for floral characteristics existing in European germplasm, while simultaneously investigating how robust phenotypes are to environmental variation, by carrying out multi-year and multi-location field trials (Chapter 3).
3. To identify genomic regions associated with phenotypic variation in floral characteristics by conducting an association mapping study using field data collected for objective 2 (Chapter 4).

4. To identify anther extrusion quantitative trait loci (QTLs) by carrying out anther extrusion phenotyping and subsequent linkage mapping within a bi-parental population (Chapter 5).
5. To ascertain whether phenotypic effects arise from mutations in candidate floral traits genes, identified within the genomic regions found to be associated with floral traits in Objectives 4 & 5, through phenotyping TILLING mutant lines (Chapter 6).

1.7. Key hypotheses

1. The time and labour investment required for phenotyping will vary between floral traits (Chapter 2).
2. Floral traits can be accurately phenotyped under both controlled environment and field conditions using low technology approaches (Chapter 2).
3. Genotypic variation for floral traits is present within existing European germplasm and will be observed within the field panel (Chapter 3).
4. Trait stability across years and locations will vary between floral traits (Chapter 3).

5. Robust floral trait loci capable of exhibiting phenotypic effects across diverse genetic backgrounds can be detected by association mapping in the field panel (Chapter 4).
6. Clear association peaks will be observed in association mapping for traits anticipated to have a simple genetic architecture. Conversely, traits expected to have a complex architecture will yield no clear association peaks (Chapter 4).
7. Attributable to more population specific effects, genomic regions associated with anther extrusion not detected by association mapping, will be detected by linkage mapping in the bi-parental population (Chapter 5).
8. Phenotypic variation in floral characteristics will be observed in the TILLING mutants identified as possessing mutations in candidate floral trait genes (Chapter 6).

CHAPTER 2: VALIDATION OF METHODS FOR PHENOTYPING FLORAL TRAITS IN WHEAT

2.1 Introduction

Initial research into floral trait variation in *Triticum aestivum* aimed to provide insight into the feasibility of hybrid breeding; an avenue for yield improvement that had not been previously explored for the species. Despite a large extent of genotypic variation having been reported for traits that influence outcrossing ability, namely pollen production, anthesis duration, anther extrusion and pollen shedding external to the spike (Cahn, 1925, Heyne and Livers, 1967, Kherde et al., 1967, Wilson, 1968, Johnson and Schmidt, 1968, D'SouzA, 1970, Athwal and Kimber, 1970, Beri and Anand, 1971, De Vries, 1971, De Vries, 1972, De Vries, 1973, Khan et al., 1973, De Vries, 1974, Jôst et al., 1976) interest in hybrid breeding subsided subsequent to the 1980s. Until the very recent past, floral traits have received little attention from scientific and industrial researchers, compared to traits such as resource partitioning and disease resistance.

As a consequence, few phenotyping approaches for floral traits have been described in the literature. While recent field studies have successfully applied methods to assess anther extrusion that were described in earlier work (Langer et al., 2014a, Boeven et al., 2016, Muqaddasi et al., 2016, Muqaddasi et al., 2017), the feasibility of applying approaches

for other target traits has not yet been determined, neither has the prospect for developing novel approaches been explored.

In the present study, methods for phenotyping floral traits were initially applied to small genotype panels under both controlled and field conditions, with the aim of ascertaining how labour intensive, time demanding and accurate each approach was. Traits targeted are listed below:

1. Anthesis duration and pattern
2. Anther length
3. The number of anthers extruded per floret
4. The number of pollen grains per anther
5. The number of pollen grains shed external to florets

From the results obtained, the reliability, accuracy and feasibility of application is discussed for each phenotyping method. The extent of genotypic variation for floral traits observed in this pilot study is also described to support the feasibility of selecting for these traits within breeding pools.

2.2 Methods

All datasets described in the present and subsequent chapters were collected by Carus John-Bejai (CJB) with the exception of data obtained from a preliminary field trial in 2014 (Section 2.2.1). Data from this trial was collected by staff attached to KWS UK Ltd and made available to CJB, who carried out all subsequent analysis on the dataset.

2.2.1 Preliminary field trial

During the summer of 2013, a panel of 445 European winter wheat varieties (*Triticum aestivum*) were phenotyped for anthesis duration and anthesis pattern as well as for anther extrusion capacity. Field trials and phenotyping were carried out in Cambridgeshire, UK at KWS UK Ltd.'s main breeding site by KWS' staff. Each genotype was represented by a one square metre plot within a completely randomized design.

2.2.1.1 Flowering scores

The percentage of spikes within a plot showing visible signs of anthesis was estimated every second day, between 0900 h and 1300 h, once approximately ten percent of spikes were at anthesis. Assessments were continued until the end of anthesis, which was take as the point at which only ten percent of spikes remained at anthesis. The daily score

obtained for a plot is hereafter referred to as its flowering score.

2.2.1.2 Visual anther extrusion scores

Anther extrusion was visually scored using a one (1) to five (5) scale at the peak of anthesis for an individual plot as follows:

- A score of one (1) was assigned to genotypes for which no extruded anthers could be observed; any anthers trapped between the palea and lemma were deemed as non-extruded. The retention of anthers within florets was confirmed by opening several florets within an ear.
- A score of five (5) was assigned to genotypes that appeared to have extruded 90%, or more, of their anthers; with six or more anthers having been extruded at each spikelet level.
- An intermediate score of 2.5 was assigned to genotypes that appeared to have extruded between 25% and 50% or their anthers.

2.2.1.3 Hierarchical clustering of flowering score data

The flowering score data matrix provided by KWS UK Ltd. was analysed using hierarchical clustering (HCA), within the R software environment using the R-package 'ape'. HCA was

carried out to group genotypes based on the initiation, duration and pattern of anthesis within whole plots. The complete link method was used for HCA; wherein each genotype is initially treated as belonging to an individual cluster, the cluster pair with the greatest similarity is first combined and the process repeated until n clusters are generated. The optimal value of n was visually determined; n=15 was found to give the best separation of genotypes into groups.

2.2.2 Initial evaluation of phenotyping methods for anther length and pollen production

2.2.2.1 The number of pollen grains per anther

The effectiveness of water and ethanol extraction of pollen grains from dehiscing anthers was tested using glasshouse grown material. Sampling regions were first defined as follows; the top and bottom regions were taken as the first two fully formed spikelets at each respective end of the spike, while the two central florets between these two regions was defined as the middle region (Figure 2.1 A).

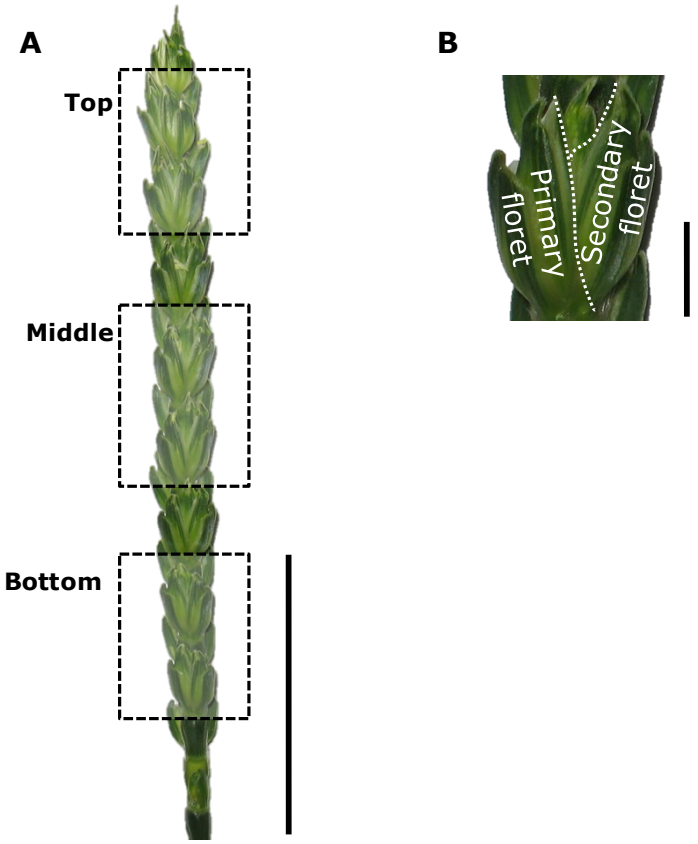

Figure 2.1: **A)** Photograph of wheat spike with sampling regions defined by dashed boxes. **B)** Photograph of wheat spikelet showing primary and secondary florets.

A) Vertical bar, 5 cm; **B)** Vertical bar, 1 cm.

Three anthers were collected immediately prior to anthesis from three biological replicates of *T. aestivum* cv. Paragon, with sampling restricted to primary florets (Figure 2.1 B) within the middle region of a spike. Anthers were placed into an empty 0.5ml Eppendorf tube and rolled gently along the side of the tube with a pair of forceps to open them along the lines of dehiscence.

A drop of fuchsine dye was added to each tube and the volume made up to 0.5ml either using 50:50 ethanol: purified water or purified water. Tubes were vortexed for 15 minutes and an aliquot of 2.5μl from the resulting suspension was placed onto a microscope slide. The number of pollen grains within the drop (n) was used to estimate the total number of pollen grains per anther as follows; (n*(500/2.5))/3.

2.2.2.2 Anther length

Plant material used in Section 2.2.2.1 was also used to assess the effect of storing anthers in 3:1 ethanol/glacial acetic acid. Three anthers were collected from ten replicate plants and immediately stored in the fixative. Sampling was restricted to yellow/green anthers from primary/secondary florets with a branched stigma, within the middle region of the ear (Figure 2.2).

Average anther length of the pooled anther samples was determined daily for five days. Anthers were photographed next to a scale bar under a stereomicroscope and their length measured using the image processing and analysis software ImageJ (National Institute of Health, Maryland, USA). A calibration was first set using the scale within each image and a segmented line was fitted along one of the two anther lobes; the length of this line was taken as the length of the anther (Figure 2.2).

Figure 2.2: Drawing of a wheat anther; the length of the solid black line along the bottom lobe is taken as the length of the anther.
Horizontal bar, 1 cm.

2.2.3 Application of phenotyping methods in glasshouse trials

Plants of the following varieties of *T. aestivum* were grown in five litre pots filled with Levington's C2 compost; Cadenza, Copain, Cyber, Isidor, Stigg, KWS Ferrum, Piko, R2, R6 and R44. The first five varieties, were selected based on the results of preliminary field scoring while the remaining five have independently been observed to exhibit characteristics

that encourage outcrossing. All plant material was obtained from KWS UK Ltd., Thriplow, UK.

Two glasshouse trials were carried out; with the first and second sowing being carried out on in July 2015 and November 2015 respectively. Plants were vernalized for 56d at 5^{o}C; subsequent to which they were grown within a controlled temperature glasshouse (18^{o}C day and 15^{o}C night) within a completely randomized design and irrigated with an automatic drip system.

2.2.3.1 Pollen grains per anther and anther length

Eight replicate plants of each genotype were used in the assessment of both pollen grains per anther and anther length; with sampling being carried out as described in Section 2.2.2.1 and 2.2.2.2. Samples were taken from the top, middle and bottom regions of a spike, from either primary or secondary florets (Figure 2.3 B), with all three anthers from a sampled floret being used.

2.2.3.2 The number of anthers extruded per floret

The number of anthers extruded per floret was determined using the eight replicates. The spike of the main tiller was harvested after all regions within the spike had undergone anthesis. The number of non-extruded anthers was

determined for four florets from the middle region (Figure 2.2 A); this was then extended to all primary and secondary florets along the length of the spike. Assuming three anthers per floret, the number of extruded anthers was determined and this value was expressed relative to the total number of florets counted. This gave the average number of anthers extruded per floret (0-3 scale).

2.2.3.3 Statistical analysis of phenotypic data

Analyses of variance (ANOVA) were carried out for assessed traits with genotype, and where applicable position within the spike, as fixed factors using GenStat16 (VSN International Ltd., Hemel Hempstead, UK). Correlations between traits and phenotypic values between trials were tested with Pearson's product moment correlation coefficients; carried out within the R environment using the R package 'Hmisc'.

2.2.4 Application of phenotyping methods in field trials

A panel of 15 genotypes were used to test approaches towards phenotyping additional floral traits in field trials. Assessments were carried out in 2014 at KWS' main breeding site within Cambridgeshire, UK using field grown material. Genotypes were sown in a randomized complete block design consisting of two blocks of 1.5m wide and 6m long plots.

Daily environmental variables, maximum/minimum air temperature (^{o}C) and rainfall (mm), during the field trial were accessed from the Ickleton weather station and provided by staff attached to KWS UK Ltd (Appendix 1).

2.2.4.1 The extent of pollen shedding external to florets

Pollen traps were constructed by coating a glass microscope slide with petroleum jelly on one surface; slides were then suspended approximately 5 cm below spikes using a bulldog clip and a wooden stake, with the coated side facing upward (Figure 2.3). Once approximately 10% of the spikes within a plot were at anthesis, slides were set out at 0800 h and collected at 1200 h daily over a 5d period at three positions within the centre of a plot. For each slide, the number of pollen grains within three random fields of view were counted and the average number of pollen grains per 10 cm^{2} was determined.

Figure 2.3: Photograph showing pollen trap suspended beneath flowering spikes within a field plot.

2.2.4.2 Anthesis duration and the area under the progress curve (anthesis pattern)

An alternative approach towards flowering score assessments (Section 2.2.1) was developed for use in this trial. The number of flowering spikes within a 0.09 m^2 quadrat were counted at two randomly chosen points near the centre of each plot; with a flowering spike being defined as one with dehiscing anthers on the day of observation. Observations were started when approximately 10% of spikes within a plot were at anthesis and repeated once daily until flowering ceased; the number of days between these two time points was taken as the anthesis duration.

Averaging across the two counts, the number of flowering spikes was expressed relative to the total number of spikes

within a quadrat; the derived percentage was taken as the flowering score for that given day. In order to quantify the duration and synchronicity of anthesis within a plot, the trapezoidal method was adopted. This approach entailed calculating the average flowering score between pairs of adjacent time points, which were then summed to give an area under the flowering progress curve (AUFPC) value for the flowering period. This method is the same used in calculating the area under the disease progress curve (AUDPC) by plant pathologists (Jeger and Viljanen-Rollinson, 2001) .

2.2.4.3 Anther length and anther extrusion

Anther length and anther extrusion assessments were carried out as described in Section 2.2.3.2; with 12 and 10 spikes per plot used for each trait respectively.

2.2.4.4 Statistical analysis of phenotypic data

For pollen trap data, variance components were estimated by the restricted maximum likelihood (REML) method within the R software environment using a random model. The significance of variance components was determined using model comparisons with likelihood ratio tests (Stram and Lee, 1994). Analysis of variance (ANOVA) were carried out for anther length, anther extrusion and AUFPC; with genotype as a fixed factor using GenStat16 (VSN International Ltd., Hemel Hempstead, UK).

2.3 Results

2.3.1 Preliminary field trial; flowering score and visual anther extrusion assessments

Using an n value of fifteen (15), flowering score profile clusters were obtained. Genotypes appeared to be grouped by both the initiation of flowering (anthesis) and the shape of flowering score curves (Figure 2.4). All clusters except 13 and 14 contained more than one genotype. We observed similar flowering score profiles (Figure 2.4) in early and late flowering clusters; for example, clusters 11 and 15 (Figure 2.1).

A dendrogram was generated with the 15 clusters in order to visualize how similar flowerings score profiles were between clusters. Alongside this, a heat map displaying the distribution of visual anther extrusion scores for each cluster was plotted; this was done to estimate the extent to which anther extrusion biased flowering score data. In some instances, clusters which had a similar distribution of anther extrusion scores exhibited distinct flowering score profiles; for example, clusters 6, 9 and 11 (Figure 2.5).

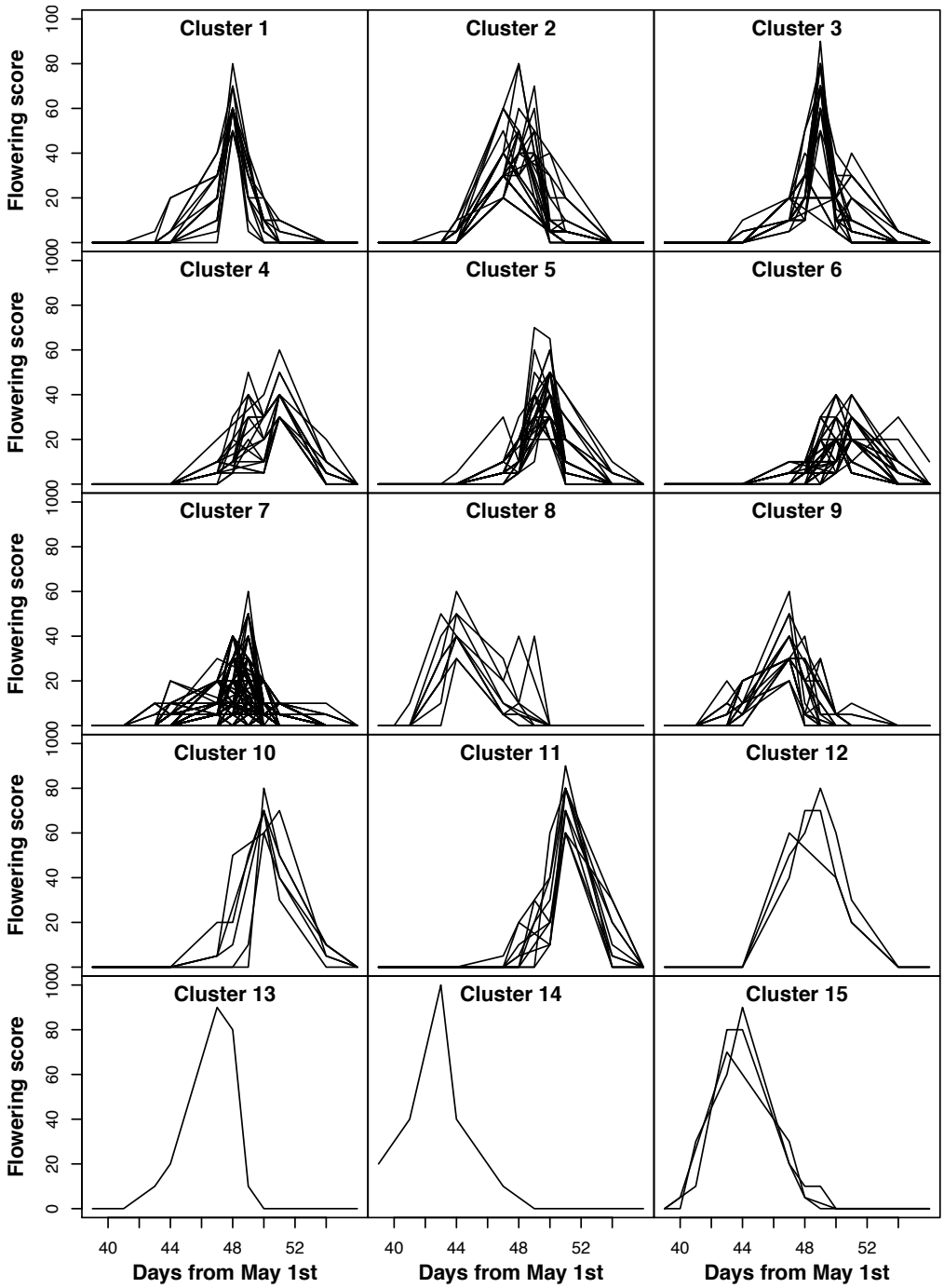

Figure 2.4: Plots of flowering scores against days from May 1st for each flowering score cluster.

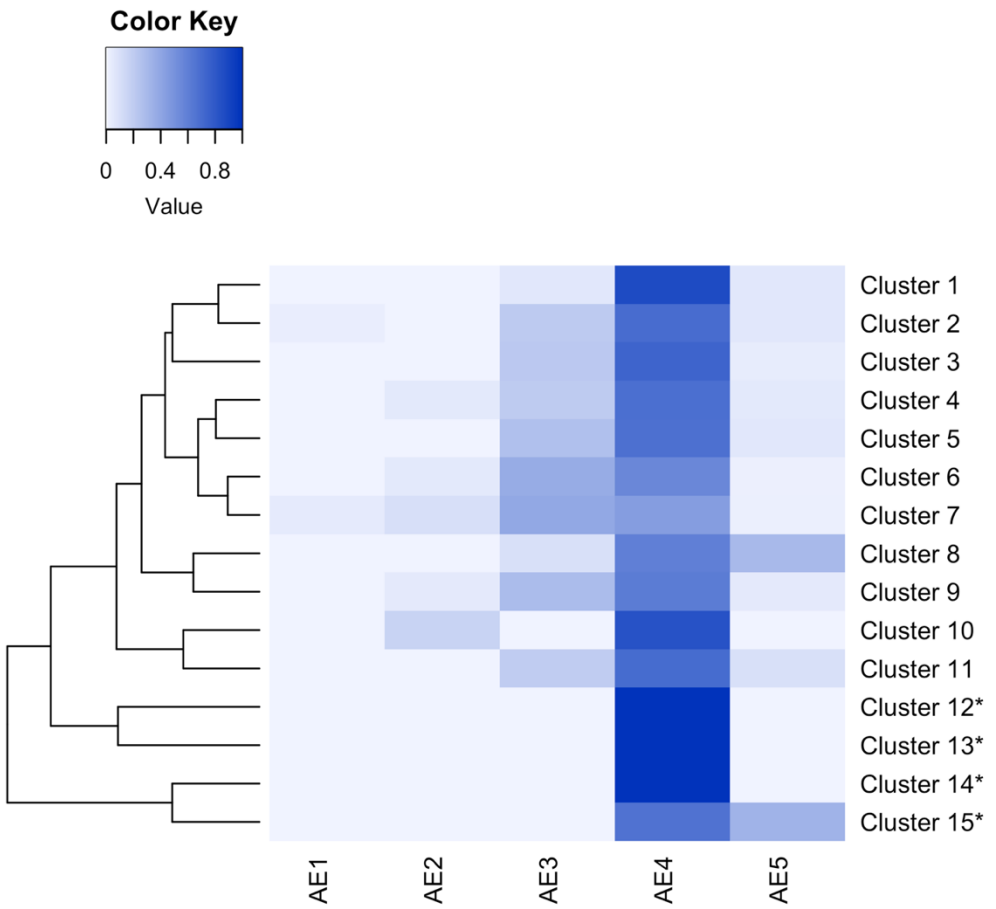

Figure 2.5: Heatmap of visual anther extrusion scores for each flowering score cluster.
The colour scale (top left) shows the proportion of individuals falling into each visual anther extrusion category (AE 1-5), described in Section 2.2.1.2. The dendrogram to the left illustrates the similarity amongst clusters with respect to flowering score profiles.

From the results of clustering, five genotypes with contrasting anthesis patterns (Figure 2.6) were selected for subsequent phenotypic assessments in glasshouse trials described in Section 2.2.3.

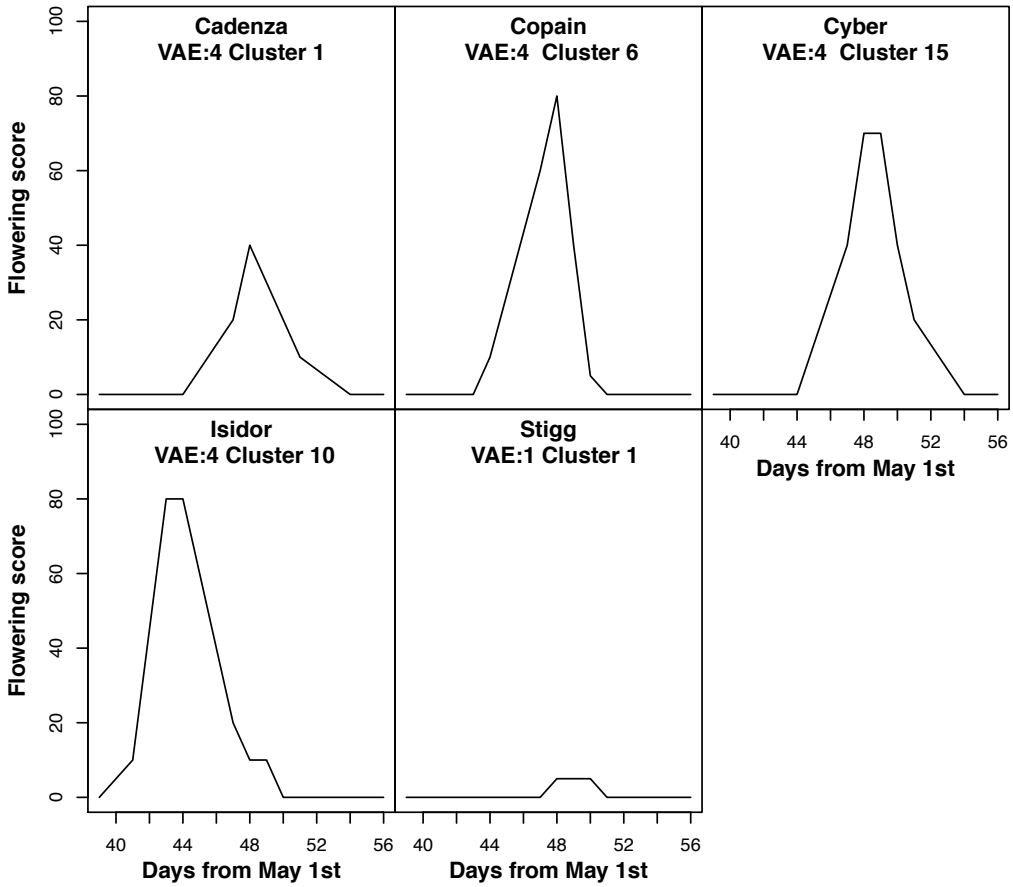

Figure 2.6: Plots of flowering score against days from May 1st for five genotypes selected for glasshouse trials.

2.3.2 Initial evaluation of phenotyping methods for anther length and pollen production

Storage of anthers in fixative at room temperature resulted in minor reductions in anther length (Figure 2.7 A). Averaging across the five days gave a low standard error of 0.01. We assumed that these differences would not mask genotypic variation in this trait.

Water extraction of pollen grains appeared to be most effective. On average, the number of pollen grains extracted using ethanol was considerably lower than that obtained by water extraction (Figure 2.7 B). Observations of anthers subsequent to vortexing confirmed conclusions drawn from pollen count data; with ethanol treated anthers retaining a larger proportion of pollen grains.

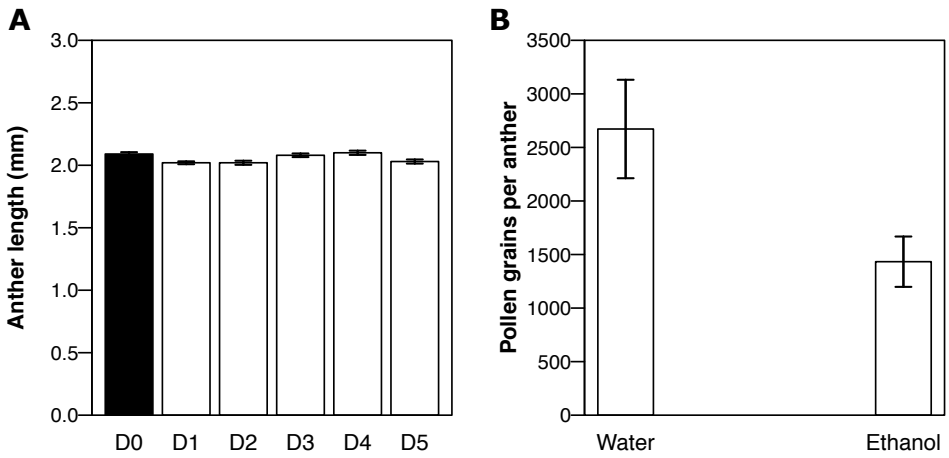

Figure 2.7: Prelimiary testing of phenotyping methods using *T. aestivum* cv. Paragon; **A)** effect of storage in fixative on anther length; **B)** pollen grains per anther extracted using water and ethanol.
Means and +/- S.E. presented. D0-D5; Day1 – Day5.

2.3.3 Genotypic variation for pollen production and anther length in glasshouse trials

Significant genotypic differences were detected for the number of pollen grains per anther and anther length (Figure 2.8 A and B) ($F_{9,\ 479}$= 39.45, *P*=<0.001+ $F_{9,\ 479}$= 28.57, *P*=<0.001). Variation related to sampling position within the spike (Figure 2.8) was also found to be significant for both traits ($F_{2,\ 479}$= 44.08, *P*=< 0.001+ $F_{2,\ 479}$= 6.38, *P*=0.002). The largest differences were observed between top florets and those at the middle and bottom.

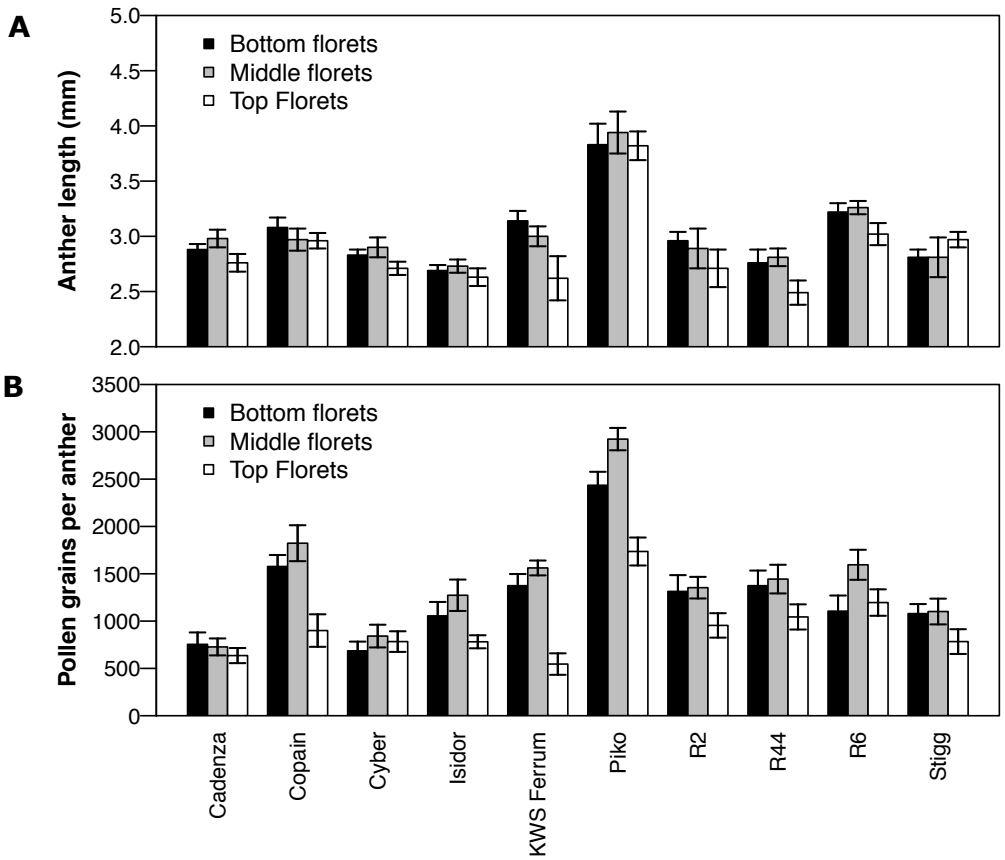

Figure 2.8: Genotypic variation in **A)** anther length and **B)** pollen grains per anther within different regions of the wheat spike.
Means and +/- S.E. presented.

Positive significant (P < 0.01) correlations were observed between anther length and pollen grains per anther in all three regions of the spike (Figure 2.9). Exclusion of the genotype Piko, which exhibited an extreme phenotype for both anther length and pollen grains per anther, resulted in insignificant correlations between the two traits; p>0.05 in all regions of the ear.

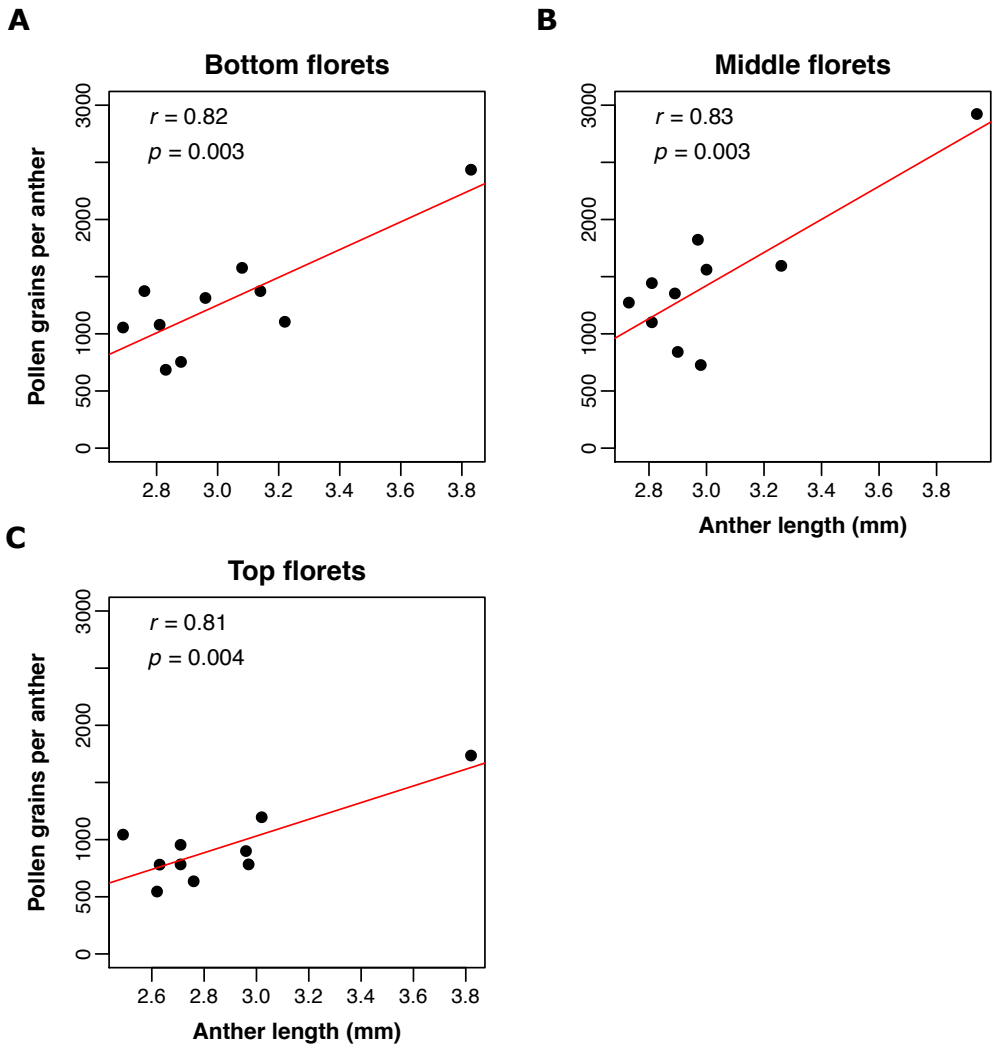

Figure 2.9: Pollen grains per anther in relation to anther length within florets from different regions within the wheat spike
A fitted regression line (red) and the Pearson correlation coefficient (r) are presented for each comparison.

2.3.4 Genotypic variation for anther extrusion in glasshouse trials

Significant genotypic differences were detected for the number of anthers extruded per floret (Figure 2.10 A) ($F_{9, 159}$= 38.32, P<0.001) derived from florets in the middle region of the spike. Values from middle florets were significantly correlated (P<0.001) with the average for the whole spike, in both trials (Figure 2.10 B and C).

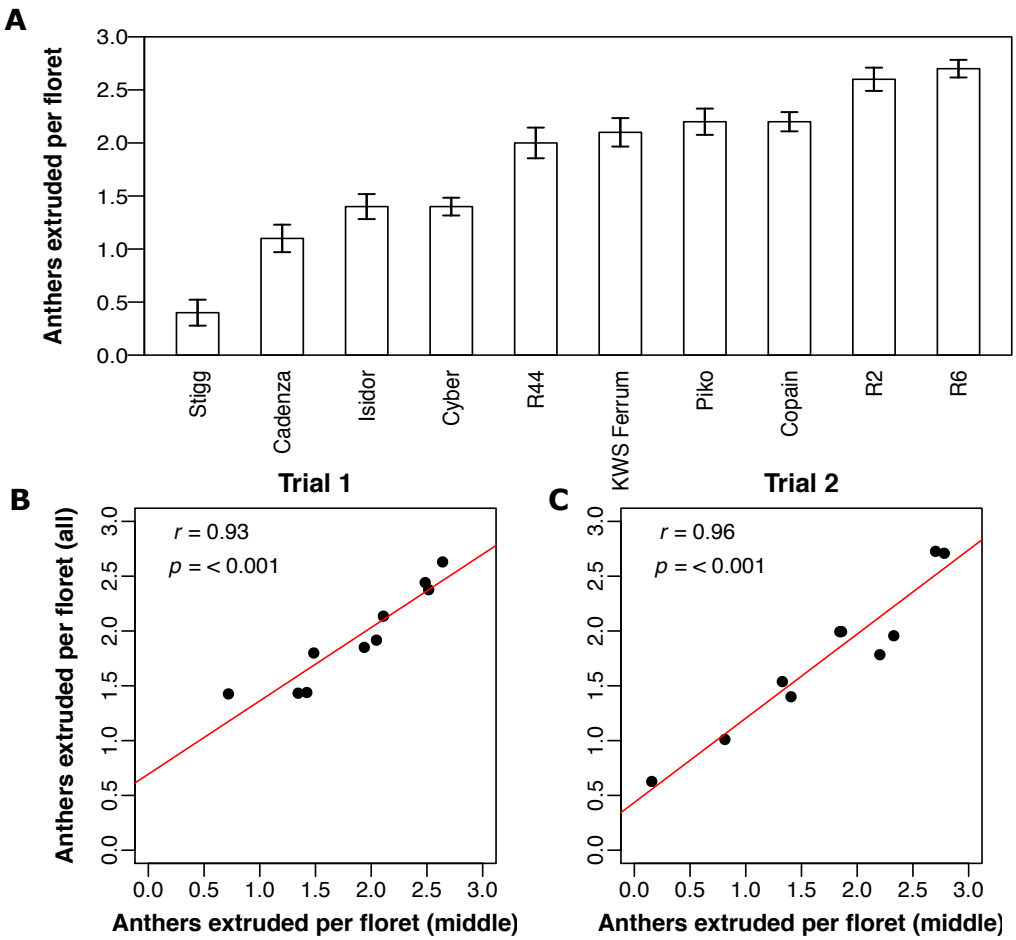

Figure 2.10: **A)** Genotypic variation in the number of anthers extruded per floret in middle of the spike. Correlation of anther extrusion values from middle florets and from whole spikes for trial **B)** one and **C)** two.
A) Means and +/- S.E. presented. **B)/C)** A fitted regression line (red) and the Pearson correlation coefficient (r) are presented for each comparison.

2.3.5 Variation in floral phenotypes between glasshouse trials

Significant correlations were observed between trials for anther length, anthers extruded per floret and pollen grains per anther ($P<0.05$) (Figure 2.11). Exclusion of Piko from the dataset resulted in an insignificant correlation between anther lengths from the two trials; this was not observed for pollen grains per anther or anthers extruded per floret.

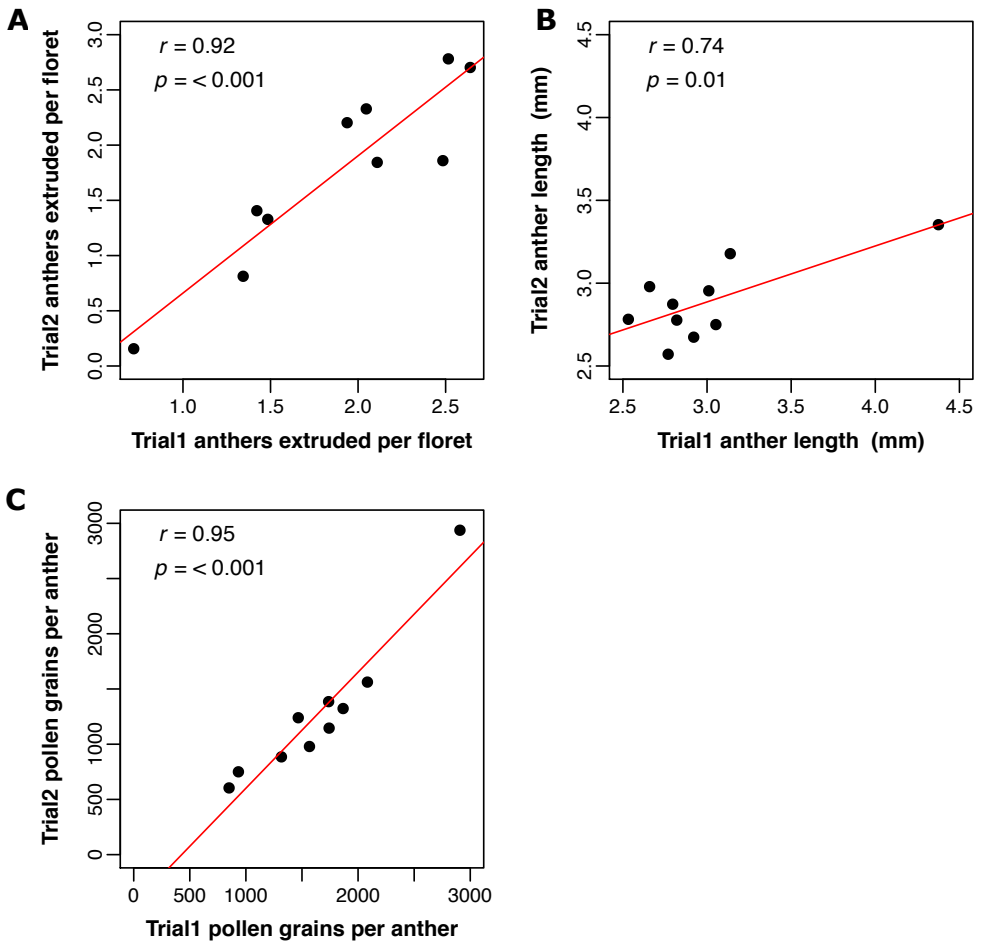

Figure 2.11: Correlation of phenotypic values between the first and second trial for **A)** anthers extruded per floret; **B)** anther length and **C)** pollen grains per anther.

A fitted regression line (red) and the Pearson correlation coefficient (r) are presented for each comparison.

2.3.6 Traits phenotyped in field trials

The number of pollen grains shed per 10 cm^2, estimated using the adhesive surface based pollen traps described in Section 2.2.4.1, showed significant genotypic variance (σ^2_G) (*P* <0.001). Additionally, the variance component associated with the response of genotypes over time (σ^2_{GxT}) was significant (*P*<0.001) (Table 2.1).

The proportion of variance explained by differences between replicates was found to be considerably smaller than that explained by genotype or the response of genotypes over time and was not found to be statistically significant (Table 2.1).

Table 2.1: Summary statistics for pollen grains shed per cm^2 by fifteen (15) genotypes.

Parameter	Pollen grains shed per cm^2
σ^2_G	989.6 ***
σ^2_{GxT}	1045.6***
σ^2_R	1.6
σ^2	489.1

Genotypic variance (σ^2_G), genotype-by-time interaction variance (σ^2_{GxT}), replicate variance (σ^2_R) error variance (σ2), and heritability (h2). *** significantly different from zero at the 0.001 level of probability.

Clear genotypic differences with respect to the number of pollen grains shed per 10 cm^2 over time were observed (Figure 2.12). For the genotype Piko, a high level of pollen was shed throughout the assessment period (Figure 2.12). Whereas in the case of the genotype Oakley, a sharp increase in pollen shedding was observed on the 32nd day from May 1st that was not sustained subsequently (Figure 2.12).

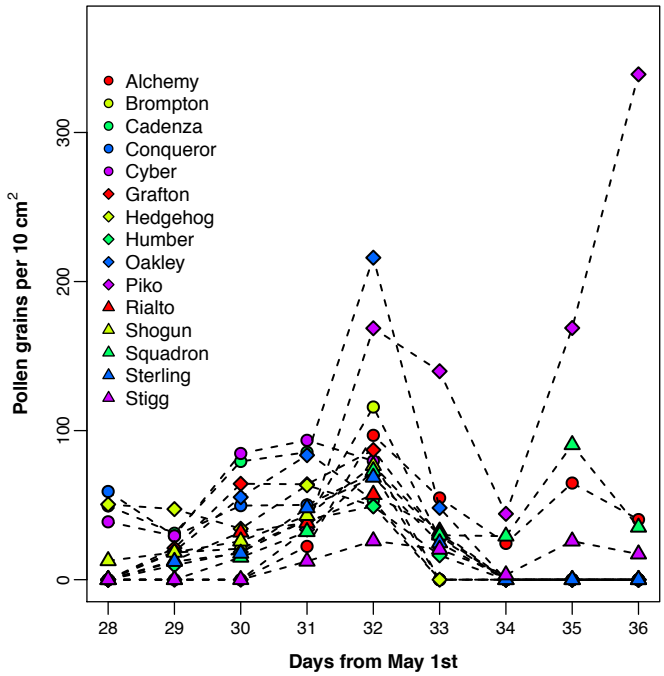

Figure 2.12: Pollen grains shed per 10 cm^2 over the observational period for fifteen (15) genotypes.
Means are presented.

Across genotypes the general pattern observed over 5d was a gradual increase and subsequent decrease in pollen shedding (Figure 2.13). Piko showed no decrease in this time (Figure 2.13 B), apart from on the 33rd day from May 1st during which rainfall occurred during the day of assessment (Table 2.2)

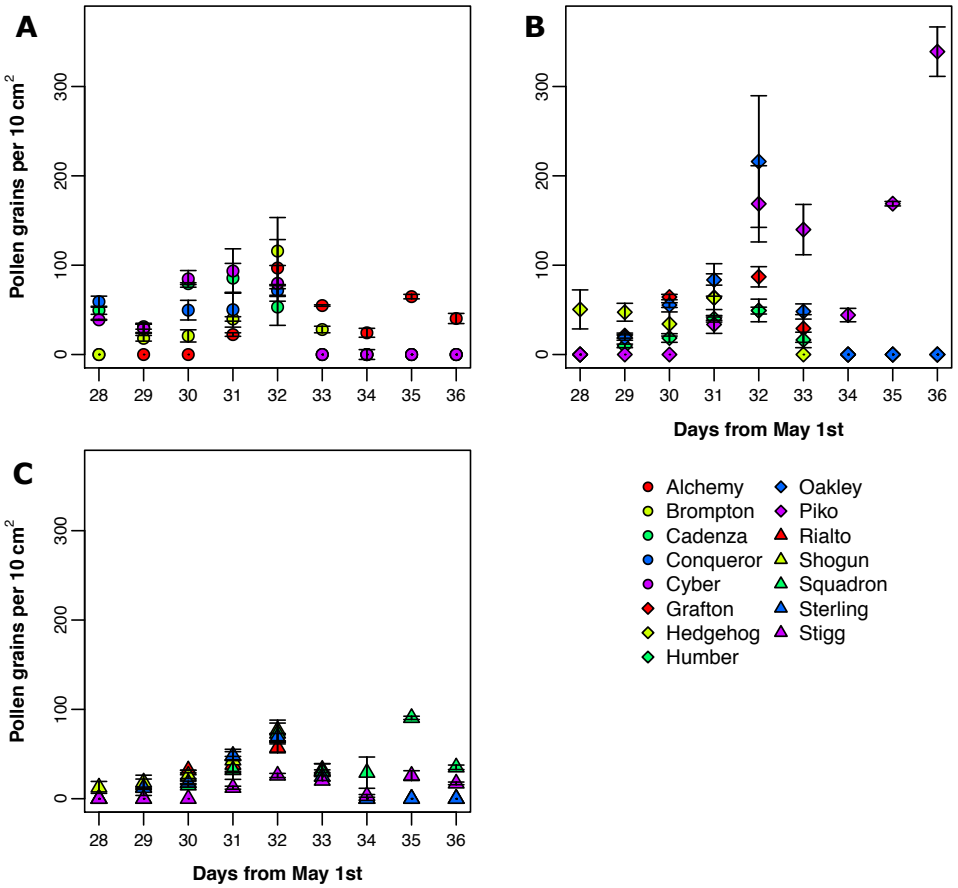

Figure 2.13: **A-C**; Pollen grains shed per 10cm^2 over the observational period for fifteen (15) genotypes.
Means and +/- S.E. presented are presented for each time point.

Table 2.2: Daily maximum/minimum temperature and rainfall during the five (5) of pollen trap assessments.

Days from May 1st	Maximum air temperature (°C)	Minimum air temperature (°C)	Rainfall (mm)
28	14.9	9.7	9.8
29	18	11.5	0
30	14.8	6.6	0
31	17.6	3.6	0
32	20.6	8.8	0
33	19.9	10.8	2.8
34	20.4	11.5	6.8
35	14.4	6.7	1.8
36	18.1	8.1	0

Significant genotypic differences were detected for anther length ($F_{14,\ 29}$= 7.87, P<0.001), anthers extruded per floret ($F_{14,\ 29}$= 68.81, P<0.001) and area under the flowering progress curve ($F_{14,\ 29}$= 5.48, P=0.001); all were significantly correlated (P<0.05) with the total number of pollen grains shed (Figure 2.14). Exclusion of Piko resulted in an insignificant correlation with anther length (P>0.05); correlations with the two other traits remained significant.

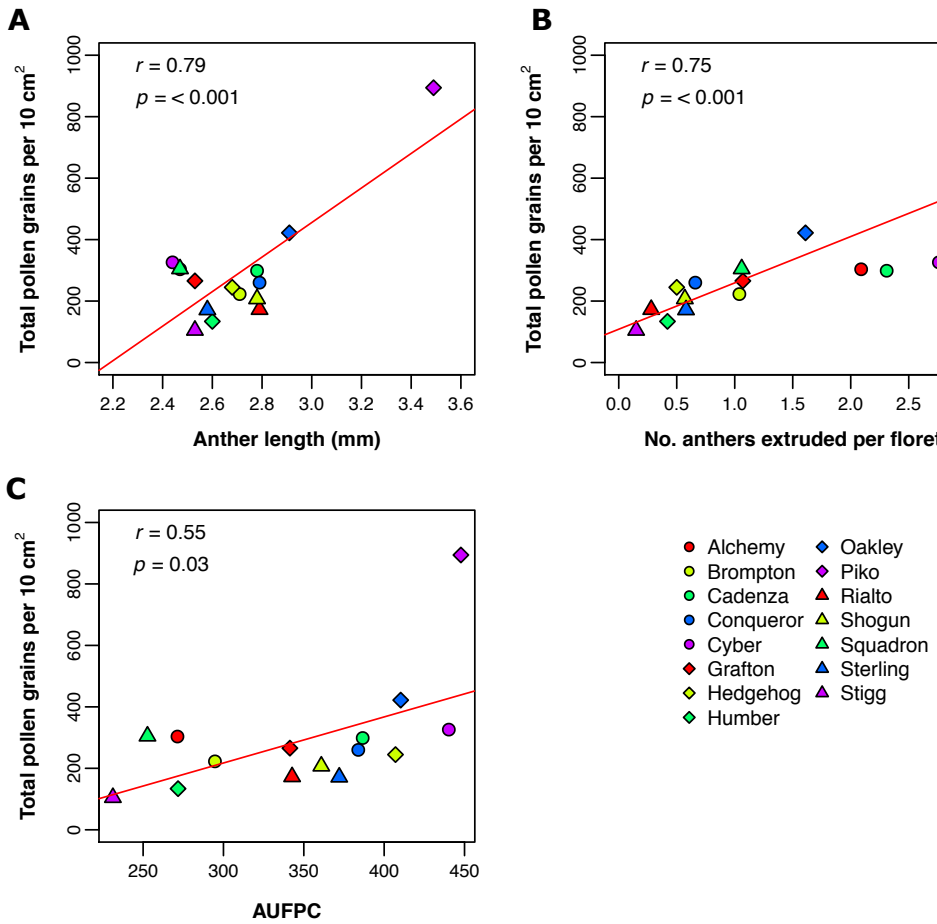

Figure 2.14: Scatterplots of the total number of pollen grains shed per 10cm^2 over five days in relation to; **A)** anther length, **B)** the no. of anthers extruded per floret and **C)** the area under the flowering progress curve (AUFPC).
A linear regression line (red) and the Pearson correlation coefficient (r) is presented for each comparison.

2.4 Discussion

Outcrossing potential can be described as a function of a genotype's capacity to produce pollen, extrude anthers, duration of its flowering period, aerodynamic properties of its pollen grains and pollen longevity/viability of the pollen. With the exception of the last two traits, which are challenging to analyse, genotype level variation has been reported prior to this study.

Phenotyping methods used in initial reports were applied to relatively small panels and the applicability of these methods for screening a large number of lines was unknown. The present study involved initially testing these methods, to determine how accurately phenotyping for floral traits can be conducted and to gain insight into how time consuming and labour intensive each method is.

2.4.1 Flowering scores and visual anther extrusion scores

Prior to this study, De Vries (1973) attempted to assess flowering patterns within glasshouse trials through visual assessment of anther extrusion on a daily basis. This approach is similar to that taken within the preliminary field trial of our study. De Vries (1973) reported genotypic differences in flowering pattern; with some genotypes

exhibiting a slower progression of flowering through the first, second and third tillers.

The variation in flowering patterns observed in our preliminary field trial (Figure 2.4) is likely attributable to genotypic effects as opposed to environmental variance. This assertion is supported by distinct patterns being observed between early flowering genotypes, the same can be said for late and intermediate flowering.

Genotypic variation for the extent of anther extrusion was anticipated within the relatively large panel (445 genotypes) used in the preliminary trial. As flowering score assessments were based on visual detection of freshly extruded anthers, variation in flowering score patterns could have been attributable to differences in anther extrusion capacity.

While this possibility cannot be definitely ruled out, the distribution of visual anther extrusion scores, which were done at the peak of anthesis and not based solely on freshly extruded anthers, suggest that anther extrusion capacity did not completely bias the data. Within clusters that are characterised by relatively high flowering scores, sustained for a prolonged period of time, variation in visual anther extrusion scores was observed (Figure 2.5) suggesting that this is not the only factor influencing flowering score patterns.

2.4.2 Method evaluation and application under controlled conditions

Conducting anther extrusion assessments using eight (8) central florets was found to provide an accurate representation of the anther extrusion capacity of the spike as a whole (Figure 2.10 A and B); which is essential if selecting for improved anther extrusion capacity. Within field trials, using middle florets is more feasible for screening a large number of lines; the results of the present study indicate that it provides a reliable estimate.

Empirically measuring anther length within a field trial is constrained by the narrow window of time within which it must be carried out. Coupled to this, there is the potential inaccuracy associated with carrying out measurements in the field. Collection and storage of anthers for subsequent measuring using imaging tools, provides a more accurate and feasible approach (Figure 2.7). Storage of anthers in a fixative of 3:1 ethanol/glacial acetic acid facilitated this within our glasshouse trials without masking genotypic differences (Figure 2.8 A).

An alternative approach towards assessing anther length or size would be to carry out visual scoring, which can be done relatively quickly for a large number of lines. Subsequent to

the application of this approach in their study, Langer et al. (2014a) reported a low heritability for visually scored anther size. They identified the same limitation described in the present study, the narrow time window within which accurate scoring can be achieved.

While not labour intensive, determining the number of pollen grains as described in this study requires processing to be carried out on the same day as sampling. In our initial tests, the integrity of pollen grains deteriorated if stored for more than three hours at room temperature. As a consequence, sampling can only be carried out in instances where a light microscope and laboratory vortex are available. Refrigeration of pollen grains, or storage in a medium that would prevent lysis of grains, until such time that they can be processed could provide a means of overcoming this limitation. These possibilities were not explored in the present study and should be investigated in future work.

Assessments of pollen grains per anther were therefore not attempted in the field trials that were conducted subsequent to this initial study; it was anticipated that this could impose a constraint on the number of additional traits that could be phenotyped for. Efforts were instead focused on assessing

anther extrusion capacity and anther length, together with anthesis characteristics.

2.4.3 Phenotyping approaches applied in field trials

Statistically significant genotypic and temporal variation in pollen shedding was detected using the pollen trap method applied in the present study (Table 2.1). The low amount of variance explained by individual replicates of a genotype (Table 2.1) suggest that surrounding plots had a minimal contribution to the amount of pollen detected on adhesive surface and that true genotypic effects were detected. The effectiveness of the method was dependent on weather conditions; a considerable dip in the number of pollen grains adhered to the traps occurred during a day of heavy rainfall for all genotypes. This inherent limitation of using an adhesive coated surface has previously been described (De Vries, 1972).

An alternative approach was described by Langer et al. (2014a), wherein the authors assessed pollen shedding capacity by bagging spikes prior to anthesis and subsequently measuring the mass of pollen collected within bags once anther extrusion had ceased. The authors reported a high heritability for pollen mass across three locations; suggesting

that environmental conditions have a relatively minor influence on the effectiveness of the approach. To date, no investigations have been undertaken in order to assess whether bagging of spikes affects the extent of anther extrusion and ultimately estimations of pollen shedding capacity under field conditions.

The simultaneous application of both pollen trap and pollen mass assessments may be of value. In addition to genotypic variation, pollen traps can provide insight into temporal differences in pollen shedding provided conducive weather conditions prevail. If this is not the case, estimates of pollen shedding capacity can still be obtained from pollen mass assessments.

As anticipated, the amount of pollen shed external to the wheat spikes showed positive relationships with both anther length, previously described as a proxy for pollen grains per anther, and anther extrusion (Figure 2.11). A modest positive correlation was also observed with the area under the flowering progress curve (Figure 2.11 C). These observations emphasise the need for the simultaneous selection of multiple traits towards the development of effective male parental ideotypes.

As compared to visual flowering score assessments, a more empirical approach towards characterizing anthesis duration

and pattern was applied in Section 2.2.4.2. Significant genotypic differences were detected for the value derived from this approach, the area under the flowering progress curve (AUFPC) (Section 2.3.6). Phenotyping AUFPC required constant field presence and a disproportionate investment in time compared to other floral traits; this approach may therefore be unfeasible in screening a large number of genotypes. While time demanding, the capacity to detect significant genotypic differences in AUFPC supports the accuracy of this method. Further to this, the trait was found to be correlated with the total number of pollen grains shed over a 5d period (Figure 2.14 C) and may be applicable in detailed characterization of small genotype panels.

The application of phenotyping methods for anther length and anther extrusion in the field trial confirmed conclusions drawn from their initial testing under controlled conditions; the application of each being found to be feasible under field conditions. Additional traits could also be assessed simultaneously, namely area under the flowering progress curve and the number of pollen grains shed per m^2. A decision was made to not carry out assessments of pollen grains shed per m^2 in subsequent field trials, due to the large time investment required, to set out pollen traps, coupled to a limited availability of labour.

2.4.4 Genotypic variation in floral traits within glasshouse trials

Seven European winter wheat varieties and three male parental lines were phenotyped with the aim of assessing the extent of variation for floral traits. Even within a small set of European lines, significant genotypic variation was detected for the number of pollen grains per anther, anther length and anther extrusion (Figure 2.8 and 2.10).

Genotypic variation for the number of pollen grains per anther ranged from 546 to 2923 (Figure 2.8 B), which is similar to that reported in prior studies (Beri and Anand, 1971, Heyne and Livers, 1967, Khan et al., 1973). No genotypes exhibited pollen production at levels comparable to that of *Secale cereal* an obligate out-crosser (Nguyen et al., 2015).

A number of lines were however comparable to R2, R6 and R44 (Figure 2.8 B) which are currently used as male parental lines in hybrid breeding. The highest pollen production was observed for Piko (Figure 2.8 B); subsequent to being released as a commercial variety, Piko has been used as a male parental line in commercial hybrid production by Saaten Union (Jacob Lage, KWS UK Ltd., personal communication, 2017).

Across genotypes, the number of pollen grains per anther was lower in florets within the top region of the spike (Figure 2.8 B). In high pollen producers such as Copain, KWS Ferrum and Piko this difference was more apparent; suggesting severely impaired pollen production capacity in this region. This observation can be related to the developmental gradient within the wheat spike; with smaller less advanced florets being found within the top region (Kirby and Appleyard, 1987). As the bottom and middle regions are likely to make the greatest contribution towards outcrossing potential; phenotyping within a large panel could be restricted to either region.

Significant genotypic differences were detected for anther length within the panel (Figure 2.8 A). Similar temporal differences within the spike were observed as for pollen grains per anther; florets from the top of the spike produced shorter anthers. The length of anthers in middle and bottom florets was relatively similar; suggesting that florets from the bottom region can be used for phenotyping anther length if the middle of the spike has passed the appropriate sampling stage.

Significant genotypic variation in anther extrusion was detected within the panel, with extremes at both tails of the

distribution of phenotypic values being observed (Figure 2.10 A). Lack of variation for anther extrusion capacity, anther length and pollen grains per anther cannot be described as a limiting factor in the development of male ideotypes with improved outcrossing potential.

For all three traits assessed within glasshouse trials, significant correlations ($P<0.01$) were observed between phenotypic values obtained in the first and second trial (Figure 2.11). This observation suggests that a reasonable level of accuracy can be achieved using the methods described and further supports the feasibility of phenotypic selection for these traits, as phenotypes appear to be robust over time.

2.4.5 Simple traits as proxies for traits that are more laborious to assess

Assessing the number of pollen grains per anther as done in the present study may not be feasible for field trials. The presence of a genetic link between the trait and anther length/size could facilitate indirect screening for pollen production. While significant correlations were observed between the two traits in the present study, care should be taken in drawing conclusions due to the small number of

genotypes investigated in glasshouse trials and the fact that only one genotype (Piko) exhibited extremes for both traits.

Prior reports of a positive correlation do not necessarily lend support for the validity of this phenotyping approach either, as these conclusions have been drawn from relatively small genotype panels as well (De Vries, 1974, Beri and Anand, 1971). Establishing the presence of a genetic link between anther length and pollen grains per anther in European material would require that a larger diverse population be screened for both traits.

The pollen trap method described in the present study enabled detection of significant genotypic variation in pollen shedding but also temporal variation. Further, differences in the responses of genotypes over time could be discerned. Provided that environmental conditions are conducive to their application; pollen traps can be seen as a more time effective approach towards characterising the anthesis profiles of genotypes, compared to the flowering score method. Once collected, pollen traps could be stored and the number of pollen grains on their surface determined at a later point in time (one month in the present study). The limitations imposed by environmental conditions could be overcome by suspending a mesh surface over the canopy where pollen

traps are located, in order to minimize the impact of rainfall on the effectiveness of the adhesive surface.

2.4 Summary

All phenotyping approaches described in this chapter were demonstrated as capable of detecting consistent significant genotypic variation in floral characteristics. The level of accuracy obtained together with the level of genotypic variation observed support the feasibility of selection for these traits towards improving the outcrossing potential of wheat varieties.

The feasibility of applying some of these approaches in large field trials was supported by the relatively low investments in time and labour required. While methods appear robust enough to meet the requirements of the research areas of the present study, there certainly exists scope for improving them towards the development of high throughput phenotyping pipeline for floral traits.

CHAPTER 3: GENOTYPIC VARIATION IN FLORAL TRAITS WITHIN FIELD TRIALS

3.1. Introduction

While a number of field studies have been published during the course of the present study, at its initiation the feasibility of applying floral phenotyping methods discussed in the previous chapter and those described in much earlier studies were not known. Further to this, the extent of variation for floral traits and the relationship between them as well as with agronomic traits, had not been described for European material.

The focus of the current chapter is to simultaneously address these questions with the aim of guiding the design of breeding programmes that aim to modify floral characteristics. Using a large diverse panel in UK based field trials and a smaller panel in field trials across Europe, phenotyping approaches were applied and the results reported. In addition to the objectives described, the results of the larger field trial were used in a genome wide association study targeting floral traits, which is described in Chapter 4.

3.2 Methods

3.2.1 Phenotyping a genotype panel for floral traits in field trials

3.2.1.1 Plant material and field trials

A panel of 111 varieties from Great Britain, France, Germany, Netherlands, Sweden and Switzerland released between 1921 and 2012, were used in this association study (Appendix 1). The panel was selected based on the results of a preliminary field trial carried out in 2013, with the aim of having variation present for the duration and synchronicity of anthesis, while simultaneously minimising variation with respect to heading date. Additionally, predominantly modern semi-dwarf varieties were selected in order to increase the relevance of the present study to active breeding programmes.

Daily environmental variables, maximum/minimum air temperature (^{o}C) and rainfall (mm), during the field trial were accessed from the Ickleton weather station and provided by staff attached to KWS UK Ltd (Appendix 1).

Table 3.1: Mean maximum and minimum temperature for the month preceding heading and during the course of the field studies.

Year	Month	Maximum (°C)	Minimum (°C)
2014	May	17.7	8.7
	June	20.8	10.7
	July	24.4	13.5
2015	May	16.5	7.5
	June	20.3	9.5
	July	22.5	12.3

Varieties were sown in a randomized complete block design consisting of two blocks i.e. two replicates per variety; at a seed rate of 275/m2 in 6.5 m^2 yield plots. Trials were carried out in 2014 and 2015 within Cambridgeshire, UK (KWS: Fowlmere 2014; Newton 2015).

3.2.1.2 Phenotyping

Days to heading from May 1st (HD) and plant height (HT; cm from the stem base to the top of spikes during grain filling) were scored. Phenotypic data related to anthesis and anther characteristics were also collected, namely the area under the flowering progress curve (AUFPC), anthesis duration (AD), anther length (AL) and the number of anthers extruded per floret (AE).

Assessments of AUFPC and AD were carried out as described in Chapter 2, Section 2.2.4.2. For AE, assessments were

carried out as in Chapter 2, Section 2.2.3.2 with 12 ears being used per plot. Assessments of anther length (AL) were carried out only in the 2015 trial on a subset of 91 varieties, using the approach described in Chapter 2, Section 2.2.3.1; with ten anthers being sampled within a plot from the middle region of five ears.

In addition to floral traits, additional phenotypic datasets were collected; 1) the number of spikelets per spike (SP), 2) the number of grains per spikelet (GPS), 3) the number of grains per ear (GPE), 4) the presence/absence of awns (AWNS) and 5) the number of fertile shoots per m^2 (FS). During the grain filling period, thirty ears were collected from each plot for the assessment of SP, GPE and GPS. A 1 m^2 quadrant was used to determine the number of fertile shoots per m^2 at three positions per plot.

3.2.1.3 Phenotypic data analysis

Using a one-stage mixed model (REML) analysis with GenStat v16 software (VSN International, Hemel Hempstead, Hertfordshire, UK), the best linear unbiased estimates (BLUEs) were generated for each variety across trials; variety was treated as a fixed effect, while replicate/trial and interaction terms were treated as random. Additionally, best linear unbiased predictors (BLUPs) were generated for each

trial. The correlation between traits (BLUEs) and trials (BLUPs) was tested with the Pearson's product moment correlation coefficient; this was carried out within the R environment using the R package 'Hmisc'.

Variance components were estimated by the restricted maximum likelihood (REML) method within the R software environment using a random model. The significance of variance components was determined using model comparisons with likelihood ratio tests (Stram and Lee, 1994). Heritability (H^2) on an entry-mean basis was calculated as the ratio of genotypic to phenotypic variance as described by Melchinger et al. (1998) (σ^2_G / (σ^2_G+ σ^2_{GxE}/e+ σ^2 / r)); where σ^2_G represents the genetic variance, σ^2_{GxE} the genotype by environment variance, σ^2 the error variance, e the number of environments and r the number of replicates.

3.2.2 Assessing anther extrusion in multi-location trials

3.2.2.1 Plant material

A panel of 24 genotypes were selected for field trials in 4 locations during the summer of 2016. In assembling the panel, variation for days to heading (from May 1st) was minimized (29-37 days from May1st in 2015) while variation for the number of anthers extruded per floret (0.7-2.7) was maximized.

Each genotype was represented by two replicate $1m^2$ plots within randomised complete block designs at three KWS sites; Thriplow, Cambridgeshire (UK); Torxé, Charente-Maritime (France) and Allonnes, Sarthe (France) and by a single replicate within a completely randomised design at Wohlde, Bergen (Germany).

3.2.2.2 Phenotyping and phenotypic data analysis

AE assessments were carried out as described in Chapter 2, Section 2.2.3.2, with five spikes harvested per plot 8-10d after the initiation of anthesis. Phenotypic data was analysed within the R environment using the R package lme4. A liner mixed-effects model was applied to the dataset with genotype, location and their interaction being treated as random effects. The significance of variance components was

determined using model comparisons with likelihood ratio tests (Stram and Lee, 1994).

Heritability (H^2) on an entry-mean basis was calculated as the ratio of genotypic to phenotypic variance as described by Melchinger et al. (1998); σ^2_G / (σ^2_G+ σ^2_{GxE}/e+ σ^2 / re); where σ^2_G represents the genetic variance, σ^2_{GxE} the genotype by environment variance, σ^2 the error variance, e the number of environments and r the number of replicates.

3.3 Results

3.3.1 Assessment of floral traits in UK field trials

3.3.1.1 Phenotypic values across trials

Estimates of genotypic variance were significant for all traits assessed (P<0.001) (Table 3.2). Genotype by environment (trial) variation was significant at the P<0.05 level for all traits except the number of fertile shoots per m^2 (Table 3.2).

Table 3.2: Summary statistics for traits investigated in panel of 111 genotypes.

Parameter	AD	AE	AUFPC	GPE	GPS	HD	HT	SP	FS
Min	4.0	0.0	133.9	21.2	1.2	15.0	64.0	10.8	36.7
Mean	8.5	1.6	340.1	55.1	2.9	29.0	79.1	19.3	54.1
Max	14.0	2.9	606.4	72.9	5.1	39.0	138.0	23.0	87.7
σ^2_G	0.6 ***	0.3 ***	1746 ***	34.6 ***	0.1 ***	2.8 ***	64.9 ***	1.2 ***	18.8 ***
σ^2_{GxE}	0.7 ***	0.1 ***	1634 ***	6.8 ***	0.02 **	2.2 ***	2.4 *	0.2 **	0.0
σ^2	0.8	0.1	2358.0	13.2	0.1	0.4	8.7	0.6	54.9
H^2	0.53	0.80	0.55	0.84	0.81	0.70	0.95	0.81	0.58

Genotypic variance (σ2G), genotype-by-environment interaction variance (σ2GxE), error variance (σ2), and heritability (H^2). ***, **, *significantly different from zero at the 0.001, 0.01 and 0.05 levels of probability. AE, number of anthers extruded per floret; AD, anthesis duration; AUFPC, area under the flowering progress curve; GPE, grains per ear; GPS, grains per spikelet; FS, fertile shoots per m^2; SP, spikelets per spike; HD, days to heading from May 1st; HT, plant height.

For all assessed traits, phenotypic values (BLUPs and arithmetic means) were significantly correlated between trials (P<0.001) (Table 3.3). As anticipated, the strength of the correlations observed between arithmetic means was notably

lower than that observed between BLUPs. The most pronounced differences were observed for AD, area under the flowering progress curve (AUFPC) and FS.

Table 3.3: Correlation of BLUP and arithmetic mean phenotypic values between trials for each evaluated trait.

	AE	**AD**	**AUFPC**	**GPE**	**GPS**	**FS**	**SP**	**HD**	**HT**
2014 BLUP vs 2015 BLUP	0.82 ***	0.72 ***	0.75 ***	0.96 ***	0.96 ***	1.00 ***	0.95 ***	0.86 ***	0.99 ***
2014 mean vs 2015 mean	0.67 ***	0.36 ***	0.38 ***	0.72 ***	0.73 ***	0.46 ***	0.68 ***	0.81 ***	0.93 ***

***, significant at the 0.001 probability level. AE, number of anthers extruded per floret; AD, anthesis duration; AUFPC, area under the flowering progress curve; GPE, grains per ear; GPS, grains per spikelet; FS, fertile shoots per m^2; SP, spikelets per spike; HD, days to heading from May 1^{st}; HT, plant height.

Across both trials, the range of phenotypic values for AUFPC was relatively similar with some outliers being observed in 2014 towards the upper end of the distribution (Figure 3.1 D). For anthesis duration (AD) and the number of anthers extruded per floret (AE) the range of values differed being relatively smaller for AE and larger for AD in the 2015 trial (Figure 3.1 A and C). The mean phenotypic value was also higher for AD and lower for AE in the 2015 trial.

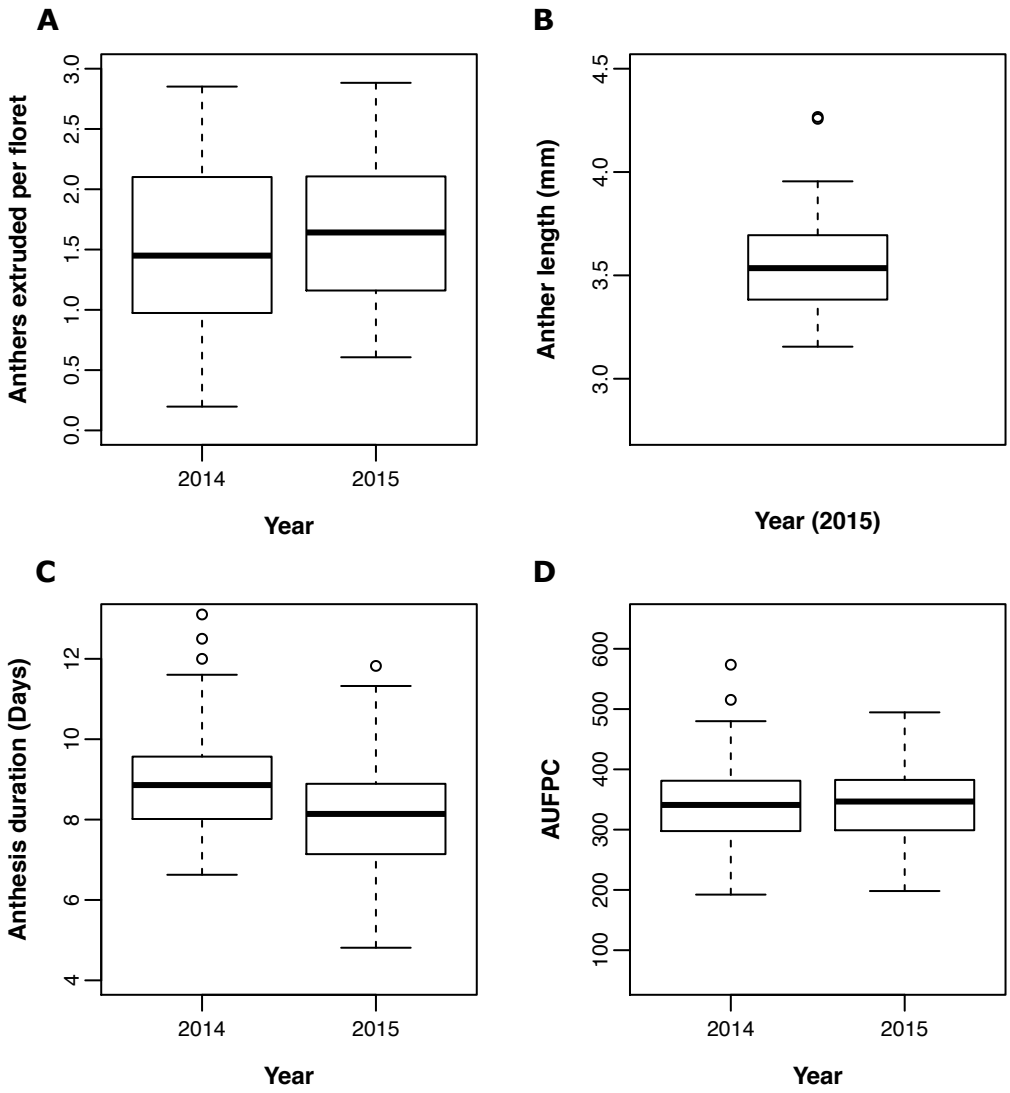

Figure 3.1: Boxplots of the distribution of phenotypic values (BLUPs) for traits; **A)** the number of anthers extruded per floret; **B)** anther length; **C)** anthesis duration and **D)** area under the flowering progress curve.

Lower mean values for grains per ear (GPE) and per spikelet (GPS) were observed in 2015 (Figure 3.2 A and B). For both days to heading from May 1st (HD) and plant height (HT), a narrower range of phenotypic values was observed in the 2015 trial. As expected due to higher temperatures (Table 3.1), earlier heading characterized the 2014 trial (Figure 3.2 C and D); shorter plants also characterized this trial.

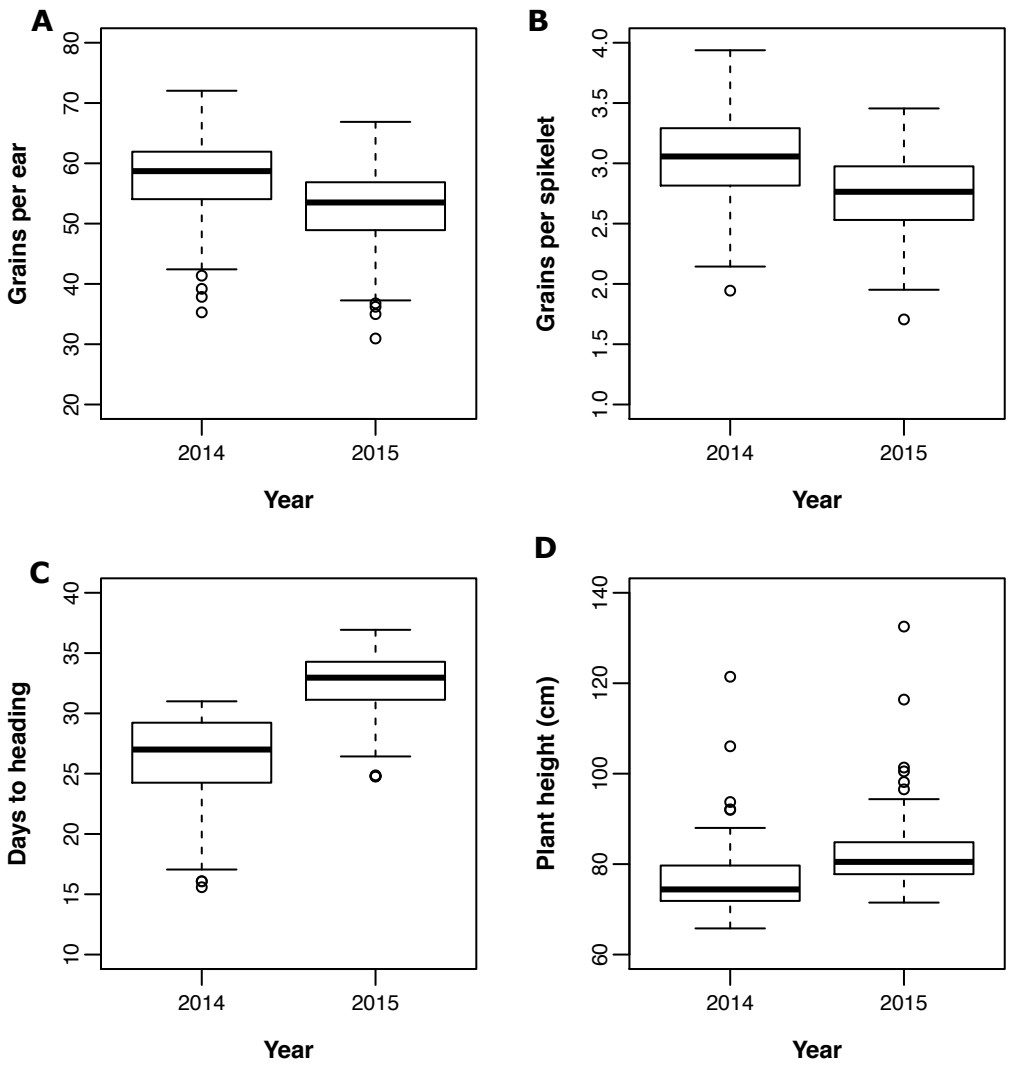

Figure 3.2: Boxplots of the distribution of phenotypic values (BLUPs) for traits; **A)** the number of grains per ear; **B)** the number of grains per spikelet; **C)** days to heading from May1st and **D)** plant height.

Across trials the range of phenotypic values for number of spikelets per spike (SP) and fertile shoots per m^2 (FS) were relatively similar in size; the mean number of spikelets per spike was however lower in the 2015 trial (Figure 3.3 A and B).

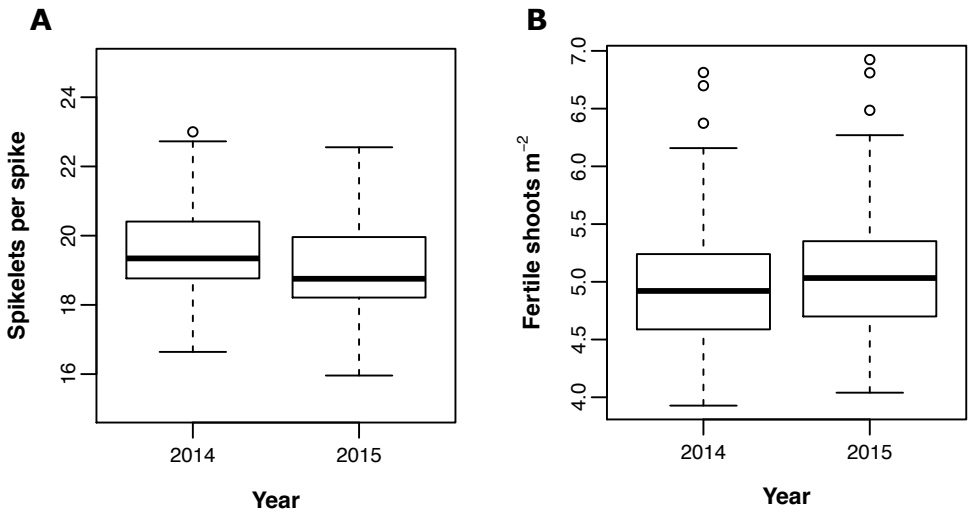

Figure 3.3: Boxplots of the distribution of phenotypic values (BLUPs) for traits; **A)** the number of spikelets per spike and **B)** the number of fertile shoots per m^2.

3.3.1.2 Distribution of phenotypic values (BLUEs)

Best linear unbiased estimates (BLUEs) generated for the numbers of anthers extruded per floret (AE), anthesis duration (AD) and area under the flowering progress curve (AUFPC) followed a near normal distribution. Mean phenotypic values for anther length (AL) from the 2015 trial showed a normal distribution as well (Figure 3.4 A, B, C and D).

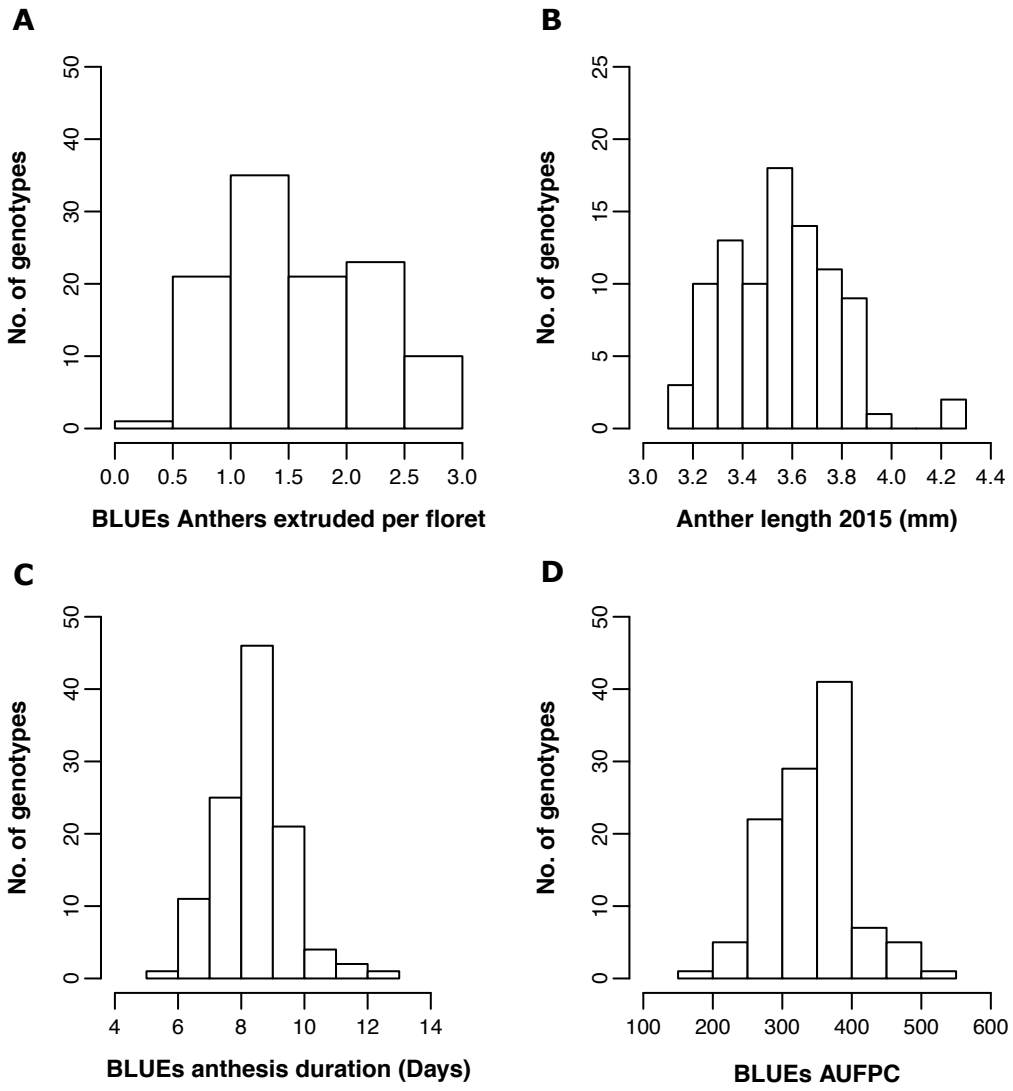

Figure 3.4: Distribution of phenotypic values (BLUEs) for **A)** no. anthers extruded per floret; **B)** anther length; **C)** anthesis duration and **D)** area under the flowering progress curve (AUFPC).

BLUEs for the number of grains per ear (GPE) and per spikelets (GPS) showed a near normal distribution (Figure 3.5 A and B). However, days to heading from May 1st (HD) showed a skewed distribution, with late flowering genotypes being more predominant (Figure 3.5 C). For plant height (HT) the majority of the panel (greater than 50%) fell within the lower end of the distribution (Figure 3.5 D).

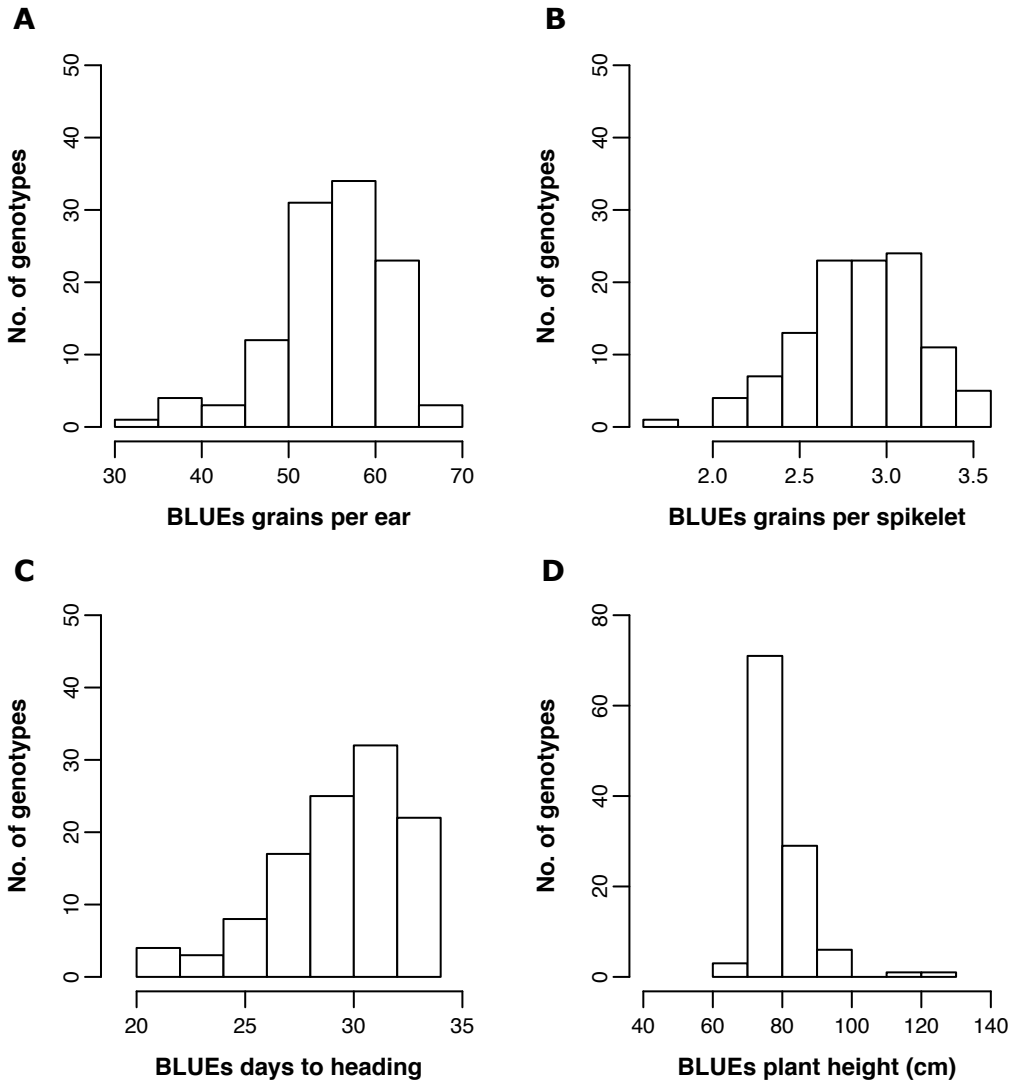

Figure 3.5: Distribution of phenotypic values (BLUEs) for **A)** grains per ear; **B)** grains per spikelet; **C)**heading date and **D)**plant height.

BLUEs for the number of spikelets per spike (SP) and fertile shoots per m^2 followed a near normal distribution; in the case of SP, a large proportion of genotypes fell within the lower end of the distribution (Figure 3.6 A & B).

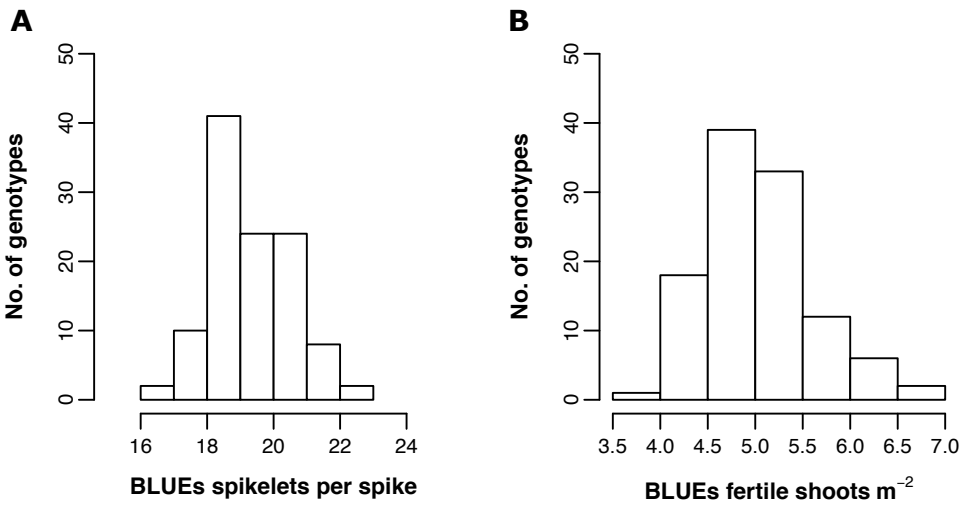

Figure 3.6: Distribution of phenotypic values (BLUEs) for **A)** spikelets per spike and **B)** fertile shoots per m^2.

3.3.1.3 Phenotypic correlations

Significant correlations were observed between floral traits and between floral and agronomic traits (Table 3.4). Correlations between floral traits and grains per m^2 were calculated but excluded from the correlation table (Table 3.4) and the correlation matrix (Figure 3.7) due to the absence of any significant correlations.

Table 3.4: Phenotypic (BLUEs) correlations between the evaluated floral and agronomic traits.

	AD	AL	AUFPC	GPE	GPS	FS	SP	HD	HT
AE	0.34 ***	-0.01	0.4 ***	-0.47 ***	-0.4 ***	0.31 ***	-0.13	-0.44 ***	0.45 ***
AD		0.36 ***	0.8 ***	-0.05	0.08	0.01	-0.25 ***	-0.58 ***	0.06
AL			0.43 ***	0.13	0.09	-0.22 *	0.1	0.2 *	-0.05
AUFPC				0.03	0.15	-0.13	-0.22 **	-0.36 ***	0.04
GPE					0.88 ***	-0.48 ***	0.28 **	0.34 ***	-0.5 ***
GPS						-0.42 ***	-0.22 *	0.2 *	-0.45 ***
FS							-0.14	-0.2 *	0.07
SP								0.27 **	-0.1

*, **, *** Significant at the 0.05, 0.01 and <0.001 probability level. AE, number of anthers extruded per floret; AD, anthesis duration; AL, anther length; AUFPC, area under the flowering progress curve; GPE, grains per ear; GPS, grains per spikelet; FS, fertile shoots per m^2; SP, spikelets per spike; HD, days to heading from May 1^{st}; HT, plant height.

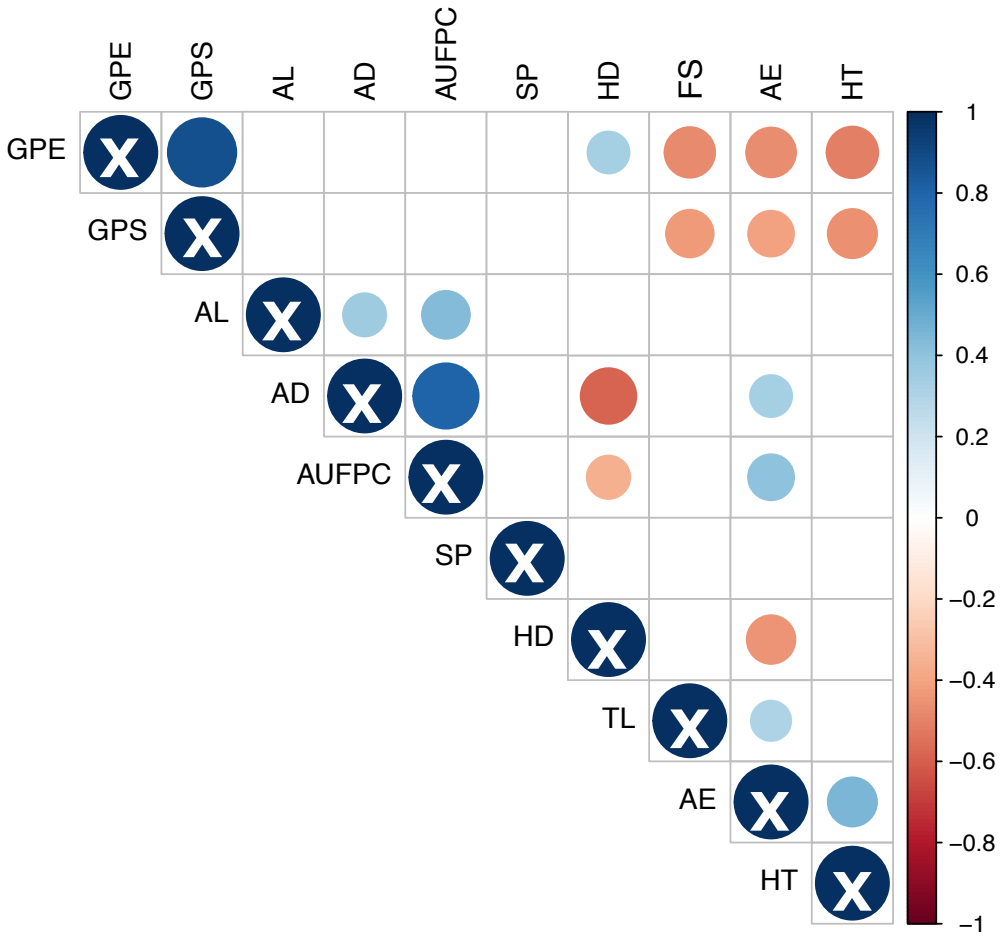

Figure 3.7:Phenotypic (BLUEs) correlations between evaluated traits. The colour scale at the right of the correlation plot corresponds to the r value of a correlation; the size of coloured dots indicates the significance of the correlation, with lower *P* values giving a larger circle. All correlations were found to be significant at the 0.05 probability level.

3.3.2 Anther extrusion in multi-location trials

Significant correlations ($P<0.001$) were observed between the anther extrusion scores of genotypes in Thriplow and that in the other three locations (Figure 3.8). The weakest correlation was observed with phenotypic values from Wholde, Germany.

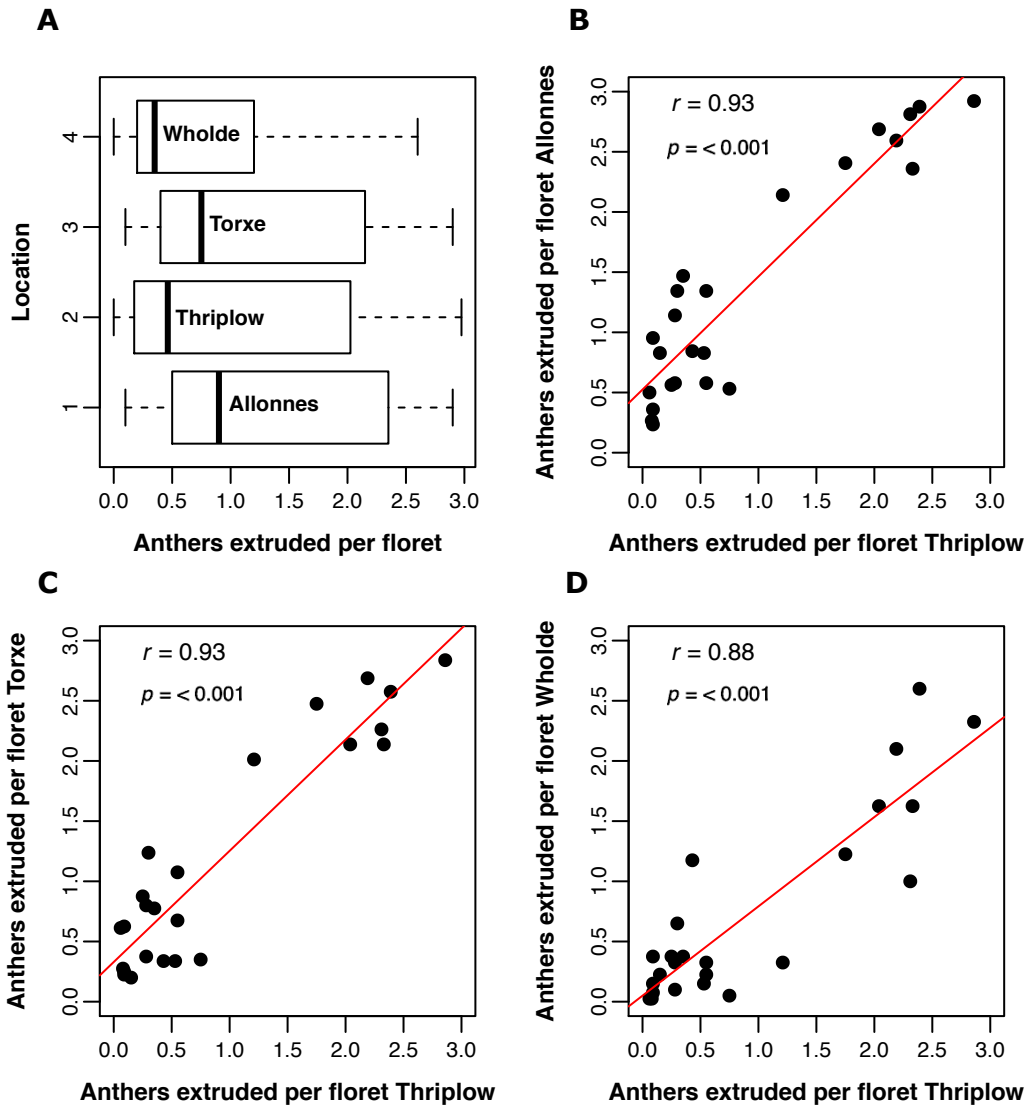

Figure 3.8: **A)** Boxplot of the distribution of phenotypic values for no. of anthers extruded per floret at three trial locations. **B, C, D)** Mean number of anthers extruded per floret at three locations, compared to the mean value observed in Thriplow.
A fitted regression line (red) and the Pearson correlation coefficient (r) are presented for each comparison.

Mean anther extrusion scores varied across locations, with significant genotype by environment (GxL) variance being detected (Table 3.5). The range of phenotypic values were relatively similar for Allonnes, Thriplow and Torxe; but notably smaller for Wholde where higher anther extrusion scores were less common (Figure 3.8 A).

Anther extrusion scores gave a high heritability H^2 (0.96); genotype by environment variance while significant, was smaller compared to the genotypic variance observed (Table 3.5).

Table 3.5: Summary statistics for the number of anthers extruded per floret (AE) for the 24 genotype panel across locations.

Parameter	**AE**
Min	0.00
Mean	1.10
Max	2.98
σ^2_G	0.75***
σ^2_{GxL}	0.04**
σ^2	0.07
H^2	0.93

Genotypic variance (σ^2_G), genotype-by-environment interaction variance (σ^2_{GxE}), error variance (σ^2), and heritability (H^2). ***, ** significantly different from zero at the 0.001 and 0.01 levels of probability.

3.4 Discussion

3.4.2 Phenotypic values across trials

For all traits assessed, significant correlations were observed between phenotypic values in the first and second trial (2014 and 2015); with correlations varying from r= 0.7 to 1.00 amongst BLUPs (Table 3.3). The high heritability values observed for days to heading from May 1st (HD) and plant height (HT) can be taken as indicative of the high precision of both trials (Table 3.2).

The heritability of the number of anthers extruded per floret was the highest amongst floral traits (H^2=0.8) and was comparable to that of the agronomic traits and values reported for anther extrusion in recent studies (Muqaddasi et al., 2017, Muqaddasi et al., 2016, Boeven et al., 2016, Langer et al., 2014a). Lower heritability was observed for anthesis duration (AD) and area under the flowering progress curve (AUFPC), H^2=0.5 and 0.6 respectively. For AD and AUFPC, genotype by environment variance (GxL) was found to be relatively high (Table 3.2).

In addition to the influence of GxL, the lower heritability observed for AD and AUFPC may have been due to the phenotyping approaches applied being more error prone as compared to that applied for AE. Examining the variance

component associated with genotype replications within trials suggests this is not the case. In each instance the variance associated with replicates within a trial was insignificant ($P > 0.05$) and the variance accounted for by this component was relatively small; 0.3516 and 0.1215 for AD and AUFPC respectively (compare with data in Table 3.2). These observations indicate that variation in phenotypic values between genotype replicates was minimal and supports the accuracy of the applied phenotyping approaches.

With respect to anther extrusion capacity, the reliability of the phenotyping approach taken in the present study is validated by the heritability observed. Further to this, the feasibility of using MAS for this trait is supported, provided that molecular markers targeting causative loci can be identified and validated (Xu and Crouch, 2008).

The efficiency of selection would be expected to be lower for AD and AUFPC, as these traits appear to be more responsive to environmental conditions; possibly due to a more complex genetic architecture (Xu and Crouch, 2008). The simultaneous inclusion of environmental variables together with molecular marker data into statistical models offers an approach towards understanding genotype by environment interactions and in turn improve levels of selection efficiency (Yin et al., 2003).

3.4.2 Phenotypic correlations

Within the genotype panel a number of significant correlations were observed amongst floral traits and also between floral and agronomic traits (Figure 3.7 and Table 3.4). Some of these correlations have been previously reported in the literature and are discussed below. It would be remiss to not emphasise that care must be taken in the interpretation of results such as these from diverse panels, as correlations between traits may simply be an artefact of selection history and not necessarily imply a causative relationship. Confirmation of genetic linkage between traits would necessitate either work in bi-parental mapping populations or genome wide association studies with adequate statistical power to account for the co-inheritance of traits.

3.4.2.1 Anther extrusion and associated traits

Within European germplasm, correlations between plant height and anther extrusion capacity have been described in recent studies; Boeven et al. (2016) and Langer et al. (2014a) reported correlations of r = 0.22 and 0.27 respectively. In the present study the strength of the correlation between these two traits was found to be slightly higher (r =0.45).

The authors of the aforementioned studies attribute these observations to the *Rht-1* dwarfing loci exerting an influence on floral characteristics that influence anther extrusion and ultimately pollen shedding capacity. This assertion is supported by co-localization of the *Rht-B1* dwarfing locus and and an anther extrusion QTL (Lu et al., 2013) and relationships reported for filament length and the number of dwarfing genes (Beri and Anand, 1971).

The results of past investigations suggest a pleiotropic effect of the *Rht-1* dwarfing loci on stamen development. The *Rht-1* loci exhibit an orthologous relationship to *SLN1* which in barley has been demonstrated to act as a repressor of gibberellic acid (GA) induced genes, including the anther development gene *HvGAMYB* (Gómez-Cadenas et al., 2001); application of exogenous GA has been found to induce higher expression of *HvGAMYB* and the repression of *SLN1* (Murray et al., 2003). Further to this in *A*. *thaliana*, anther filament elongation is known to be promoted by GA and repressed by DELLA proteins, which show an orthologous relationship to the *T. aestivum Rht-1* gene products (Cheng et al., 2004).

Murray et al. (2003) proposed that *HvGAMYB* plays a key role in anther development in barely; expression of the gene is observed in the epidermis, endothecium, middle layer and tapetum during early anther development; sustained

expression is observed in the epidermis throughout development.

Taking into consideration the high level of conservation reported between the *T. aestivum* and *H. vulgare* GAMYB genes (Haseneyer et al., 2008); it is plausible that a similar interaction occurs in bread wheat, wherein the *Rht* loci plays an orthologous role to that of *SLN1* in barley. A model can be proposed to provide an explanation for the association of the dwarfing (GA insensitive) alleles of the *Rht-1* loci with greater anther retention; in GA insensitive mutants anther and filament elongation is repressed during early development by the over expression of DELLA proteins.

Boeven et al. (2016) and Langer et al. (2014a) also reported anther extrusion as being negatively correlated with heading date; r=-0.51 and -0.63 respectively; i.e. earlier heading is associated with higher anther extrusion. In the present study a significant negative correlation between the two traits was observed (r = -0.44). To date no research has been carried out to determine whether earlier flowering genotypes tend to extrude more anthers as compared to their later flowering counterparts.

However, the results of the present study appear to suggest that this is not the case; despite the 2015 trial being characterised by much later heading dates, anther extrusion

was on average higher during this trial and the distribution of phenotypic values was towards higher anther extrusion (Figure 3.1 A and 3.2 C). Thus it does not appear that earlier flowering is pre-requisite for a high anther extrusion capacity.

Within the present study, no significant correlation between anther extrusion and the number of spikelets per spike was observed. Boeven et al. (2016) and Langer et al. (2014a) both reported significant, albeit weak, negative correlations between the two traits (r= -0.15 and -0.48 respectively). The presence of this relationship between two traits is supported by earlier reports that lax headed wheats, i.e. with less spikelets per unit of spike length, tend to exhibit greater separation angles between the palea and lemma during anthesis (Johnson and Schmidt, 1968). Within our panel, variation in spike length concurrent to variation in the number of spikelets per spike may explain why no significant correlation was detected as in these past studies. Phenotypic data for spike length was however not collected.

The present study did detect significant negative correlations between anther extrusion and the number of grains per ear (r= -0.49) and grains per spikelet (r= -0.40). Assuming that these two traits provide an estimate of the number of fertile, and therefore fully formed florets within the spike/ear; these correlations could be attributed to anther extrusion in the

primary and secondary florets being constrained by the presence of a greater number of higher order florets within individual spikelets. De Vries (1971) has previously described that compromised floret gaping is associated with a greater number of higher order florets.

3.4.2.2 Anthesis profiles and associated traits

As anticipated the area under the flowering progress curve (AUFPC) was strongly correlated with anthesis duration; r=0.8. This suggests that selection for longer anthesis duration may be adequate in attempting to breed males with anthesis profiles most likely to cover the period of receptivity in female parents, i.e. profiles with high levels of synchronised anthesis sustained for a relatively long period of time.

However, the negative relationship observed between anthesis duration and heading date (r = -0.58) questions the feasibility of selecting for longer anthesis duration and ultimately higher AUFPC values. The 2015 trial (Section 3.3.1.1) was characterised by later heading and shorter anthesis durations (Figure 3.1 C); suggesting that genotypes with long anthesis durations achieve this by flowering earlier in the season when milder temperatures may enable slower completion of anthesis. The absence of any significant correlation between the length of the flowering period and

heading date in the study carried out by Langer et al. (2014a) may be attributable to a narrower range of heading dates or the fact that their assessments of flowering duration were based on visual scoring.

Interestingly, no major difference in the range of AUFPC values was observed between the trials; apart from the absence of outliers in the 2015 trial (Figure 3.1 D). The negative correlation between AUFPC and heading date (r=-0.36) observed is likely an artefact of the relationship between AUFPC and anthesis duration described previously. This supports the validity of selecting for anthesis profiles, one approach being the AUFPC, and suggests that screening for this trait would need to be done in addition to or instead of anthesis duration. With the advent of crop imaging platforms designed for field trials, it is likely that more high throughput and accurate approaches will be developed towards screening for this trait.

3.4.2.3 Anther length and associated traits

For anther length, as phenotypic data were only collected in the 2015 trial, correlations with other traits cannot be conclusively stated. However, anther length was found to be positively correlated with both anthesis duration and AUFPC (r = 0.36 and 0.43 respectively). This relationship is of interest

as physiological links between the length of the vegetative growth phase and anther have previously been described (Komaki and Tsunewaki, 1981, Jôst et al., 1976). In theory, while not explicitly within the vegetative phase, having a slower progression of anthesis could enable a genotype to achieve longer anthers and potentially produce more pollen grains per anther.

3.4.3 Extent of genotypic variation for traits assessed (BLUEs)

In designing the panel, an attempt was made to minimise variation with respect to heading date in order to assess the floral characteristics of all genotypes under relatively similar climatic conditions. Approximately 70% of the panel headed within a five-day period between 29 and 34 days from May 1st (Figure 3.5 C). Later heading predominated within the panel, with few genotypes exhibiting very early heading dates. The genotype panel was compromised of predominantly semi-dwarf varieties; with approximately 70% of the panel growing to a height of 70-80 cm (Figure 2.4 D).

For all remaining traits a considerable amount of phenotypic variation was observed within the panel (Figure 3.4, 3.5 and 3.6). The results of the present study and those of recent investigations suggest that it it is possible to select for floral traits within European elite germplasm (Muqaddasi et al.,

2017, Muqaddasi et al., 2016, Boeven et al., 2016, Langer et al., 2014a). This will be a key step towards the development of efficient hybrid wheat breeding programmes in Europe.

Furthermore, despite the general trends observed across the panel, for some genotypes variation in floral characteristics appear to be uncoupled from agronomic traits that could affect the adaptability of genotypes to specific environments. For instance, a high capacity for anther extrusion was not entirely restricted to taller or early heading genotypes. Similarly, later heading genotypes in some instances exhibited anthesis durations and areas under the flowering progress curve values that were comparable to those that headed much earlier.

3.4.4 Heritability of anther extrusion across locations

Across four locations, a similar range of phenotypic values for the number of anthers extruded per floret (AE) was observed with our smaller panel of 24 genotypes (Figure 3.8 A). Anther extrusion phenotypic values from the Thriplow (UK) trial was found to be a reliable predictor of performance in the other three locations; with *r* ranging from 0.88–0.93 (Figure 3.8). This observation supports the feasibility of exchanging male parental lines between environments, at least within Europe,

and suggests that the development of male pools can be carried out outside of the target environment if needed.

Further to this, the variance explained by the interaction between genotype and location, while significant, was considerably smaller than that explained by genotype (Table 3.2). The effective exchange of male parental lines would require appropriate agronomic characteristics be present for the new environment, such as heading date and plant height. The availability of perfect markers tagging the causative loci governing these traits can help facilitate this process.

The heritability of AE observed with this panel was found to be higher than that observed with the larger panel phenotyped in 2014 and 2015; 0.93 vs 0.80 respectively. This could be either be attributable to greater precision of field trials, owing to the relatively smaller number of lines, or more likely the fact that genotypes all headed within a relatively smaller time frame than in the larger field trials. The high heritability observed in these multi-location trials further supports the effectiveness and accuracy of the approach taken in the present study towards assessing anther extrusion capacity.

3.4.5 Assessment of phenotyping strategies

Phenotyping approaches for floral traits applied in the current study appear to be adequately robust and accurate to detect

genotypic variation; for each trait significant genotypic variance was detected with modest to high heritability.

Further, the feasibility of applying these methods in field trials has been demonstrated. In the larger field trials, 222 plots were screened for all target traits within a period of a month without any apparent compromise to accuracy. With specific reference to anther extrusion capacity, assessments could be carried out at multiple locations within the same field season, due to the ability to store ears for later assessment and the relative speed at which phenotyping can be carried out (five minutes per ear).

However as mentioned previously, there exists a need for higher throughput approaches towards screening for anthesis duration and profiles. For both anthesis duration (AD) and the area under the flowering progress curve (AUFPC), the approaches taken in the present study were associated with higher trait heritability (0.5 and 0.6 respectively) as compared to that of the flowering duration approach taken by Langer et al. (2014a) which was based on visual scoring (0.31). These AD and AUFPC methods however require daily presence in the field and imposed a constraint on the number of additional traits that could be screened. Viable alternatives that can potentially achieve greater accuracy and are more amendable to high throughput screening, include adhesive surfaces for

assessment of pollen shedding over time or automated imaging platforms capable of discerning ears that are at anthesis.

3.5 Summary

Phenotyping methods, targeting anther extrusion capacity, anthesis duration, anthesis profile and anther length, were successfully applied to a large diverse panel of European winter wheat varieties in two field trials. The correlations observed between phenotypic values from each trial, together with the heritability observed for these traits suggest that phenotyping was carried out with adequate accuracy and phenotypic selection is feasible for these traits. In the case of anther extrusion capacity, multi-location trials using a small panel confirmed conclusions drawn from the larger panel. Nevertheless, it is clear that it is possible to improve upon applied phenotyping approaches and to develop higher throughput methods for screening larger populations.

The significant genotypic variance observed for each trait supports the feasibility of selecting for these traits. Areas that require further research include the relationship between floral traits and agronomic traits, an understanding of which is prerequisite for the effective exploitation of existing diversity. Towards the identification of genomic regions associated with variation in these traits a genome wide association study using the phenotypic data described in this section is presented in the following chapter.

CHAPTER 4: AN ASSOCIATION MAPPING STUDY TARGETING FLORAL TRAITS

4.1 Introduction

The effective exploitation of the genotypic variation in floral traits described in Chapter 3 can be facilitated by identification of genomic regions and causative loci that underlie observed variation. Association mapping offers a feasible approach towards achieving this goal, as it enables multiple traits to be simultaneously investigated while potentially offering a relatively high level of genomic resolution (Yu and Buckler, 2006).

The modest-high heritability of traits and the large amount of phenotypic diversity observed suggest that data described in Chapter 3 would be appropriate for use in a genome wide association study. The panel of 111 genotypes used in the field trials was therefore genotyped using the iSelect 90K SNP array for which a high density SNP map has been published (Wang et al., 2014).

Efforts towards identifying genomic regions associated with variation in anthesis duration, anther length, area under the flowering progress curve (AUFPC) and anther extrusion using association mapping are described in the current chapter. The applicability of the findings of the present study towards development of marker assisted selection targeting floral traits is discussed.

4.2 Methods

4.2.1. Phenotypic data

Phenotypic data (BLUEs & BLUPs) from the panel of genotypes described in Section 3.2.1 were used to carry out a genome wide association study (GWAS). Marker trait associations were investigated for the following traits; the number of anthers extruded per floret (AE), anthesis duration (AD), the area under the flowering progress curve (AUFPC), anther length (AL), days to heading from May 1st (HD), plant height (HT), spikelets per spike (SP) and awn presence/absence (AWNS).

4.2.2 Genotyping

Genotypic data was generated by single nucleotide polymorphism (SNP) analysis on the Illumina Infinium iSelect 90,000 SNP wheat array and made available by KWS UK Ltd. Wang et al. (2014) previously published the development of this platform and its corresponding consensus map. Additionally, the panel was genotyped using diagnostic markers for *Rht1*, *Rht2* and *Ppd-D1* and the presence of the 1BL.1RS translocation. This gave 26,008 SNPs that were polymorphic within the association panel. Markers with >10% missing data or a minor allele frequency (MAF) < 5% were then excluded from analysis.

Of the remaining 23,233 SNPs, those for which a map position was not available within the consensus map were excluded. For the association mapping study 17,608 markers were retained after filtering (Table 4.1); imputation of marker scores was then carried out using the software Tassel; with genotypic data first being converted to numerical format and imputation by mean then carried out.

Table 4.1: Distribution of 17,608 mapped SNP markers across the three bread wheat genomes (A,B,D) and groups.

Chromosome	Genome			Total by group
	A	**B**	**D**	
1	1084	1912	599	3595
2	972	1302	561	2835
3	899	1226	256	2381
4	610	577	79	1266
5	1033	1599	200	2832
6	1090	1103	206	2399
7	1246	881	173	2300
Total by genome	6934	8600	2074	
Total				17608

4.2.3 Linkage disequilibrium and population structure

The correlation coefficient (r^2) between marker loci pairs was calculated as a measure of linkage disequilibrium (LD) using the R package 'popgen' (Marchini and Marchini, 2013) within the R environment. Plots of r^2 against genetic distance were generated in order to investigate LD decay within the association panel; this was done for all three genomes collectively, as well as for each individual genomes.

The critical value of r^2 estimated using the procedure described by Breseghello and Sorrells (2006) by taking the 95% percentile of the square rooted r^2 distribution of unlinked marker pairs (inter-chromosomal). A second degree Loess (locally weighted scatter plot smoothing) curve was fitted in R; the point (genetic distance) at which this curve intercepted the critical value of r^2 was used as the estimate of significant LD decay.

A single marker from each pair displaying an absolute correlation coefficient (r^2) >0.9 was excluded, giving 4695 markers which were subsequently used in the generation of a marker-based Losielle kinship matrix (Loiselle et al., 1995). The number of markers retained was 1961, 2124 and 610 for the A, B and D genomes respectively. This filtering was

carried out in order to prevent the biasing of the matrix by groups of tightly linked markers within small genomic regions. The extent of population structure was assessed using principal coordinate analysis (PCA) using markers at least 1 cM apart. The number of markers retained was 3566 in total, with 1511, 1894 and 161 being localized to the A, B and D genomes respectively. The proportion of variance explained by each PC was plotted in order to determine the optimum number of PCs required to adequately control for population structure.

4.2.4 Association Mapping

Association mapping (AM) was carried out using a mixed linear model in efficient mixed-model association (EMMA) within the genome association and prediction integrated tool (GAPIT) within the R environment (Lipka et al., 2012).

4.2.4.1 Model testing

Three different AM models were tested; 1) a naïve model with no correction for kinship or population structure 2) a kinship model into which the kinship matrix generated from the subset of 4695 markers was incorporated and 3) a full model into which PCs were included as covariates along with the kinship matrix.

QQ plots were generated for each model in order to assess the extent to which each controlled for the rate of false positives. Subsequent to model testing, results from the model which provided the most effective control for the rate of false positives are reported in the present study.

4.2.4.2 Association mapping for agronomic and floral traits

AM was conducted with the selected model using BLUEs across the two trials and BLUEs from each individual trial. The threshold of Bonferroni correction for multiple testing displayed on plots was obtained by dividing 0.01 and 0.05 (*P*) by the number of markers used in this association study.

Association mapping for awn presence/absence and spikelet number was carried out with the aim of ascertaining whether the rate of LD decay within the genotype panel was sufficient for the detection of significant MTAs. Strong MTAs have been reported previously within the literature from studies applying the same genotyping platform, as in the present study, on larger European winter wheat panels (Mackay et al., 2014, Boeven et al., 2016).

4.2.4.3 Association mapping using functional marker scores

Additionally, the marker scores for each perfect marker were treated as a phenotype and AM carried out with the naïve model. This was carried out in order to ascertain whether any SNPs co-localized with the *Rht1*, *Rht2* and *Ppd-D1* loci and also to provide some insight into the suitability of the genotype panel, the distribution of SNPs and their corresponding map for AM. Marker-marker associations detected were then compared to those reported previously using the iSelect 9k or 90k SNP platforms (Sukumaran et al., 2016, Lopes et al., 2015).

4.2.4.4 Identification of candidate floral genes within high LD regions

For each floral trait, SNPs localizing to the same chromosome and exhibiting high LD with the respective MTA ($r^2 > 0.3$) were extracted. These SNPs were assigned a physical position their respective chromosome using their best alignment with the IWGSC NRGene assembly (RefSeqv1.0). The TGACv1 gene models, aligned to the RefSeqv1.0, that overlapped each SNP was then determined. SNP and gene model alignments were made available by Ricardo Ramirez-Gonzalez (John Innes Centre, UK) and Gemy Kaithakottil together with David

Swarbreck (Earlham Institute) respectively via www.wheat-training.com.

From the list of genes obtained, their corresponding transcript stable IDs were entered into http://www.wheat-expression.com to retrieve their expression levels in the spike, using the log2 transcripts per million (tpm) scale, from the developmental time-course of Chinese Spring dataset (Borrill et al., 2016).

Any transcripts showing expression within the spike were identified for subsequent investigation by bulking the expression profiles across all developmental stages. In order to mitigate against the inherent limitations of applying such a filter, expression profiles for the three groups of above ground tissues, the spike, leaves and the shoot and the mean expression at each developmental stage was extracted from the dataset. Further, the spike developmental stages within the dataset included stages relevant to anther development.

Two-way repeated-measures analysis of variance (ANOVA) were then carried out to determine significant differences between the expression profiles of each transcript with differences between means being determined by Tukey's HSD test at a significance threshold of $P<0.05$.

Any transcript exhibiting higher expression in the spike as compared to at least one of the two remaining tissue types

was listed as a putative candidate. *Arabidopsis thaliana* genes orthologous to these candidates were obtained using the EnsemblPlants Blast Server database (http://plants.ensembl.org/) and their expression patterns and biological role obtained using the UniProt database (http://www.uniprot.org).

4.3 Results

4.3.1 Extent of population structure and linkage disequilibrium

A principal component analysis was run using a reduced marker dataset, in order to investigate the population structure within the association panel. The first 10 eigenvectors (PCs) from PCA explained 45.08% of genetic variation within the association panel; with the first three accounting for 8.08%, 5.94% and 5.72% (Figure 4.1).

From the sharp drop in percentage explained variance observed between PC3 and PC4 in the scree plot (Figure 4.1), it was concluded that inclusion of the first three PCs would adequately correct for population structure.

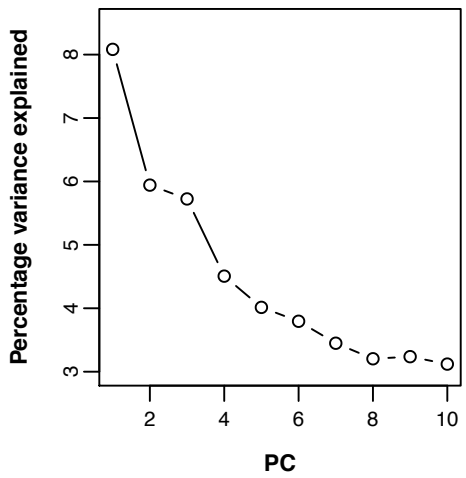

Figure 4.1: Scree plot showing proportion of explained variance of the first ten principal coordinates.

Separation of genotypes into distinct sub groups was not observed when principal coordinates (PCs) one to three were plotted against each other (Figure 4.2). Genotypes possessing the 1BL.1RS translocation were observed to cluster together to a small extent.

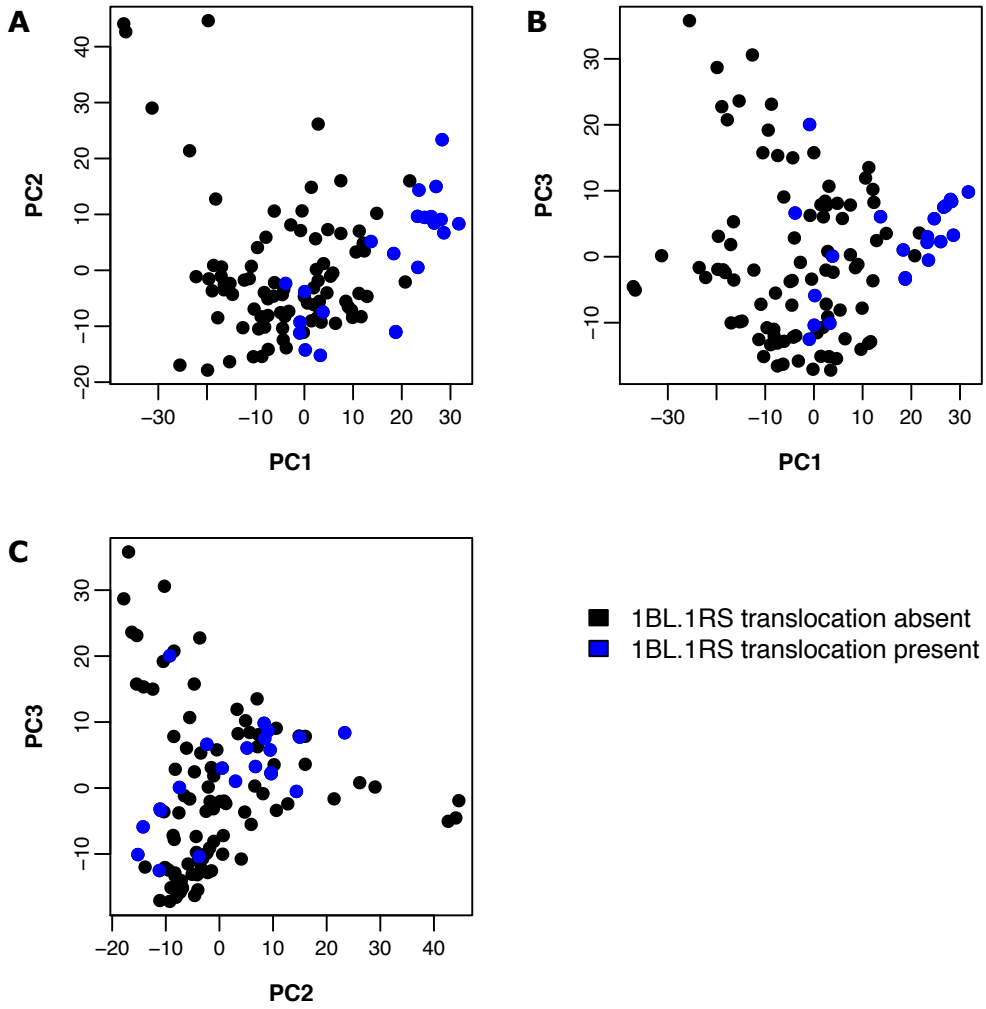

Figure 4.2: Population structure of the panel of 111 winter wheat varieties as determined by PCA (coloured by the presence/absence of the 1RS translocation), showing PC1-3 plotted against each other.

No clear pattern was observed when country of origin was taken into consideration; genotypes originating from England were found scattered throughout the plots (Figure 4.3).

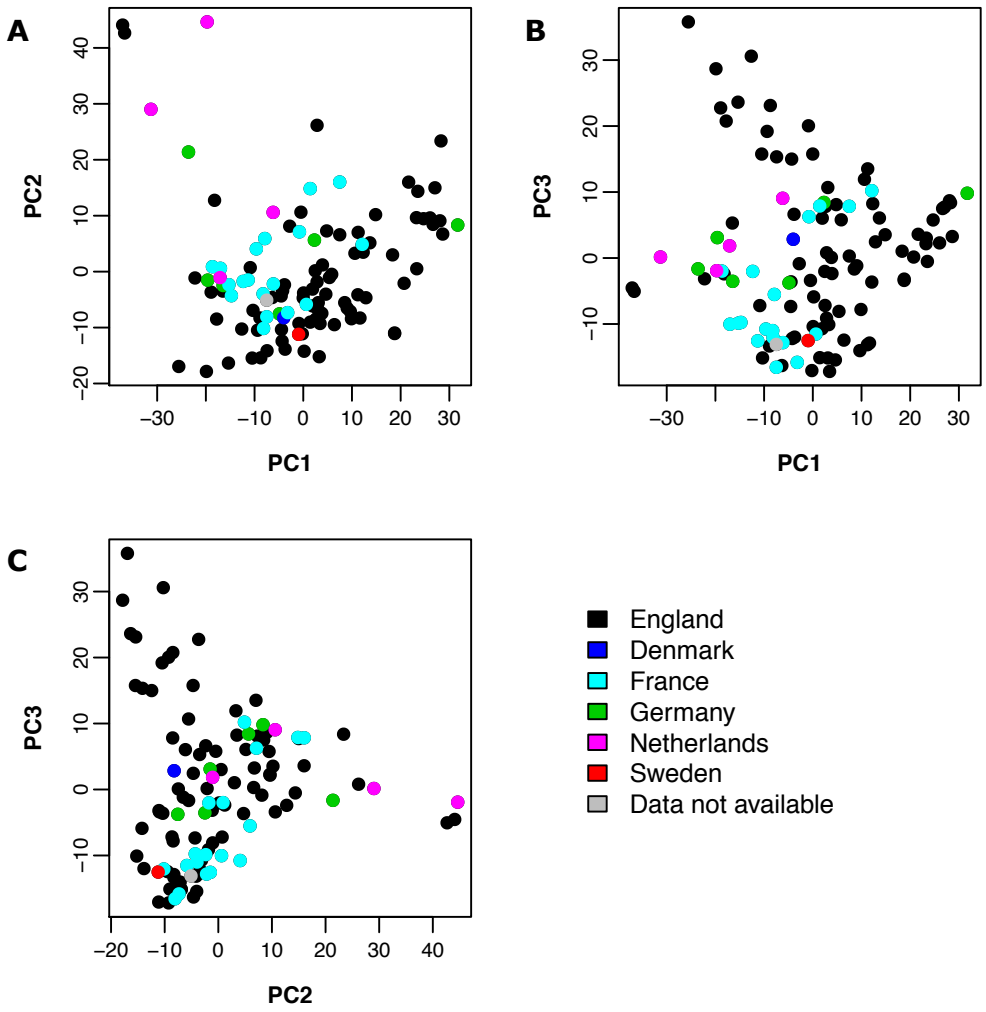

Figure 4.3: Population structure of the panel of 111 winter wheat varieties as determined by PCA (coloured by country of origin), showing PC1-3 plotted against each other.

In the plot of PC1 vs PC2 the detached genotypes towards the upper left corner, included the facultative wheat variety Axona and its descendants; Cadenza and XI-19 (Figure 4.4).

Robigus and its descendants clustered closely together within each PCA plots (Figure 4.4) and were detached from the main cluster in plots of PC1 vs PC3 and PC2 vs PC3.

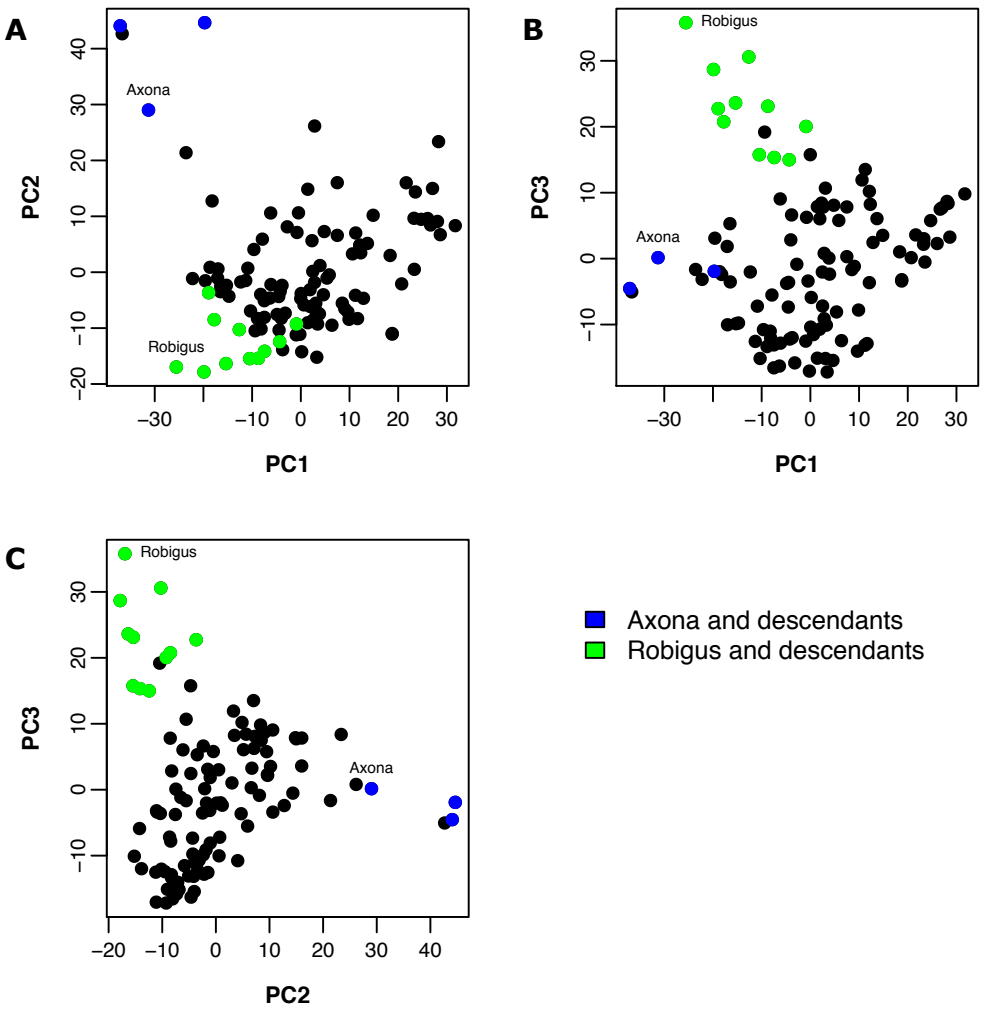

Figure 4.4: Population structure of the panel of 111 winter wheat varieties as determined by PCA, showing PC1-3 plotted against each other. Genotypes have been coloured based on their pedigree.

Pairwise LD (r^2) was calculated between all inter-chromosomal and intra-chromosomal combinations of markers. From the 95% percentile of the distribution of unlinked r^2 (intra-chromosomal) the critical r^2 value was estimated for each genome and the genome as a whole. Plots of genetic distance showed significant LD to extend to 23 cM based on the critical value of r^2 (Table 4.2 & Figure 4.5).

Table 4.2: Overview of LD in wheat genomes A, B and D.

	Critical r^2	**Extend of LD for critical r^2 in cM**	**Extend of LD for r^2 =0.2 in cM**
A Genome	0.056	17	4.0
B Genome	0.058	17	3.0
D Genome	0.050	23	9.5
All Genome	0.059	16	4.5

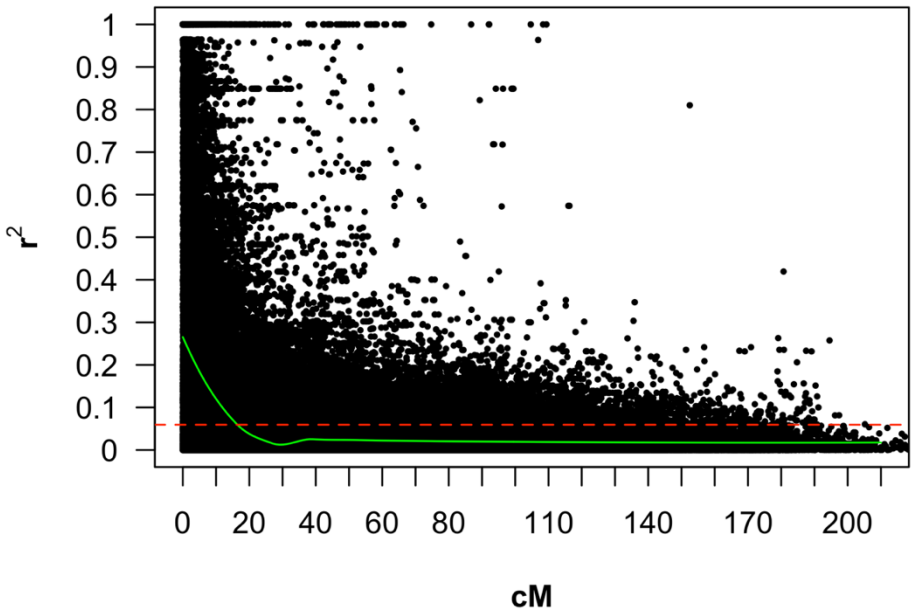

Figure 4.5: LD (r^2) plotted against genetic distance (cM) for all three genomes.
The horizontal dashed red line shows the critical r^2 value derived from the 95% percentile of the distribution of unlinked r^2 values. The green curve is the second degree LOESS curve fitted to the dataset.

Variation in the rate of LD decay with genetic distance was observed between the genomes; being fastest in the A and B genomes and slowest for the D-genome, in which significant LD extended to 23 cM (Figure 4.6 and Table 4.2).

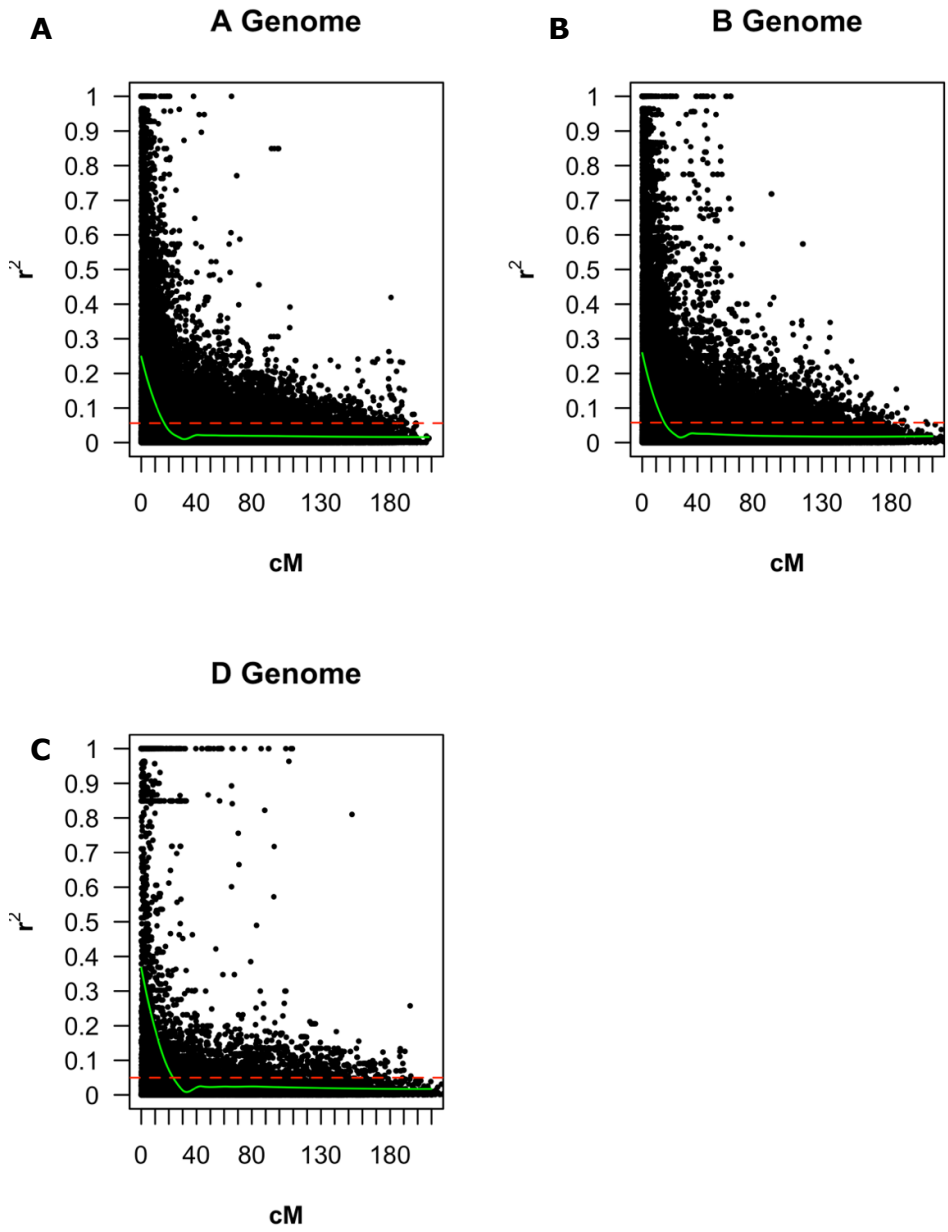

Figure 4.6: LD (r^2) plotted against genetic distance (cM) for each of the three genomes.
The horizontal dashed red line shows the critical r^2 value derived from the 95% percentile of the distribution of unlinked r^2 values. The green curve is the second degree LOESS curve fitted to teach dataset.

4.3.2 Association Mapping

4.3.2.1 Model Testing

Without population structure and kinship, *p*-values are anticipated as being uniformly distributed between 0 and 1 with only true associations leading to deviations from the diagonal line. The high proportion of *p* values above the diagonal observed for the majority of traits using the naïve model indicated an overestimation of *p* values.

The kinship, PC and full model all gave improvements as compared to the naïve model. The full model was most effective at controlling the rate of false positives for the majority of traits; with observed *p*-value distributions more closely following the expected distribution of values. Additionally, the full model considerably reduced the number of MTAs above the declared significance thresholds.

The full model was not adequate for two traits; AWNS and HT. An alternative kinship matrix was applied to both traits, in this case generated using the VanRaden algorithm, together with the first three PCs. This model was referred to as the full model_2, and was effective at reducing the number of associations declared significant.

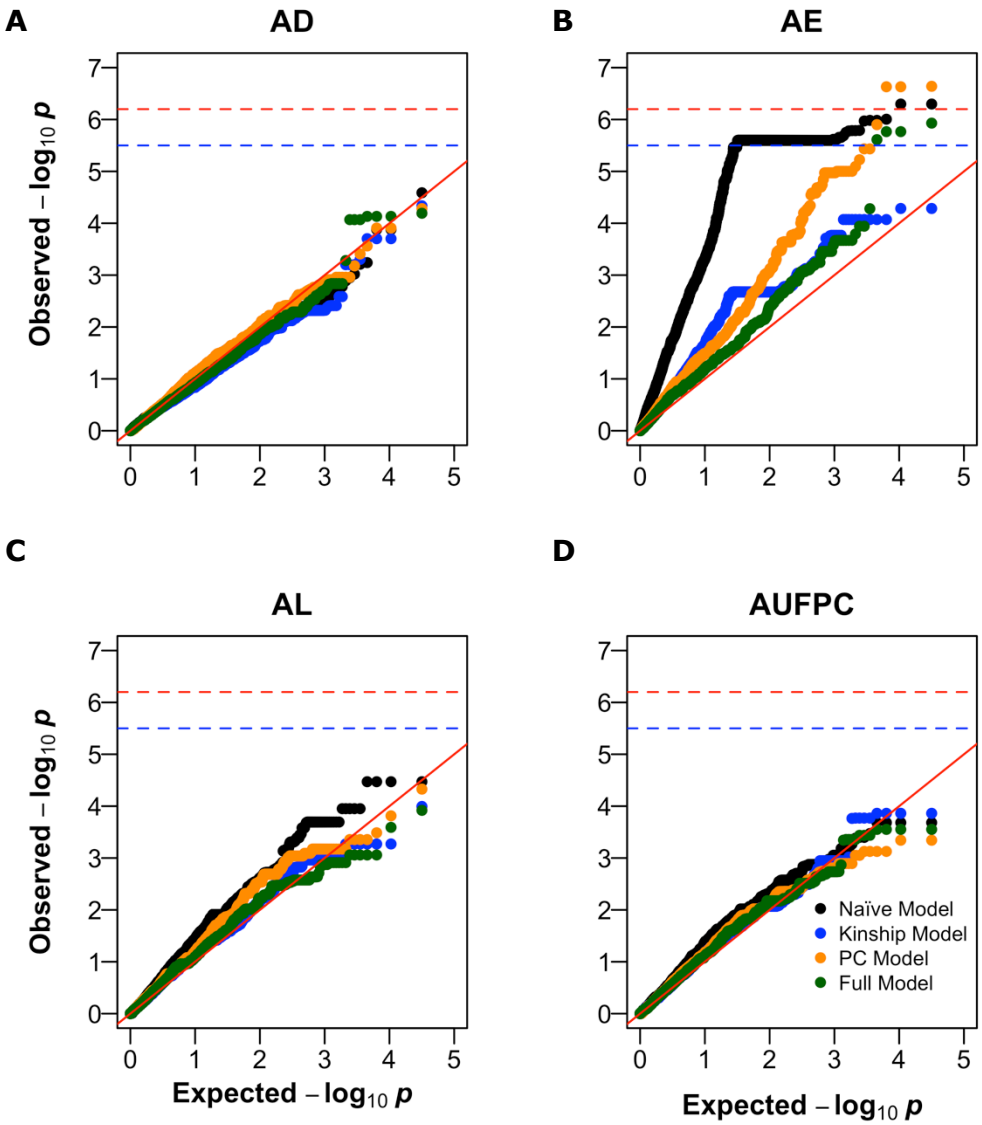

Figure 4.7: Q-Q plots of $-\log_{10} p$ values obtained from the AM models applied to anthesis duration (AD); number of anthers extruded per floret (AE); anther length (AL) and area under the flowering progress curve (AUFPC).

The red and blue dashed lines indicate the threshold for Bonferroni correction at α 0.01 and 0.05 respectively.

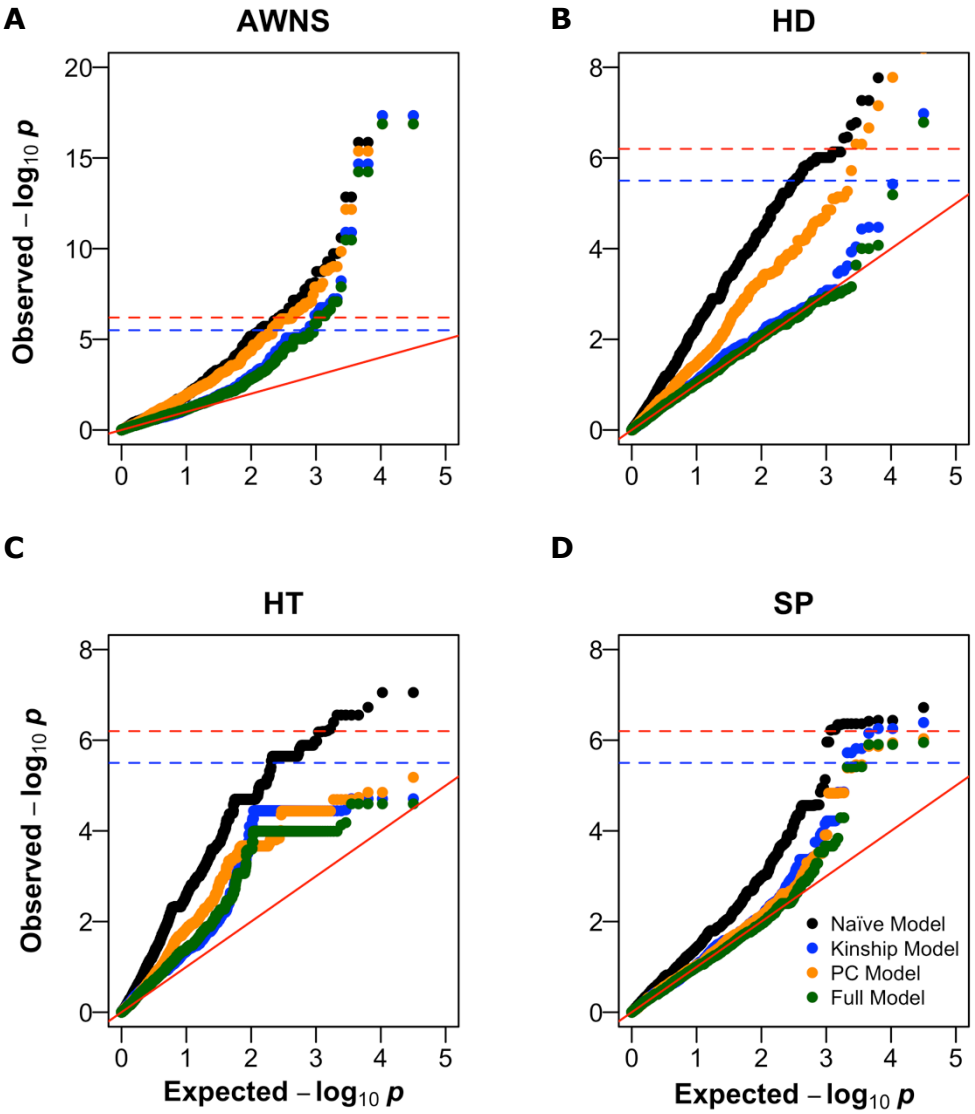

Figure 4.8: Q-Q plot of $-\log_{10} p$ values obtained from the AM models applied to awn presence/ absence (AWNS); days to heading (HD); plant height (HT) and the number of spikelets per spike (SP).

The red and blue dashed lines indicate the threshold for Bonferroni correction at α 0.01 and 0.05 respectively.

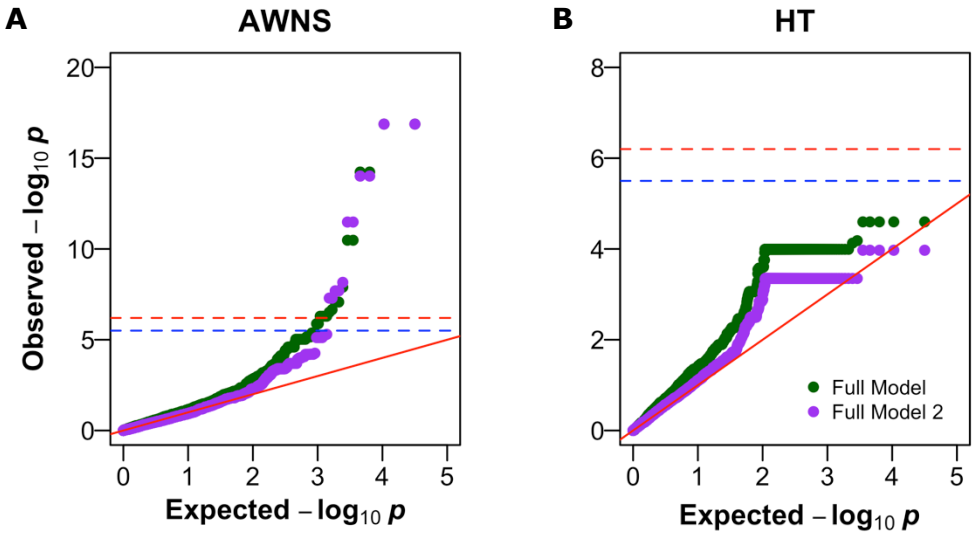

Figure 4.9: Q-Q plot of $-\log_{10} p$ values obtained from the AM models applied to awn presence/ absence (AWNS) and plant height (HT).

The red and blue dashed lines indicate the threshold for Bonferroni correction at α 0.01 and 0.05 respectively.

Table 4.3: Number of SNP markers with marker trait associations above defined thresholds and -log10 (*p*-value) for strongest MTA traits assessed using the four models tested.

Trait	-log10 (*p*-value)		No. of markers above threshold			
			α=0.05		α=0.01	
	Naïve model	Full model	Naïve model	Full model	Naïve model	Full model
AD	4.58	4.19	0	0	0	0
AE	1.20	5.93	510	4	2	0
AL	3.35	3.91	0	0	0	0
AUFPC	3.18	3.55	112	0	54	0
AWNS	8.87	16.87	47	11	9	11
HD	8.87	6.78	47	1	9	1
HT	6.55	3.97	78	0	11	0
SP	6.72	5.95	16	4	14	0

4.3.2.2 Localization of functional markers

An attempt was made to localize three functional markers to a chromosomal position *Ppd-D1*, *Rht1* and *Rht2*. The strongest marker-marker association (MMA) was compared to that reported in the literature by studies using the 90k or 9k iSelect SNP array (Sukumaran et al., 2016, Lopes et al., 2015).

Only *Rht1* localized to the position anticipated from prior investigations, the short arm of chromosome 4B as reported by Sukumaran et al. (2016) (54-64 cM). SNPs within this region showed a strong association with *Rht1* (Figure 4.10 B).

In the case of *Rht2* only relatively weak associations were observed on chromosome 4D where the gene is known to be present. No polymorphic SNPS were present (Figure 4.10 C) at the chromosomal position reported in the literature. Consequently, the strongest MMA on this chromosome was not at the expected position.

The strong association with *Ppd-D1* was not found on the expected chromosome where the gene is known to be present, chromosome 2D. However, the strongest association on 2D was detected at the anticipated position (Figure 4.10 A).

Table 4.4: Chromosome (Chr) and position (Pos; cM) of the strongest SNP associated with each functional marker genotyped in the association panel.
The reference to which each marker marker association (MMA) is compared is listed under Ref.

Gene	Ref	Strongest MMA on ref chr?	Strongest MMA at ref chr pos?	Best hit on expected chromosome	
				-log$_{10}$ (p-value)	Pos (cM)
Ppd-D1 Chr 2D	Sukumaran et al. (2016)	NO	NO	2.85	19
Rht1 Chr 4B	Sukumaran et al. (2016)	YES	YES	12.37	56
Rht2 Chr 4D	Lopes et al. (2015)	NO	NO	4.4	54

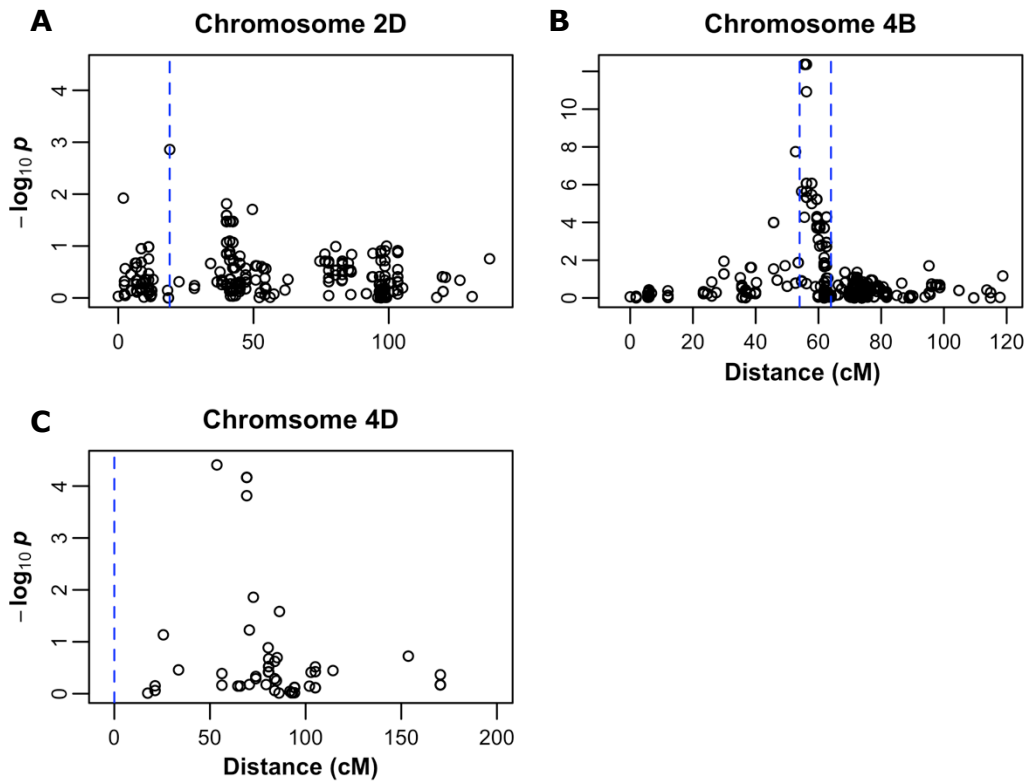

Figure 4.10: Manhattan plots for individual chromosomes of SNP markers associated with functional marker scores on chromosomes **A)** 2D **B)** 4B and **C)** 4D.
The blue dashed line indicates the position of the peak marker-marker associations reported by Sukumaran et al. (2016) or Lopes et al. (2015).

4.3.2.3 Marker trait associations (MTAs) detected for agronomic traits

Awn presence /absence and spikelet number

For awn presence/absence, 11 SNPs were found to be significantly associated with the trait at the 0.001 level. All SNPs were located on chromosome 5A spanning a region from 129.86 to 148.3 cM (Figure 4.11 A).

For the number of spikelets per spike (SP), four SNPs were found to be significant at the 0.05 level on chromosome 7A spanning a region from 152.18 to 152.78 cM (Figure 4.11 B).

The chromosomal region within which the detected awn presence/absence associations were detected on chromosome 5A corresponds to the region detected by Mackay et al. (2014). The peak SNP for spikelets per spike was the same reported by Boeven et al. (2016), BS00026622_51.

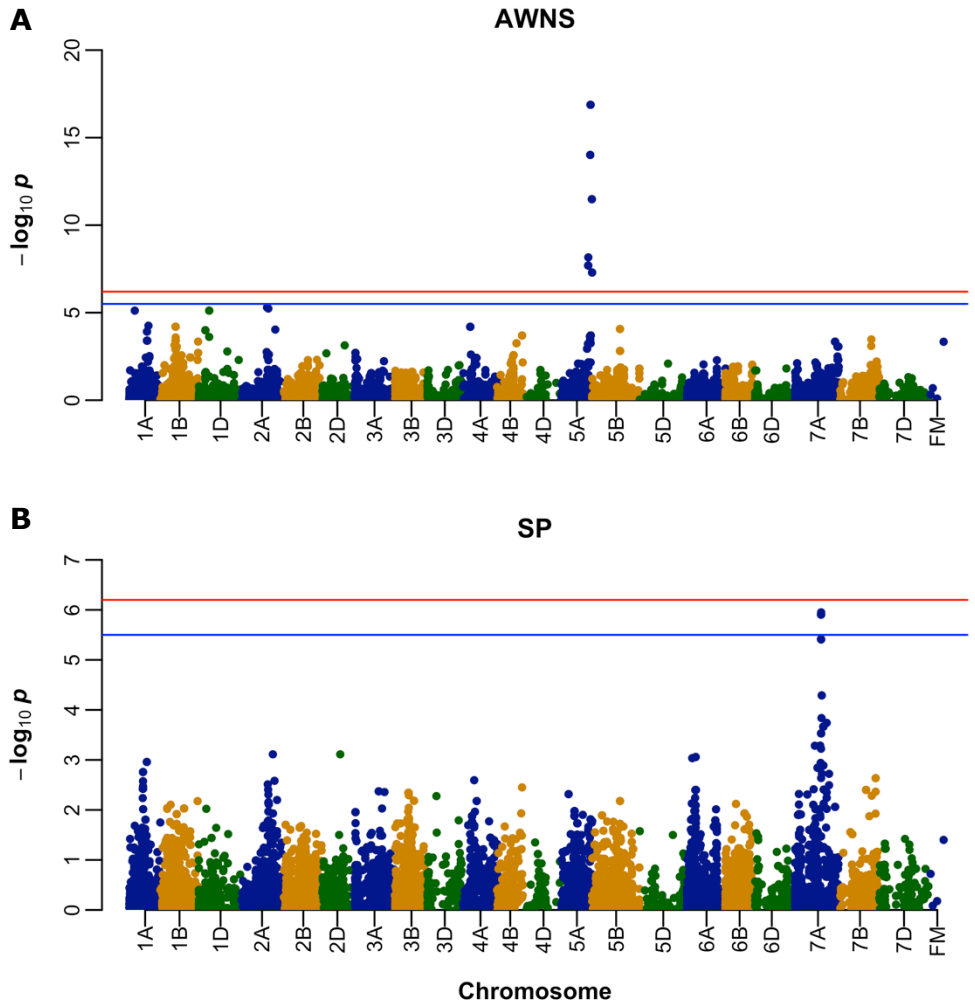

Figure 4.11: Manhattan plots of SNP markers associated with awn presence/absence (AWNS) and the number of spikelets per spike (SP). The red and blue lines indicate the threshold for Bonferroni correction at α 0.01 and 0.05 respectively. FM, functional markers.

Days to heading from May 1st

The only significant MTA with HD was the *Ppd-D1* locus; ($-\log_{10}$ (*p*-value) = 7.5) (Figure 4.12 A). A SNP on chromosome 7A fell slightly below the 0.01 Bonferroni correction threshold ($-\log_{10}$ (*p*-value) = 6.2). When *Ppd-D1* allelic status was included as a covariate in AM, the strength of association for this SNP was reduced considerably (Figure 4.12 B).

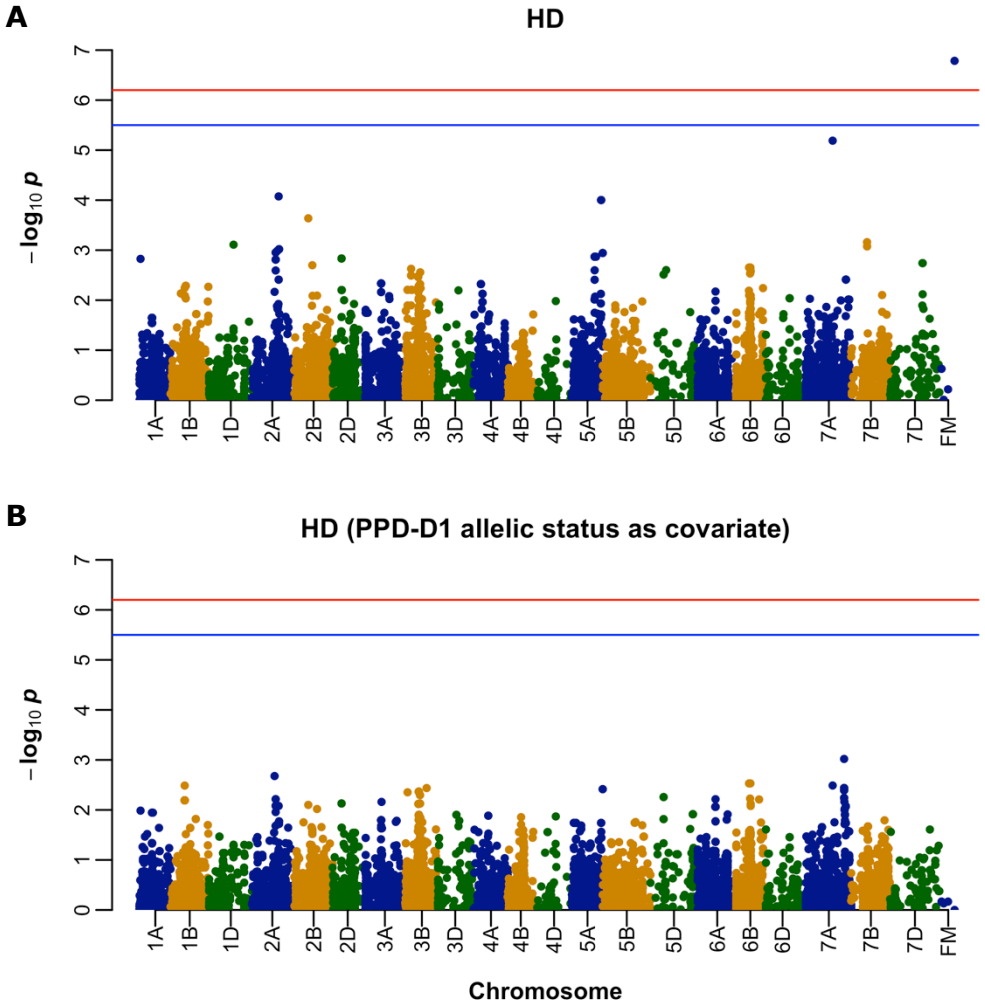

Figure 4.12: Manhattan plots of SNP markers associated with days to heading from the 1st May (HD).

The red and blue lines indicate the threshold for Bonferroni correction at α 0.01 and 0.05 respectively. FM, functional markers.

Plant height

No significant plant height MTAs were detected by AM (Figure 4.13). Genotyping was carried out for for variants at the two major *Rht* loci; *Rht1* on chromosome 4B and *Rht2* at the homoeologus position on chromosome 4D. The semi dwarf allele of *Rht2* (*Rht-D1b*) was found in 76% of genotypes within the panel. The corresponding allele of *Rht1* (*Rht-B1b*)

was more common; 90% of genotypes possessed the variant. Allelic variation at none of the *Rht* loci was significantly associated with plant height (Figure 4.13 A).

Inclusion of *Rht1* allelic status as a covariate in AM did increase the strength of association for *Rht2* (-$\log_{10}$ (*p*-value) =2.88 vs 4.4) making it the second strongest association (Figure 4.13 B).

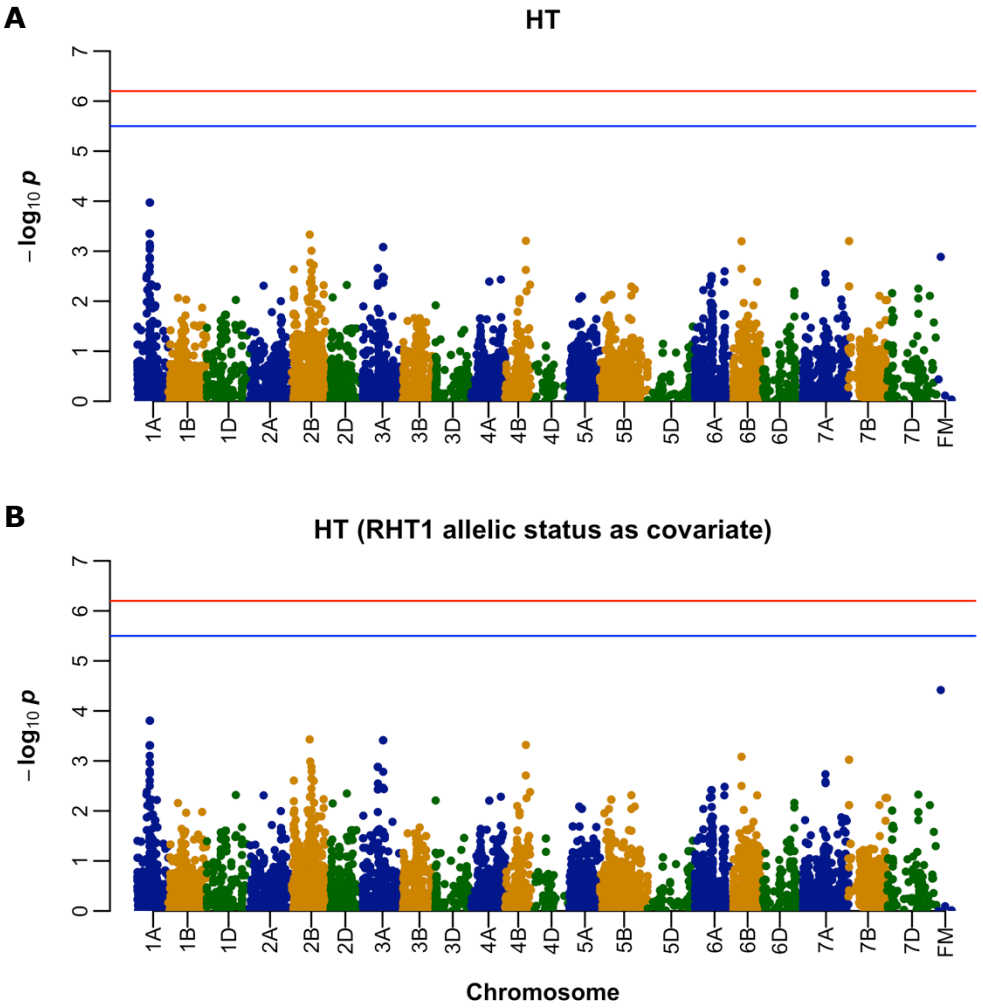

Figure 4.13: Manhattan plots of SNP markers associated with plant height (HT) ; results are presented from AM using (A) no marker covariates; (B) *Rht2* as a covariate.
The red and blue lines indicate the threshold for Bonferroni correction at α 0.01 and 0.05 respectively. FM, functional markers.

4.3.2.4 MTAs detected for anther traits

Four (4) significant SNP associations were detected for the number of anthers extruded per floret (AE) (Figure 4.14 A). Three co-localizing on 2B were designated AE_MTA1 ($-\log_{10}$ (*p*-value) = 5.61 to 5.93). The SNP on 2A, was designated AE_MTA2 ($-\log_{10}$ (*p*-value) = 6.6). No significant associations were detected for anther length (AL), association peaks were observed on 1B and 4A (Figure 4.14 B); AL_MTA1 and AL_MTA2 ($-\log_{10}$ (*p*-value) =3.91 and 3.59 respectively).

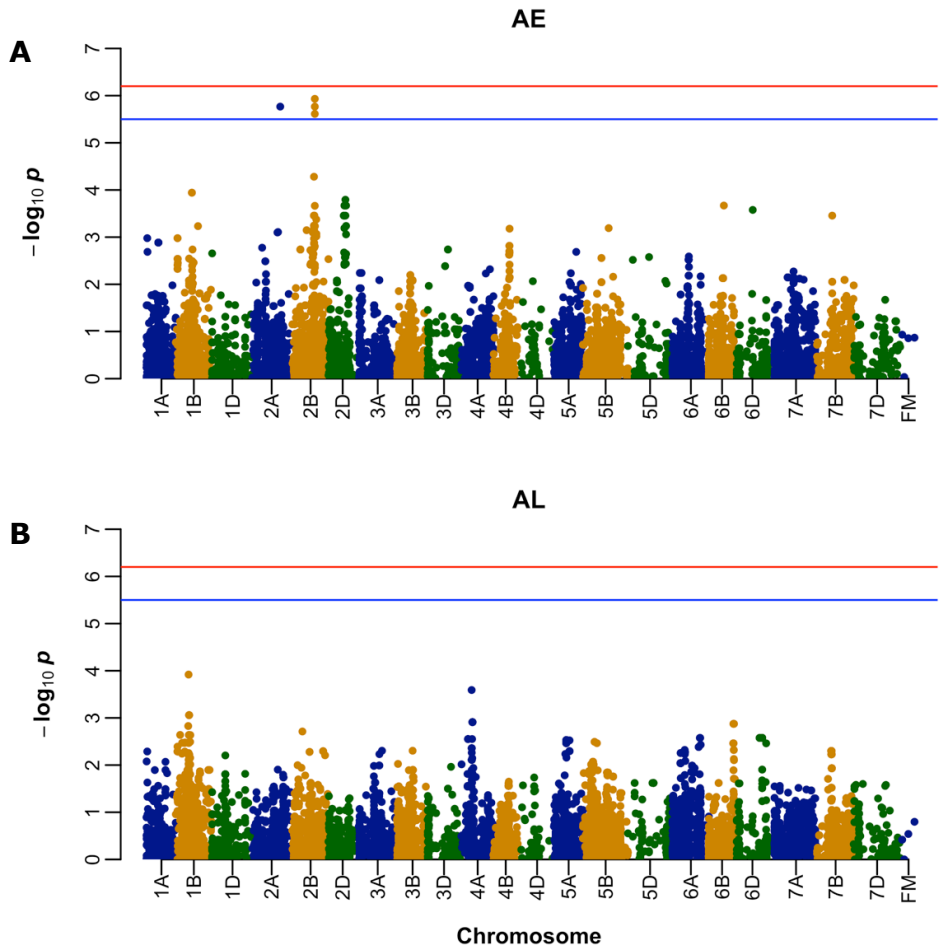

Figure 4.14: Manhattan plots of SNP markers associated with number of anthers extruded per floret (AE) and anther length (AL).
The red and blue lines indicate the threshold for Bonferroni correction at α 0.01 and 0.05 respectively. FM, functional markers.

4.3.2.5 MTAs detected for anthesis traits

No SNPs were significantly associated with anthesis duration (AD) (Figure 4.15 A). *Ppd-D1* exhibited the strongest association with AD (-$\log_{10}$ (p-value) = 4.19) (Figure 4.15 B). When *Ppd-D1* was included as a covariate, the strength of associations with AD and SNPs on 3B increased (Figure 4.15 B). Three co-localizing SNPs in this peak were designated as AD_MTA1 (-$\log_{10}$ (*p*-value) = 4.79).

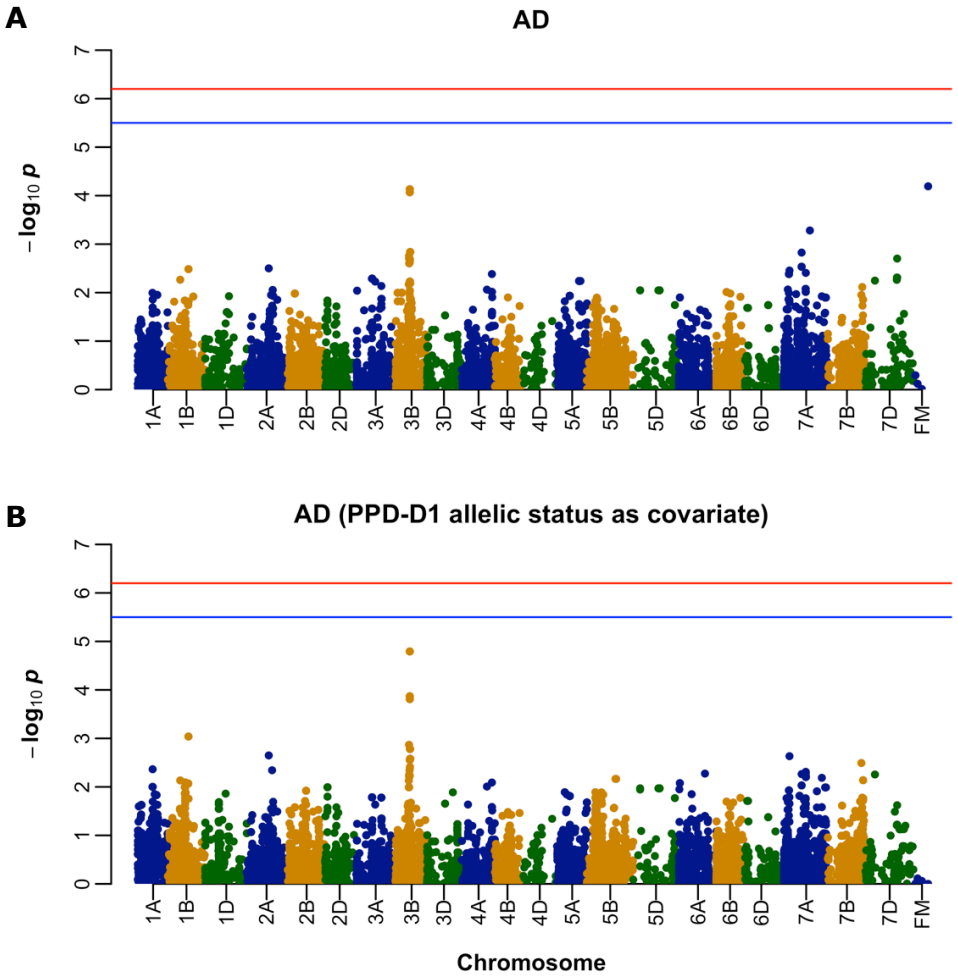

Figure 4.15: Manhattan plots of SNP markers associated with anthesis duration (AD). Results are presented from AM using (A) no marker covariates; (B) *Ppd-D1* as a covariate.

The red and blue lines indicate the threshold for Bonferroni correction at α 0.01 and 0.05 respectively. FM, functional markers.

No MTAs were declared for the trait area under the flowering progress curve (AUFPC) due to the presence of multiple weak association peaks (Figure 4.16).

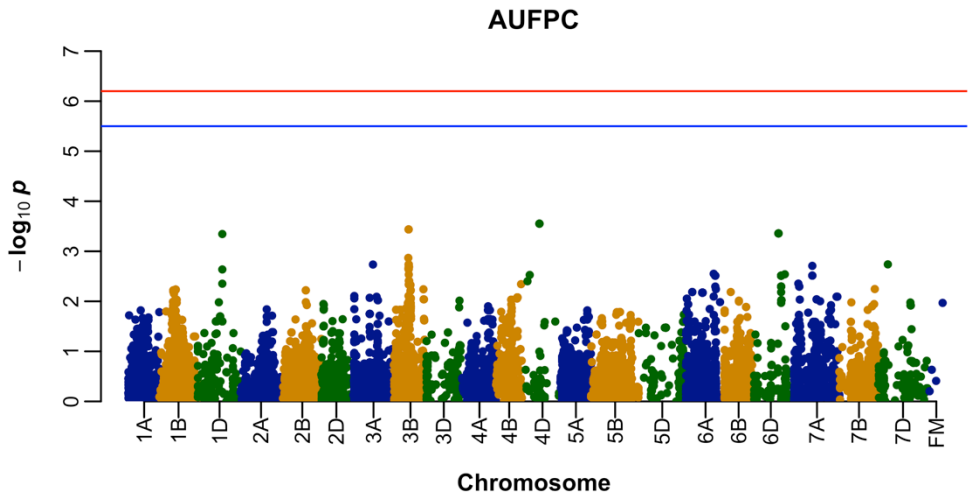

Figure 4.16: Manhattan plots of SNP markers associated with area under the flowering progress curve (AUFPC).

The red and blue lines indicate the threshold for Bonferroni correction at α 0.01 and 0.05 respectively. FM, functional markers.

4.3.2.6 Allelic effects of floral trait MTAs

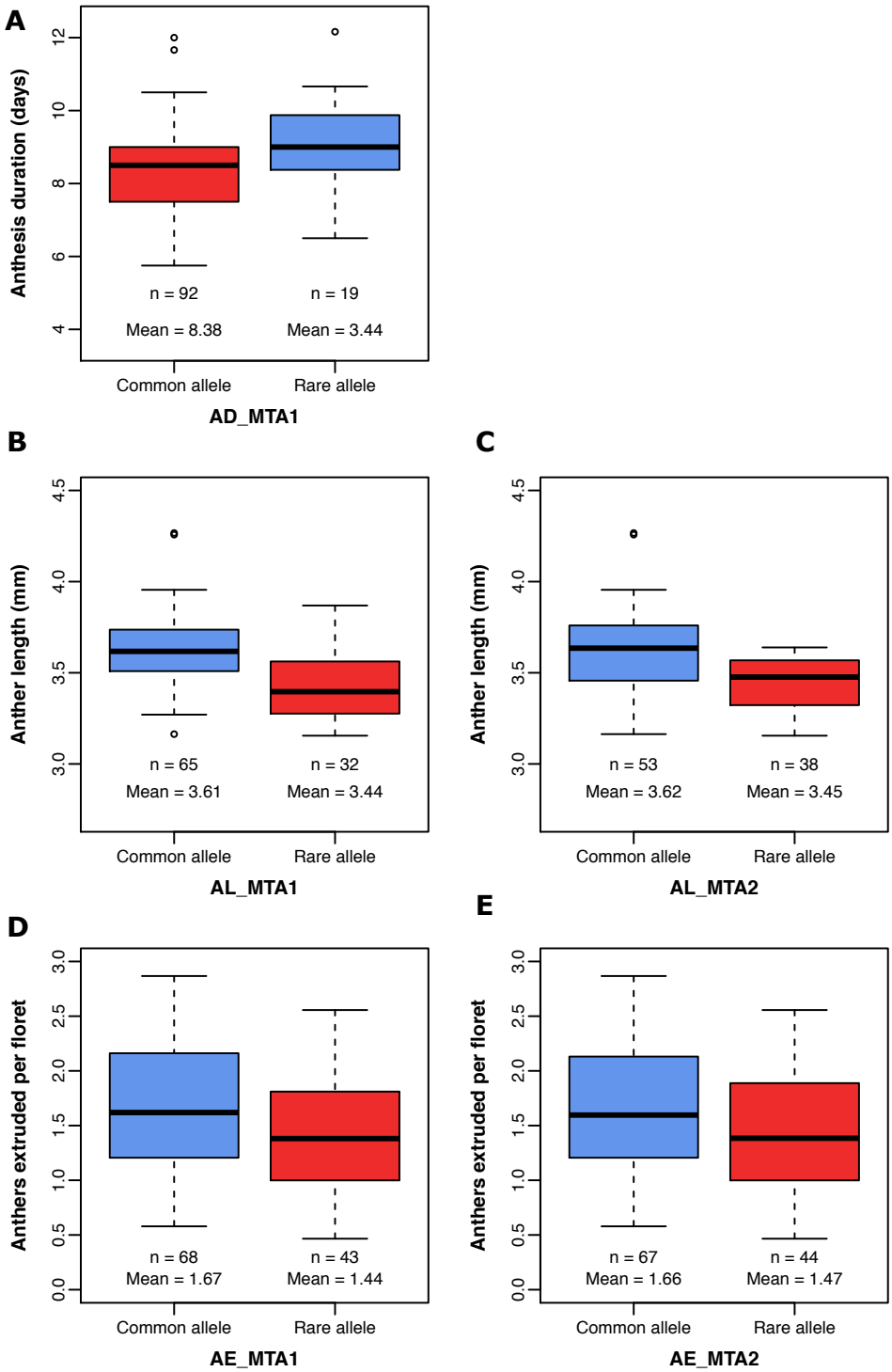

Figure 4.17: Boxplots showing phenotypic variation between genotypes carrying different alleles at declared MTAs for **A)** anthesis duration; **B)** & **C)** anther length **D)** & **E)** anthers extruded per floret. The number (n) of genotypes within each group and the mean values for each group are given below the box plot. Each SNP was bi-allelic in nature; the rare allele was therefore defined as the one being less frequent in the genotype panel.

4.3.2.7 Summary of MTAs

Table 4.5: Marker trait associations (MTAs) detected.

MTA	Chr	-log$_{10}$ *p* BLUEs	-log$_{10}$ *p* Trial1	-log$_{10}$ *p* Trial2	MAF	Effect BLUEs	*Pg*
AD_MTA1	3B	4.79	3.62	4.05	0.17	0.76	16
AD_*Ppd-D1*	FM	4.40	6.00	2.10	0.06	1.11	16
AE_MTA1	2B	5.93	5.96	4.40	0.40	-0.26	12
AE_MTA2	2A	5.77	5.60	4.40	0.40	-0.26	12
AL_MTA1	1B	3.92	NA	NA	0.42	-0.10	16
AL_MTA2	4A	3.59	NA	NA	0.34	-0.10	16
AWNS_MTA1	5A	16.87	NA	NA	0.08	NA	48
HD_*Ppd-D1*	FM	6.78	7.27	5.12	0.06	-3.11	19
HT_*RHT2*	FM	4.41	4.11	4.66	0.25	4.71	11
SP_MTA1	7A	5.95	5.75	6.03	0.36	0.62	20

Abbreviations; AD anthesis duration, AE number of anther extruded per floret, AL anther length, AUFPC area under the flowering progress curve, HD Days to heading from May 1st, HT plant height, Chr Chromosome, Pos chromosome position (cM), MAF minor allele frequency, Effect BLUEs allele substitution effect, pG proportion of genotypic variance explained by the MTA in percent.

Table 4.6: SNPs designated as marker trait associations (MTAs).

MTA	SNP	Chr	Pos
AD_MTA1	BS00060073_51 BS00066467_51 BS00078127_51	3B	71.34
AE_MTA1	Excalibur_c95944_641 BobWhite_c13525_262 RAC875_c29913_139	2B	119.07
AE_MTA2	wsnp_Ex_c45468_51254832	2A	135.56
AL_MTA1	wsnp_Ra_c8506_14401408	1B	62.31
AL_MTA2	RAC875_c1022_3059	4A	57.89
AWNS_MTA1	BobWhite_c8266_227	5A	140.59
SP_MTA1	BS00026622_51 RAC875_c19111_628 wsnp_Ku_rep_c104159_90704469	7A	152.78

Abbreviations, see Table 4.5.

4.3.2.8 Comparison of detected anther extrusion MTAs to previously reported MTAs

Using the consensus genetic map, SNPs within the current dataset that co-localized with anther extrusion MTAs previously reported within the literature (Boeven et al., 2016, Muqaddasi et al., 2017) were identified. These SNPs are hereafter referred to as Boeven_AE (Boeven et al., 2016) and Muqaddasi_AE (Muqaddasi et al., 2017) SNPs. In addition to generating Manhattan plots using the genetic consensus map, plots were generated with the IWGSC RefSeq v1.0 physical map positions (Section 4.2.4.3). In both instances the Boeven_AE and Muqaddasi_AE SNPs were highlighted.

Boeven_AE and Muqaddasi_AE SNPs localized to a number of chromosomes across the genome; clear association peaks were not apparent on all of these in the present study (Figure 4.19). With the physical map, the chromosomal location of SNPs was reassigned from that indicated by the genetic consensus map; Boeven_AE and Muqaddasi_AE SNPs appeared proximal to clear association peaks on chromosomes 1A, 1B, 2B, 2D and 6D and 7A (Figure 4.19 B).

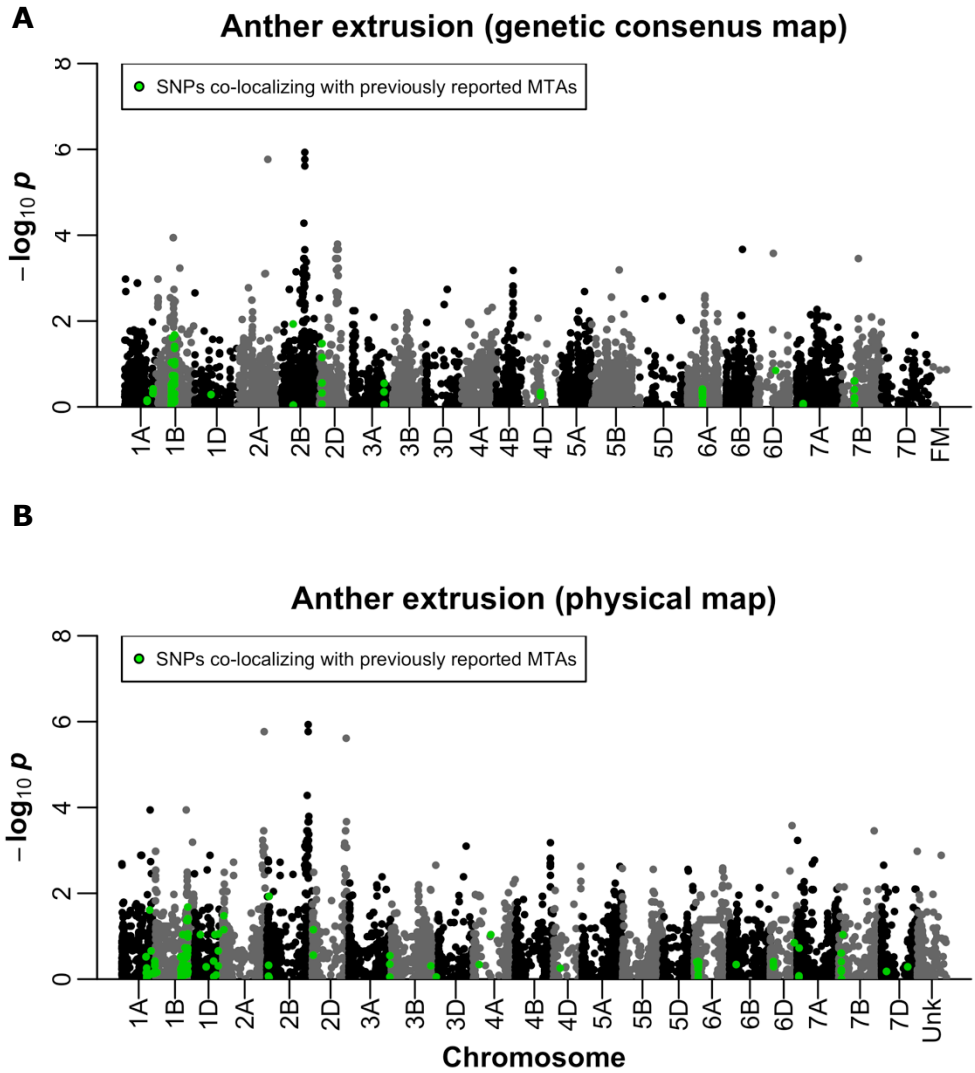

Figure 4.18: Manhattan plots of SNP markers associated with the number of anthers extruded per floret (AE), using **A)** the genetic consensus map and **B)** the physical map positions.

SNPs co-localizing to previously reported MTAs have been highlighted in green. Abbreviations; Fm, functional markers; Unk, unmapped SNPs.

On chromosome 1A the strongest association was detected at 5.00^{e+08}bp (Figure 4.18 A); a number of Boeven_AE SNPs localized proximal to this (4.83^{e+08} to 5.17^{e+08}bp) (Figure 4.19 A). Similarly, on 1B Boeven_AE SNPs were in close proximity to the strongest association (5.43^{e+08}bp) (Figure 4.19 B); ranging from (5.05^{+08} to 5.73^{e+08}bp).

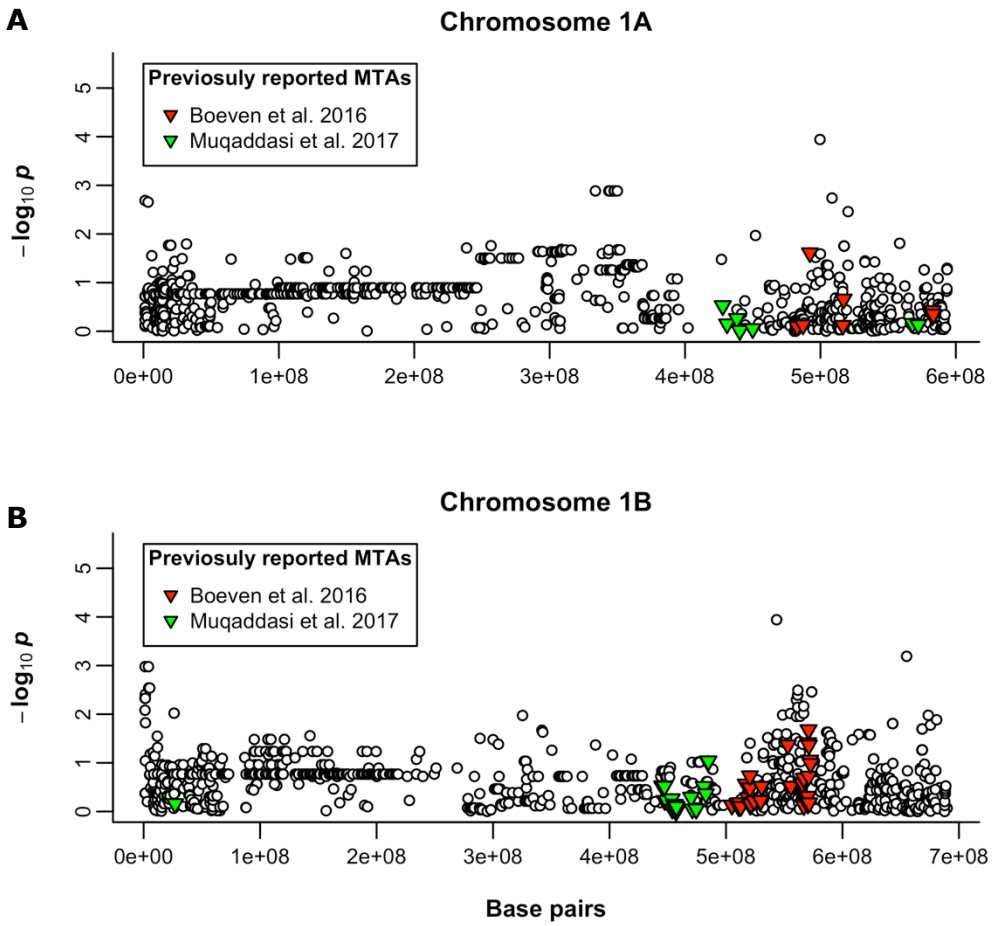

Figure 4.19: Manhattan plots of SNP markers associated with the number of anthers extruded per floret (AE) for individual chromosomes; **A)** 1A and **B)** 1B.

SNPs co-localizing to previously reported MTAs have been highlighted with green or red arrows.

On both chromosomes 2B and 2D Boeven_AE and Muqaddasi_AE SNPs were located on the short arm of the chromosome; the major association peaks on these chromosomes detected in the present study are found on the long arm (Figure 4.20 A). Near to the Boeven_AE and Muqaddasi_AE SNPs a minor peak was observed on 2B and 2D (3.14^{e+08}bp to 4.22^{+08}bp) and (2.08^{+08}bp to 2.22^{+08}bp) respectively (Figure 4.20 B).

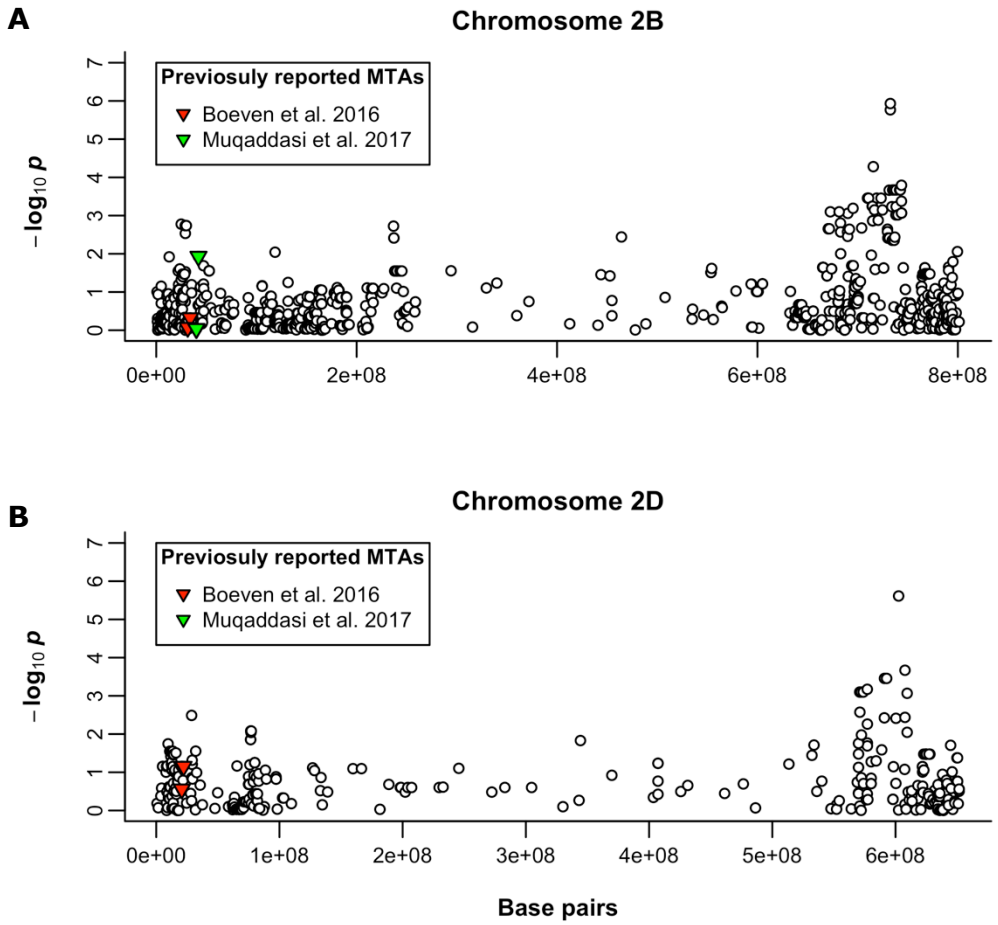

Figure 4.20: Manhattan plots of SNP markers associated with the number of anthers extruded per floret (AE) for individual chromosomes; **A)** 2B and **B)** 2D.
SNPs co-localizing to previously reported MTAs have been highlighted with green or red arrows.

On chromosome 6D, a Boeven_AE SNP (4.22 $^{+08}$bp) was proximal to the strongest association detected on this chromosome (3.84^{+08}bp) (Figure 4.21 A). On chromosome 7A, Boeven_AE SNPs were located between 3.20 $^{+08}$bp to 3.65^{+08}bp relatively close to the strongest association detected on this chromosome (3.08^{+08}bp) (Figure 4.21 B).

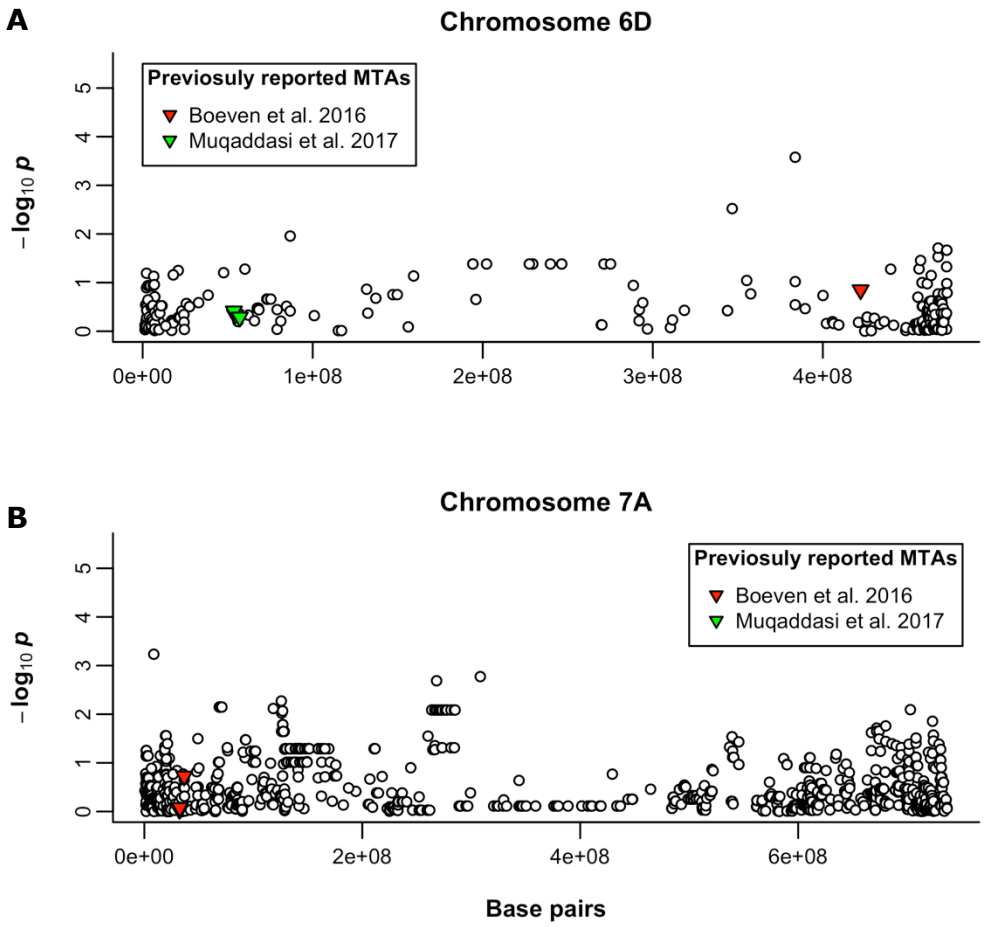

Figure 4.21: Manhattan plots of SNP markers associated with the number of anthers extruded per floret (AE) for individual chromosomes; **A)** 6D and **B)** 7A.
SNPs co-localizing to previously reported MTAs have been highlighted with green or red arrows.

4.3.4 Gene content of genomic regions associated with floral trait MTAs

For the floral traits investigated, genes within the TGACv1 assembly that overlapped SNPs in high linkage disequilibrium (LD) with MTAs were identified. Of the genes identified, a subset was made of those exhibiting expression within spike tissue (Table 4.7).

Table 4.7: The number of SNPs and overlapping genes falling into different categories during the process of narrowing the list of putative candidate genes for each MTA.

Trait	Chr MTA observed	No. SNPs r^2 >0.3	No. genes overlapping SNPs	No. genes with expression in spike
AD	3B	24	15	12
AE	2A	9	7	7
	2B	45	29	24
AL	1B	35	24	24
	4A	4	2	2

Abbreviation, Chr Chromosome; MTA marker trait association; SNP single nucleotide polymorphism; r^2 measure of linkage disequilibrium; AD anthesis duration; AE anthers extruded per floret; AL anther length.

Within this subset, genes showing significant differences in expression between above ground tissue types were investigated further (Table 4.8). Those exhibiting the highest expression within the spike as compared to at least one of the other two tissue types i.e. leaves or shoots were identified (Table 4.8) and described as floral trait influencing candidate genes.

Table 4.8: Transcripts of candidate genes showing higher expression in the spike as compared to leaves and/or shoots.

MTA	Transcript stable ID	Spike expression greater than leaves	Spike expression greater than shoot
AD	TRIAE_CS42_3B_TGACv1_221539_AA0743410.1	**X**	
AE	TRIAE_CS42_2AL_TGACv1_095354_AA0309820.1	**X**	**X**
AE	TRIAE_CS42_2AL_TGACv1_098045_AA0325550.6	**X**	**X**
AE	TRIAE_CS42_2AL_TGACv1_095354_AA0309800.1	**X**	
AE	TRIAE_CS42_2AL_TGACv1_094127_AA0292960.1	**X**	
AE	TRIAE_CS42_2BL_TGACv1_130072_AA0402800.2	**X**	**X**
AE	TRIAE_CS42_2BL_TGACv1_130008_AA0401480.1	**X**	**X**
AE	TRIAE_CS42_2BL_TGACv1_131885_AA0433240.1	**X**	
AE	TRIAE_CS42_2BL_TGACv1_130657_AA0415450.1	**X**	
AE	TRIAE_CS42_2BL_TGACv1_129463_AA0384870.1	**X**	
AE	TRIAE_CS42_2BL_TGACv1_129360_AA0380210.2	**X**	**X**
AL	TRIAE_CS42_1BS_TGACv1_050455_AA0172240.1	**X**	**X**
AL	TRIAE_CS42_1BS_TGACv1_049488_AA0154650.1	**X**	
AL	TRIAE_CS42_1BS_TGACv1_049430_AA0153280.1	**X**	
AL	TRIAE_CS42_1BS_TGACv1_049370_AA0150930.4	**X**	
AL	TRIAE_CS42_4AL_TGACv1_290824_AA0989730.3	**X**	
AL	TRIAE_CS42_4AL_TGACv1_290275_AA0983490.1	**X**	

Significant differences have been determined at the $P < 0.05$ level. Abbreviations, sp spike; lf leaves; sh shoots; exp expression.

Four anther length candidates were identified on chromosome 1B and two on 4A. For AE, four candidates were identified on chromosome 2A and five on 2B. A single AD candidate was identified on chromosome 3B (Table 4.9)

Table 4.9: Putative candidate genes for anthesis duration (AD), anther extrusion (AE) and anther length (AL). The SNPs overlapping candidates are listed along with the orthologous *Arabidopsis thaliana* (AT) genes.

Trait	SNPs	Gene ID	Gene description	AT Gene
AL	IAAV268 IACX5712	1BS_049370 _AA0150930	Kinesin-like protein	KIN13A
AL	GENE_0433_66	1BS_049430 _AA0153280	Mannosyl-oligosaccharide 1,2-alpha-mannosidase	MNS3
AL	Ra_c28610_501	1BS_049488 _AA0154650	NA	NA
AL	Kukri_c7647_1122 IACX5756	1BS_050455 _AA0172240	NA	NA
AE	IAAV1401	2AL_094127 _AA0292960	2-methoxy-6-polyprenyl-1,4-benzoquinol methylase	COQ5
AE	Tdurum_contig28795_381 Excalibur_c37753_754	2AL_095354 _AA0309800	Autophagy-related protein 101	ATG101
AE	BobWhite_c9690_94	2AL_095354 _AA0309820	NA	NA
AE	Kukri_c51540_490	2AL_098045 _AA0325550	NA	NA
AE	Excalibur_rep_c104465_155	2BL_129360 _AA0380210	NA	NA
AE	Tdurum_contig77036_338	2BL_129463 _AA0384870	TaAP2-B	AP2, TOE3
AE	BobWhite_c39437_110	2BL_130008 _AA0401480	NA	NA
AE	RAC875_c28108_523	2BL_130072 _AA0402800	NA	NA
AE	Kukri_c24378_618	2BL_130657 _AA0415450	Golgin candidate 2	GC2
AE	wsnp_Ex_c24135_33382521 wsnp_Ex_c24135_33382700	2BL_131885 _AA0433240	NA	IQD32
AD	BS00110445_51	3B_221539_ AA0743410	NA	SPHK 1-2
AL	RAC875_c1022_3059	4AL_290275 _AA0983490	NA	GN, GNL1
AL	wsnp_Ku_c3081_5776947 Ra_c3111_1623	4AL_290824 _AA0989730	NA	

4.4 Discussion

4.4.1 Population structure of the panel

The population structure within the AM panel was examined through PCA using a reduced marker dataset of 3566 non-redundant SNPs. A clear division of genotypes into subgroups was not observed when the first three principal coordinates (PCs), which collectively explained 19.74% of genotypic variation, were plotted against each other (Figure 4.2). These observations are in line with previous studies using European winter wheat panels, where little population structure was reported (Bentley et al., 2014, Reif et al., 2011, Würschum et al., 2013). Despite the absence of distinct subgroups, some genotypes were observed as being detached from the main cluster in each instance.

Prior studies have emphasized the need for using non-redundant markers when assessing population structure (Lopes et al., 2015, Bansept, 2013). Using an unfiltered dataset, Bansept (2013) observed artificial structuration within her panel due to the presence/absence of the 1BL.1RS translocation. When a spaced marker dataset was used, the division of the panel into two subgroups disappeared (Bansept, 2013). Observations such as this arise because translocations, such as 1BL.1RS, behave as a single

recombination unit. Retaining all the markers that are present in a translocation can therefore result in an overestimation of the degree of population structure.

No clear structure due to the 1BL.1RS translocation was observed within the association panel (Figure 4.2). While clustering together to some extent, genotypes without the translocation were interspersed between those possessing it. This was taken as indicative of the reduced marker dataset being appropriate for inferring population structure through PCA analysis.

In the plot of PC1 versus PC2, the alternative wheat Axona and its descendants (Cadenza and Xi19) clustered together away from the main group (Figure 4.4 A). Heines Peko (PC1= -23.57 and PC2= 21.38), an older alternative wheat, was found within close proximity to the aforementioned genotypes (Figure 4.4). Comparing the pedigree of Axona and Heines Peko revealed that they share a common ancestor, Heines Kolben which is listed as a spring wheat. Within a panel consisting of spring and winter wheats, a clear division between both groups would be anticipated as they are in effect differentiated breeding pools. This may provide an explanation for the detachment of these genotypes from the main group.

In plots of PC1 versus PC3 (Figure 4.4 A and Figure 4.4 B) and PC2 versus PC3, Robigus and its descendants clustered together away from the main group. Robigus is known to have emmer wheat (*Triticum dicoccoides*) in its pedigree but until recently the location and number of any introgressions was unknown. The results reported by Gardner et al. (2016) suggest the presence of introgressed fragments on chromosomes 3B and 4A. The clustering of Robigus derived lines would therefore result from differentiation in multiple parts of the genome, as opposed to the presence of a single introgressed fragment. Further to this, Robigus originates from a cross originally made by a Dutch breeding company and would be anticipated to be genetically distinct from older UK varieties (Jacob Lage, KWS UK Ltd., personal communication, 2017). This could explain why clustering was observed despite using the reduced marker dataset.

4.4.2 Extent of linkage disequilibrium (LD)

Significant LD was found to extend to approximately 16 cM (Table 4.2 and Figure 4.5) based on a critical r^2 value of 0.06 (Breseghello and Sorrells, 2006). The rate of LD decay observed was comparable to previous reports for European winter wheats by Bentley et. al (2014) (DArT, 20 cM) and Bansept (2013) (DArT,16.7 cM) but was slower than that reported by Würshum et al. (2013) (SNP, 8 cM). From the

rate of LD decay within the association panel, the present study would have a reasonable chance of detecting causal loci albeit potentially with low resolution.

Variation in the rate of LD decay for elite European germplasm within the literature can be attributed to differences in the genetic diversity of the panels under investigation. Chao et al. (2010) reported than LD decay was much slower within subpopulations as compared to within the entire panel. The rate observed for the association panel can be attributed to its relatively small size and possibly lower genetic diversity as compared to that used by Wurschum et al. (2013). Relatively slower decay of LD was observed for the D genome (Table 4.2 and Figure 4.6 C); this has been reported in a number of studies (Chao et al., 2010, Akhunov et al., 2010, Chao et al., 2007, Mackay et al., 2014) and has been proposed to be an artefact of the genetic bottleneck that accompanied the speciation of hexaploid wheat.

4.4.3 Effectively controlling for false positives in association mapping

Association mapping studies (AM) essentially model what could potentially be a multi-factorial trait as if it were due to single locus, ignoring background variables (Atwell et al., 2010). A concerted effort must therefore be made to account

for these variables in order to minimize the chance of detecting spurious associations. Using the naïve model in AM for the traits investigated, an excess of significant/strong associations was observed; the extent to which varying between traits (Figure 4.7, 4.8 and 4.9). This was anticipated as prior studies have reported that the extent of confounding by false positives varies between phenotypes due to differences in genetic architecture and selection history (Atwell et al., 2010).

In addition to a naïve model, two models were employed in the association mapping. Family relatedness was anticipated to be a confounding factor within the association panel, as many genotypes share the same parental lines, and therefore a kinship matrix was first included in a mixed linear model (MLM). Comparisons of the QQ-Plots from this model and the naïve model revealed that the kinship matrix had a large influence on the distribution of p-values; demonstrating its value in controlling the rate of false positives (Table 4.4 and Figure 4.7, 4.8 and 4.9).

The kinship matrix on its own did not appear to adequately control for false positives (Table 4.4). The first three principal coordinates (PCs) were then incorporated into the MLM as well. This model, the full model, was found to be the most appropriate for AM in the association panel (Table 4.4).

Application of the full model resulted in a near uniform distribution of p-values within the QQ-plots, with deviations from the diagonal occurring much later than observed using the naïve model.

It would be remiss to assert that population structure and kinship can be entirely accounted for in AM; prior studies have demonstrated that even with the most stringent control measures, confounding can still occur (Atwell et al., 2010). Care must be taken in interpreting the results of any AM study; however, with the genome wide-coverage of SNPs used in the present study the assumption can be made that the true causative loci, or at least the genomic region it localizes to, would show the strongest association (Atwell et al., 2010).

4.4.4 Suitability of genotype panel for association mapping

Further to investigating the decay of LD in order to support the suitability of the present study's panel for AM, an effort was made to ascertain whether MTAs previously reported for larger European winter wheat panels could be detected. For both spikelet number (SP) and awn presence/absence (AWNS) this was achieved with a high amount of precision (Figure 4.11) (Mackay et al., 2014, Boeven et al., 2016). This

supports the assertion that despite the relatively small size of the panel, enough diversity was captured so as to achieve a rate of LD decay needed for the detection of significant MTAs.

While LD decay appears to be adequate, marker coverage across the three genomes was not equal. Within the association panel, the D genome was characterized by a considerably lower number of polymorphic markers; an extreme case being for group 4 with 610, 577 and 79 SNPs being localized to chromosomes 4A, 4B and 4D respectively. This needs to be taken into consideration when interpreting the present AM results.

Despite being able to localize *Rht1* to the expected position on 4B (Sukumaran et al., 2016) the present study was unable to localize the *Rht2* loci on chromosome 4D via marker-marker associations. No SNPs were in fact present in close proximity to the position at which the *Rht2* locus was anticipated (Figure 4.10 C) (Lopes et al., 2015). Within the dataset, SNPs at the position expected for *Ppd-D1* were present (Figure 4.10) (Sukumaran et al., 2016); the strongest association on chromosome 2D occurred at this position however it was not the strongest within the genome. Prior studies have reported that there is a fast decay of LD in the region surrounding the *Ppd-D1* locus in European material (Langer et al., 2014a) which may help explain this observation. The rate of LD decay

around the locus may vary between populations and could explain why Sukumuran et al. (2016) were able to detect a significant marker-marker association with the *Ppd-D1* locus at the expected position on chromosome 2D.

4.4.5 Marker trait associations

MTAs were declared significant based on a Bonferroni correction threshold at α=0.01 or 0.05. The threshold, being a function of the number of markers tested, should be viewed as extremely conservative as it assumes the independence of markers. With the exception of the number of anthers extruded per floret (AE), no significant MTAs were detected for floral traits. Failure to detect significant associations does not necessarily imply that the present AM study has not detected genomic regions that have a tangible effect on the traits of interest.

The significance of the association of a given marker with phenotypic variation, is in effect a function of a number of factors. These include the heritability of the trait under investigation, the number of genes influencing the phenotype, the effect of an individual marker and the frequency at which the rare allele occurs within the panel being used (Myles et al., 2009, Brachi et al., 2011).

Use of the false discovery rate (FDR) is an alternative approach which uses the distribution of observed p-values to estimate the proportion of significant associations that will be false positives. The FDR approach (Storey, 2002) is also inherently limited as having a high number of correlated markers i.e. non-independent, skews the distribution of p-values and in turn the estimation of the FDR. Bansept et al. (2013) reported no improvement using the FDR method as compared to the Bonferroni correction method; the authors attributed this to having a large proportion of redundant markers within their dataset. As a high proportion of correlated SNPs was present within the current genotype dataset, the Bonferroni correction threshold was retained.

4.4.5.1 Effects of marker trait associations

Marker trait associations detected for floral traits explained a modest proportion of the phenotypic variation observed within the panel, ranging from 12 to 16 percent. Compare this to that observed for spikelets per spike (SP) and days to heading (HD); the strongest MTAs detected for these traits explained 20 and 19 percent respectively.

With reference to AE, the proportion of genotypic variance explained by the two MTAs described in the present study explained twelve percent each. This was comparable to that

reported for the MTAs detected by Boeven et al.(2016) (0.33 to 10.15 %) and Muqaddasi et al. (2017) (9.44 to 16.98%). The rare allele of AE_MTA1 and AE_MTA2 exhibited a negative effect on the number of anthers extruded per floret of the magnitude -0.26; which falls within the range reported by Boeven et al. (2016) (-0.29 to 0.35). As opposed to reporting the number of anthers extruded per floret (maximum of three) Muqaddasi et al. (2017) reported the number of anthers extruded by all eight florets scored (a maximum of 24). Expressing their results on the scale used in the present study, the effects of MTAs reported ranged from -0.2 to 0.25. In the case of anthesis duration (AD) and anther length (AL), MTAs described for both traits explained 16 percent of the phenotypic variance each. The effect of the rare allele of AD_MTA1 was to increase anthesis duration by 0.76d, this was comparable to that seen for the *Ppd-D1* locus (1d) in the initial analysis. In the case of AL_MTA1 and AL_MTA2, the rare allele of both had a negative effect on anther length of the magnitude -1 mm.

4.4.5.2 Genomic regions associated with variation in floral characteristics

It may be surprising that significant/strong associations were detected in the present study due to the relatively small size

of the panel used and the relatively low statistical power that would be expected as a consequence. As mentioned previously, the genetic architecture of a trait also influences detection power. This helps to explain the limited success of AM for some traits and the presence of clear associations for others in the present study.

With reference to genetic architecture, prior information was available regarding the heritability of the traits used in in AM. The heritability observed for anthesis duration (AD) and the area under the flowering progress (AUFPC) was relatively lower than that observed for the number of anthers extruded per floret (AE). The heritability observed for AE (0.8) was in fact comparable to that observed for traits for which significant marker trait associations were also observed such as the number of spikelets per spike (0.8) and the number of days to heading (0.7). As the power to detect marker trait associations increases with trait heritability, the chances of detecting associated genomic regions was not expected to be equal for all floral traits.

For the trait AD, a strong correlation was observed with the number of days to heading (Section 3.3.1.4). Unsurprisingly *Ppd-D1* showed the strongest association in AM for this trait. Upon inclusion of *Ppd-D1* allelic status as a covariate in AM, a clear association peak, albeit statistically insignificant, was

observed on chromosome 3B. In light of what appeared to be effective control for false positives, it would be remiss to disregard association peaks such as these; which are indicative of genetic factors capable of exhibiting phenotypic effects across diverse genetic backgrounds.

Prior studies have detected a flowering time QTL near the centromere of the long arm of chromosome 3B (Griffiths et al., 2009, Pánková et al., 2008, Zanke et al., 2014); the use of DArT markers in these studies prevents direct comparisons to be made between the localization of this and AD_MTA1. Zanke et al. (2014) simultaneously applied SNP genotyping, alongside DArT, in their association mapping study of flowering time in European germplasm and detected a number of co-localizing marker-trait associations at 65 cM on chromosome 3B. As AD_MTA1 localizes to 71.34 cM, the possibility exists that the same genomic region has been detected in the present study. Subsequent to identification of the causative locus of AD_MTA1, genomic resources such as TILLING populations (Uauy et al., 2009), can be used in order to elucidate whether the phenotypic effect on anthesis duration associated with AD_MTA1 is an indirect result of variation in flowering time.

Association peaks were also observed for anther length (AL); although more care should be taken in the interpretation of

these results as only a single year's data was available. On both chromosomes 1B and 4A, the association strength of SNPs increased with proximity to the strongest marker-trait association forming clear association peaks.

Prior to this study, factors capable of influencing the anthesis profiles of genotypes have not been discussed. In addition to anthesis duration (Section 3.3.1.4), it is likely that AUFPC is a function of a number of additional traits and consequently governed by a number of genes. This in addition to the relatively low level of heritability observed, may explain the absence of clear association peaks as in the case of AD or AL; a number of relatively weak peaks were observed throughout the genome.

A practical approach towards AM for complex traits such as AUFPC would be to dissect the individual traits that constitute the phenotype and carry out individual AM studies for each. In Chapter 3 while a number of additional floral and agronomic traits were investigated, the only trait that showed a strong significant association with AUFPC was anthesis duration.

With reference to AUFPC, further work is required in order to ascertain what physiological factors influence this trait; possibilities include anthesis duration within individual ears and the timing of anthesis initiation in successive tillers. Efforts towards AM for a trait such as AUFPC can also be aided

by having a relatively larger panel in which LD decays at a faster rate; the likelihood of detecting phenotypic associations with multiple genomic regions decreases with slower decay of LD.

For AE clear marker trait associations were observed at homoeologus positons on the long arm of the chromosomes 2A (2AL) and 2B (2BL). As a number of recent association mapping studies targeting anther extrusion have been carried out using the same genotyping platform as the present study, a detailed approach was taken in order to ascertain whether the major association peak and other minor peaks had previously been reported.

4.4.5.2.1 Anther extrusion: comparison of marker trait associations detected to those reported previously

From the comparisons carried out it was concluded that the genomic regions containing the major association peaks observed on 2A and 2B have not previously been reported as being associated with anther extrusion. While both of the compared studies (Muqaddasi et al., 2017, Boeven et al., 2016) utilized European germplasm, varieties bred for/within the UK were more represented in the panel used in the present study. This may help to explain why the major MTAs of the present study were not detected by Boeven et al. (2016) and Muqaddasi et al. (2017)

On chromosomes 1A, 1B, 7A and the short arms of 2A and 2B SNPs showed a greater association with AE within regions previously reported to have AE marker-trait associations (Boeven et al., 2016, Muqaddasi et al., 2017). The relatively weak strength of these associations observed in the present study may be attributable to the effect of causative loci being dependent on genetic background.

4.4.5.4 Prospects for marker assisted selection

The findings of the present study may be of benefit towards the development of marker assisted selection (MAS) programmes if causative loci can be identified. This is not necessarily straightforward, in some cases SNPs exhibiting the strongest association with phenotype are non-causal. They may instead be positively correlated with the true causative loci and with the genomic background (due to population structure which may have not been entirely accounted for) (Atwell et al., 2010). In the present study in most instances single strong association peaks that are obvious by eye were observed, suggesting that true associations have in fact been detected. Atwell et al. (2010) emphasized that background functional knowledge about pathways related to a trait can help considerably in the prediction of which loci underlie observable variation.

For each of the MTAs declared for floral traits the allele associated with the desired phenotype, i.e. longer anthesis duration, larger anthers and more extruded anthers, was not extremely rare within the association panel. The allele frequency of the desirable allele was 17% for AD_MTA1; 60% for AE_MTA1 and AE_MTA2; 58% for AL_MTA1 and and 66% for AL_MTA2.

This suggests that targeting the causative loci associated with these MTAs towards the improvement of these traits is feasible. Locally common alleles have been described as more useful in breeding programmes as they will be either be at a selective advantage or selectively neutral (Brown, 1989). Further to this, with common alleles it will be easier to devise effective strategies to increase their frequency in male parental pools (Brown, 1989).

With respect to anther length and anthers extruded per floret, it can be argued that as the desirable allele was present in the majority of the association panel, these findings will be of little practical use towards making large improvements in either of these traits.

In the case of anther extrusion, all prior studies indicate that the trait is quantitatively inherited with a number of loci underlying variation in the trait (Muqaddasi et al., 2017, Boeven et al., 2016, Lu et al., 2013, Skinnes et al., 2010).

The results of the present study suggest that anther length is also polygenic trait. Association mapping studies have a greater likelihood of detecting genomic regions with robust loci that are capable of exhibiting effects across diverse genetic backgrounds. These loci may in fact be prerequisite for enhanced expression of the desired phenotype.

For all floral traits, the desirable allele of declared MTAs resulted in a notable shift in the distribution of phenotypic values towards the upper end of the distribution (Figure 4.17). In the case AE and AL, the highest phenotypic values occurred only in the presence of the desirable allele of the MTAs declared for these trait (Figure 4.17 B, C, D & E).

This taken together with the proportion of phenotypic variance explained by each MTA (12 to 16%) and their allelic effects suggests that development of reliable markers from these MTAs and their use as a selection criterion, would be an attractive platform for screening lines prior to field evaluation in early generations.

For each AE, AD and AL no more than two (2) medium effect MTAs could be declared. Marker-assisted selection targeting just these loci, or even in conjunction with the other medium effect MTAs previously reported, is unlikely to be feasible due to the costs incurred and the high selection response anticipated for phenotypic selection; high heritability has been

observed for these traits in the present study (Chapter 3) and reported by past studies.

Markers developed from these MTAs will be most applicable during reciprocal recurrent selection in male parental pools (Boeven et al., 2016), especially in light of the need to continually introgress new germplasm into the established pool. Progenies will need to be screened for floral characteristics as well as their general combining ability (GCA) (Boeven et al., 2016) and while this can be achieved through testcrosses with lines in the female pool, requires a significant cost and time investment.

Genomic prediction using genome-wide molecular markers can provide a more feasible approach towards screening in this instance, especially taking into consideration that marker profiles are likely to be available at this stage in a wheat breeding programme. Weighted ridge-regression best linear unbiased prediction wRR-BLUP where major and medium effect MTAs are fitted as fixed effects, for example those from the present study together with others reported, offers a promising approach for genomic prediction targeting floral characteristics and GCA (Zhao et al., 2014) simultaneously. As compared to unweighted RR-BLUP, wRR-BLUP has the added benefit of overcoming the limitations of target loci not being in strong LD other markers on the respective

chromosome (Boeven et al., 2016). Using this approach, Boeven et al. (2016) reported a high cross-validated prediction ability for a number of floral traits; the highest being 0.70 for anther extrusion.

4.4.5.5 Putative causative loci within MTA genomic regions

As mentioned previously, SNPs showing the strongest association with phenotypic variation may not necessarily overlap the causative gene. For each floral trait MTA, the SNPs showing the strongest association were therefore investigated alongside those SNPs that they are in high LD with.

Analysis of MTA SNPs and markers in high LD regions (r^2 > 3.0) revealed 24, 54 and 29 distinct SNPs for the traits anthesis duration (AD), the number of anthers extruded per floret (AE) and anther length (AL) (Table 4.7).

For each individual MTA a number of genes overlapping SNPs were identified; the associated transcripts (splice variants) of some genes showed no expression in the spike at any developmental stage (Table 4.7). These genes were therefore excluded from the list of potential candidates for the observed traits.

Amongst the genes being expressed within the spike, and with significant differences detected in relation to tissue and developmental stage ($P < 0.05$), a small number were characterized by higher expression in the spike as compared to at least one of the two remaining above-ground tissues (Table 4.8). The genes contained within this subset were considered likely candidates.

The approach taken towards identifying candidate genes in the present study does possess limitations. Expression profiles and the assembly used in SNP and gene localization were both generated from the variety Chinese Spring which is likely to be genetically distinct from the varieties used in the present study. Further to this, higher expression within the spike as compared to other tissues may not be a pre-requisite for tangible phenotypic effects to be observed. Temporal variation in gene expression was accounted for by including tissue age as a factor in ANOVA analysis; however, changes in gene expression over time found to be statistically insignificant may still be sufficient for the initiation/cessation of developmental processes.

The role of putative candidates identified in the present study can be confirmed by various strategies including using genetic resources such as tilling populations; fine mapping of the genomic regions containing MTAs is however recommended.

This can either be achieved through SNP enrichment or the development of populations with recombinations in the respective genomic regions that can be used in linkage mapping studies.

Amongst the anther extrusion candidates, a *T. aestivum* ortholog of the *Arabidopsis thaliana* floral gene *AP2* was found to overlap with a SNP within the high LD region of AE_MTA1 (chromosome 2B) (Table 4.9). Allelic variation of a gene orthologous to AP2 in *Horedum vulagre* has previously been demonstrated to influence floret gaping (Nair et al., 2010). While the *T. aestivum* orthologs have been characterized, functional genetic variation has not previously been reported (Ning et al., 2013b). The 2A homoeolog of TaAP2-B (TaAP2-A) was not found in the high LD region of AE_MTA2.

For anther length, a putative ortholog of the *A. thaliana* genes *GN* (Source: UniProtKB/Swiss-Prot; Acc: Q42510) and *GNL1* (Source: UniProtKB/Swiss-Prot; Acc: Q9FLY5) was identified as a candidate on chromosome 4A (Table 4.9). In *A. thaliana* both genes are expressed in flowers and are strongly expressed in actively dividing or elongating cells.

On chromosome 3B, a putative ortholog of the *A. thaliana* genes SPHK1 (Source: UniProtKB/Swiss-Prot; Acc: Q8L7L1) and SPHK2 (Source: UniProtKB/Swiss-Prot; Acc: F2Y4A3) was present in the high LD region associated with anthesis

duration (Table 4.9). High expression of these genes in stems and flowers is observed in *A. thaliana* and both are involved in abscisic acid (ABA) signalling.

Researchers have emphasized the need for integrating multiples sources of genome-wide evidence towards narrowing the list of candidate genes to a more manageable number for further investigation (Jumbo-Lucioni et al., 2010, Civelek and Lusis, 2014). While outside the scope of the present study, one possibility would be to obtain expression data from RNA-seq, using floret specific tissue from key spike developmental stages, in order to complement the results out of AM (Ingvarsson and Street, 2011).

An approach similar to bulk-segregant analysis can be taken; wherein genotypes at both ends of the phenotypic distribution for a trait are analysed and significant differences in gene expression between the two groups identified (Ingvarsson and Street, 2011). The anchoring of annotated gene models and SNPs to the physical map of the *T. aestivum* genome, carried out by researchers attached to the John Innes Centre and Earlham Institute (www.wheat-training.com), would enable direct comparisons to be made between the results of AM and expression analyses.

4.5 Summary

The suitability of the genotype panel used for association mapping was supported by estimates of linkage disequilibrium (LD), the ability to localize functional markers and marker-trait associations previously reported for agronomic traits. The genomic resolution of the present study was however inherently limited by a relatively low SNP marker density on the D genome chromosomes and not having a faster rate of LD decay within the association mapping panel used.

AM for floral traits identified genomic regions with clear association peaks for anthesis duration, anther length and the number of anthers extruded per floret. Marker trait associations in these regions explained a modest proportion of genotypic variance and were associated with tangible phenotypic effects across genetic backgrounds. Within the association peaks observed, putative causative loci were identified through linkage disequilibrium analysis together with expression profile data. The findings of the present study can facilitate the confirmation of causative floral trait loci and their subsequent use in breeding programmes.

CHAPTER 5: IDENTIFCATION OF ANTHER EXTRUSION QTLS BY LINKAGE MAPPING IN A BI-PARENTAL POPULATION

5.1 Introduction

Robust loci, exhibiting modest phenotypic effects across diverse genetic backgrounds are undoubtedly key targets for trait improvement. For quantitatively inherited traits, these loci may in fact be pre-requisite for the expression of the desirable phenotype. Following the detection of loci meeting this criterion by association mapping (AM), described in Chapter 4, the focus of the present study shifts towards the identification of additional loci present within recently released commercial varieties that are capable of enhancing anther extrusion capacity.

Towards this effort, the present chapter describes linkage mapping (LM) of anther extrusion in an F3 derived F4 population originating from a cross between KWS Ferrum and KWS Tempo, varieties released in 2010 and 2015 respectively. LM in this bi-parental population provided an opportunity to circumvent two inherent limitations identified in the preceding AM study:

1. Use of a bi-parental population would enable estimation of phenotypic effects of loci with less confounding due to variation in phenological characteristics (Dodig et al., 2012). Loci not found to be associated with variation in

anther extrusion within the AM panel may therefore exhibit an effect within the bi-parental population.

2. The minor allele frequency threshold set for AM, in order to minimize the detection of spurious associations (Brachi et al., 2011), may have prevented the detection of anther extrusion enhancing alleles that were at a very low frequency in the association panel.

While LM was anticipated to provide these benefits, a lower level of genomic precision was expected as a trade-off, as bi-parental populations are typically characterised by a lower frequency of recombination events as compared to association panels. This factor has implications for the interpretation of the results of LM; specifically, with reference to the identification of causative quantitative trait loci (QTL) regions.

In the present chapter, the phenotyping of the aforementioned bi-parental population for the number of anthers extruded per floret and the subsequent detection of anther extrusion QTLs is reported. Avenues for the identification of causative loci within each QTL are discussed.

5.2 Methods

5.2.1 Plant material and phenotyping

A population of 307 unselected F4 lines (F3:4 lines) originating from a cross between varieties KWS Ferrum and KWS Tempo, were provided by KWS UK Ltd. for use in this experiment. KWS Ferrum has previously been characterised as having a high capacity for anther extrusion while that of KWS Tempo is moderate. The field trial was carried out in the summer of 2015 at Fowlmere, UK; with each line being represented by a one-metre-long row of 8-10 plants, all derived from a single F3 plant.

Daily environmental variables, maximum/minimum air temperature (^{o}C) and rainfall (mm), during the field trial were accessed from the Ickleton weather station and provided by staff attached to KWS UK Ltd (Appendix 1).

Phenotyping for anther extrusion, days to heading and plant height was carried out as done in Section 3.2.1.2. In the case of anther extrusion, assessments were carried out using five ears collected from each row; the row average was taken as the phenotypic value for a line. A Pearson correlation coefficient (r) between traits were calculated within the R environment using the R package ‘Hmisc'.

5.2.2 Genotyping and linkage map construction

Using genomic DNA extracted from pooled samples of young leaves from five plants per line, genotyping was carried out by KWS UK Ltd, via single nucleotide polymorphism (SNP) using a proprietary Illumina Infinium iSelect 15,000 SNP wheat array. This gave a total of 2685 polymorphic SNPs; markers with greater than 10% missing data, and greater than 40% heterozygous genotypes or a minor allele frequency (MAF) less than 5% were excluded from analysis.

A linkage map produced from filtered SNPs was provided by KWS UK. Construction of linkage groups was carried out with JoinMap 4.0, with maximum likelihood mapping being used to estimate map distance (cM). Using the high density Multiparent Advanced Generation Intercross (MAGIC) wheat map published by Gardner et al. (2016), linkage groups were then assigned to chromosomes and where applicable allocated to specific chromosomal arms by comparing arrangement of markers with that published.

SNPs fell into 19 linkage groups (Table 5.1); no groups could be assigned to chromosomes 3D, 4D, 5A and 5D (Table 5.1). Chromosomes 2B and 7A are represented by two linkage groups each (Table 5.1) corresponding to different chromosomal arms in both instances.

Table 5.17: Linkage groups constructed and distribution of 2685 polymorphic SNP markers across the three wheat genomes and groups.

Linkage Group	Chr	No. markers
1	1A	208
2	1B	217
3	1D	86
4	2A	252
5	2BS	184
6	2BL	18
7	2D	53
8	3A	116
9	3B	236
10	4A	128
11	4B	136
12	5B	297
13	6A	161
14	6B	258
15	6D	54
16	7AS	50
17	7AL	133
18	7B	91
19	7D	7

Chr	Genome			Total by group
	A	B	D	
1	208	216	86	510
2	253	202	53	508
3	116	236	-	352
4	128	136	-	264
5	-	297	-	297
6	161	258	54	473
7	183	91	7	281
Total by genome	1049	1436	200	
Total				2685

Abbreviations, Chr Chromosome.

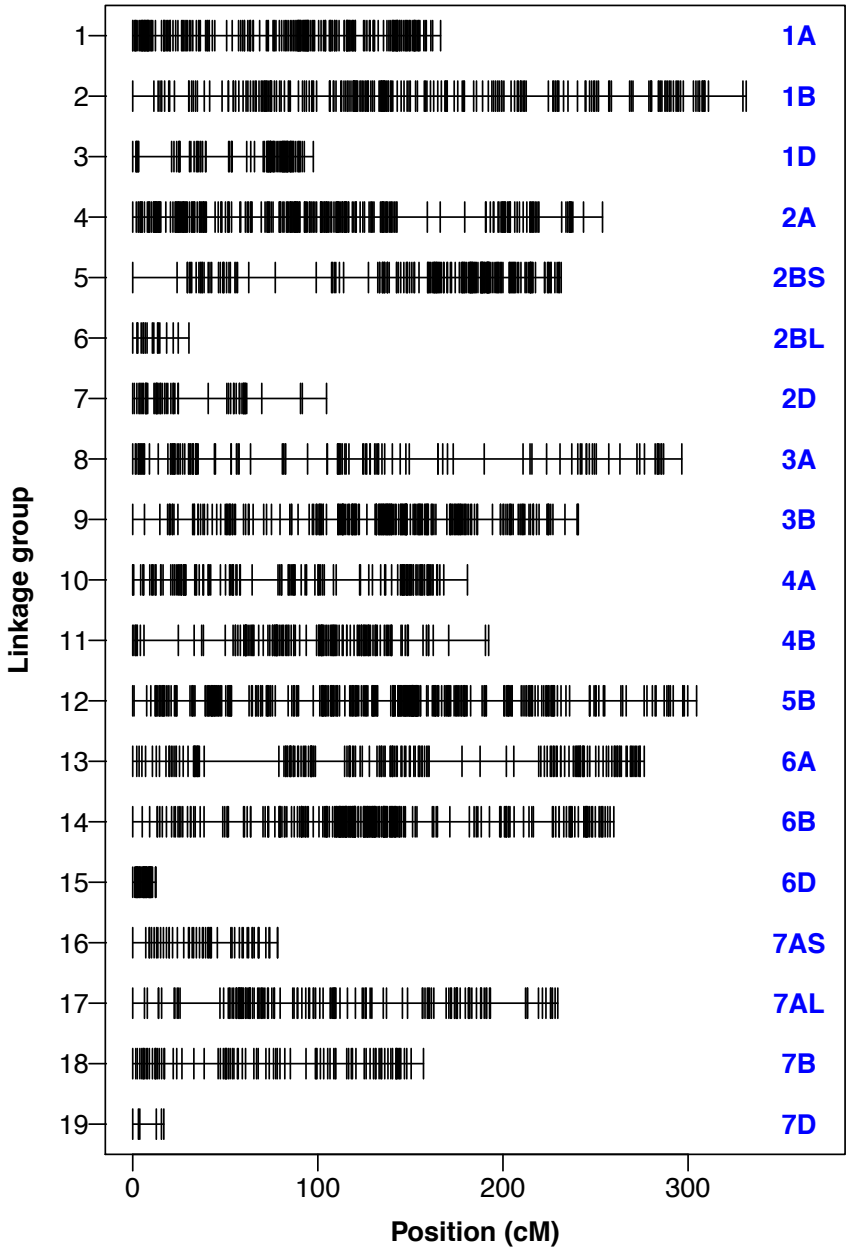

Figure 5.1: Genetic maps of each linkage group constructed. The assigned chromosome identity of each linkage group is indicated towards the right of the plot.

5.2.3 QTL detection

QTL analysis was performed in the R software environment using the R-package 'R/qtl'. Using line averages, single interval mapping (SIM) and composite interval mapping (CIM) were carried out. For CIM, missing marker data was imputed

and forward selection of covariates was done in three ways; with the exclusion window being set at 20 cM, 60 cM and 2000 cM. Stepwise selection for multiple QTLs (MQM), with forward/backward model selection was also carried out in order to confirm QTLs detected by CIM/SIM and refine their position.

For each trait, QTL significance was determined using 1000 permutations ($P<0.05$) from which a LOD-threshold was declared. The LOD support interval (confidence interval) was defined by the LOD-drop-off method (Lander and Botstein, 1989). The boundaries of the interval were defined by an increase in LOD of 1 unit, corresponding to a 96.8 CI in theory (Mangin et al., 1994).

In the initial plan of this study, F5 lines derived from the F4 population exhibiting recombinations in the genomic regions harbouring QTLs, would have been phenotyped in a second field trial and linkage mapping carried out. This would not only have provided insight into the environmental stability of QTLs but also served as the initial step towards fine-mapping of these regions. Due to unforeseen circumstances, a number of lines in this F5 population were not available for phenotyping in the 2017 field season and the second phase of the linkage mapping study was therefore not carried out.

5.2.4 Identification of putative causative loci within anther extrusion QTL regions

For detected QTLs, the physical positions of SNPs at the left and right border of the confidence interval were determined with the dataset used in Chapter 4, Section 4.2.4.2. The TGACv1 gene models overlapping the region defined by the the border SNPs was determined and candidate genes then selected based on their expression profiles, as in Section 4.2.4.2.

5.3 Results

5.3.1 Phenotypic variation

Continuous variation was observed for all traits (Figure 5.2). While a near normal distribution was observed for days to heading and plant height, the distribution for number of anthers extruded per floret was skewed towards higher anther extrusion (Figure 5.2). Transgressive segregation in AE was apparent with a number of lines extruding a lower number of anthers per floret than KWS Tempo (Figure 5.2 C).

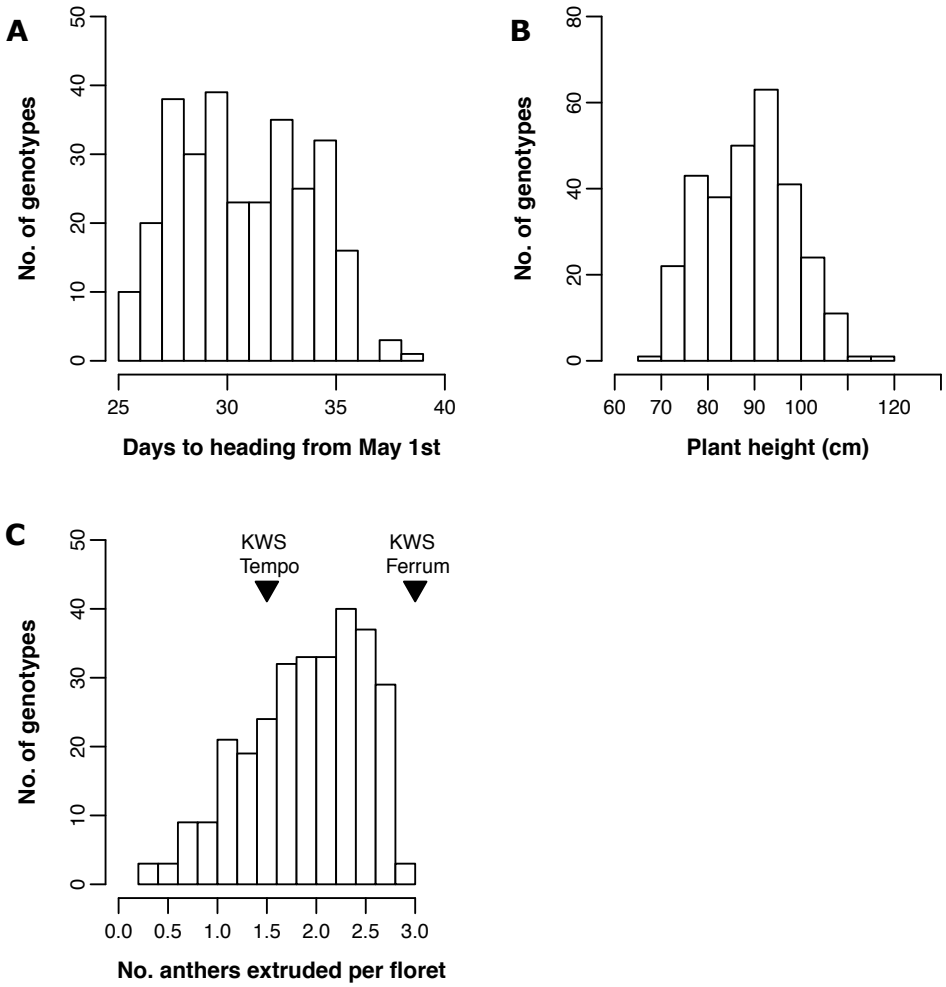

Figure 5.2: Histograms of the distribution of phenotypic values for; **A)** days to heading from May1st; **B)** plant height and C**)** no. anthers extruded per floret.

Significant (P<0.001) correlations were observed between the number of anthers extruded per floret (AE) and days to heading (HD), as well as plant height (HT). The strength of these correlations however was relatively weak r= -0.25 (HD) and 0.33 (HT) (Figure 5.3).

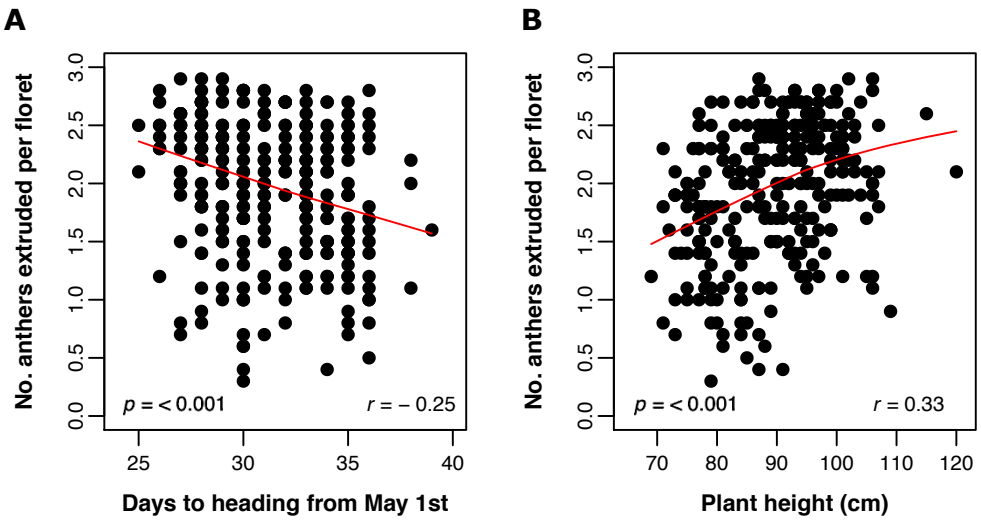

Figure 5.3: Scatterplots of phenotypic values for number of anthers extruded per floret against; **A)** days to heading; **B)** plant height.
A fitted lowess line (red) and the Pearson correlation coefficient (r) are presented for each comparison.

5.3.2 Comparison of QTL detection methods

Differences were detected between composite interval mapping (CIM), multiple QTL mapping (MQM) and single interval mapping (SIM) with respect to the number of QTLs detected as well as their chromosomal position.

CIM and MQM both detected a significant QTL for days to heading (HD) on chromosome 7B that was not detected by SIM (Figure 5.4). Conversely, SIM detected a plant height (HT) QTL on chromosome 4A than was not found to be significant in MQM or CIM analyses (Figure 5.4).

The effect of restriction window size on QTL LOD scores and confidence intervals varied for traits and between chromosomes (Figure 5.4). For all traits, significant QTLs detected by restricted CIM (exclusion window size 2000 cM) were also detected by MQM (Figure 5.4); with no variation in the position or size (cM) of the QTLs being observed. LOD scores were however found to be lower with MQM as compared to CIM (Figure 5.4, Table 5.2). The results of restricted CIM and MQM are presented in subsequent sections together with that of SIM for comparative purposes.

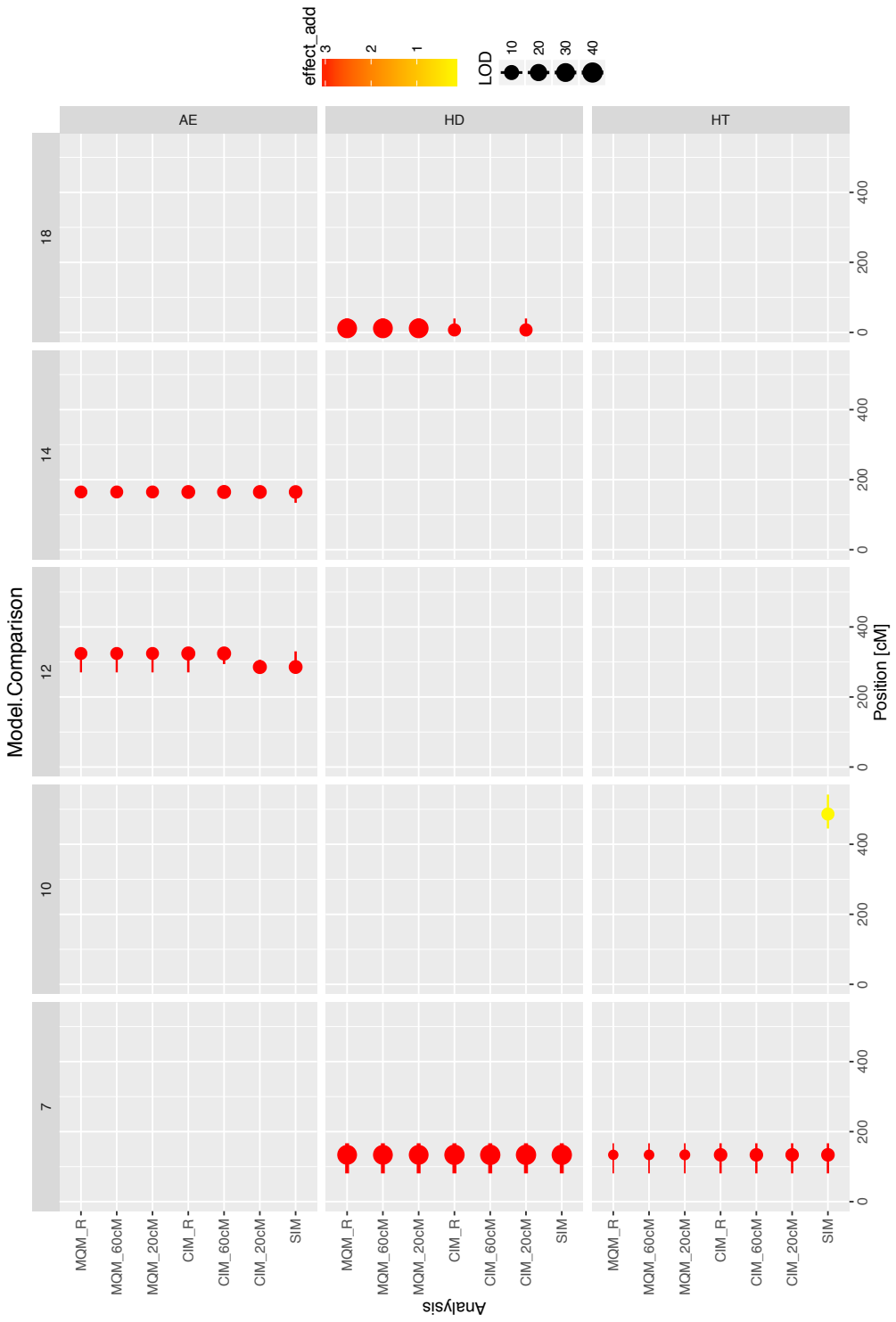

Figure 5.4: Differences in position and LOD scores of QTL detected by QTL mapping methods using different restriction window sizes for the inclusion of marker covariates. Linkage groups are indicated at the top of the plot.

5.3.3 QTLs detected

For the traits assessed, a detailed list of QTL discovered by restricted CIM, MQM and SIM is given in Table 5.2. Two AE QTLs were detected, one on chromosome 5B (AE_QTL1) and another on 6B (AE_QTL2). A single QTL on chromosome 2D was found to be associated with both HD and HT (HD +HT QTL1). CIM and MQM detected a second QTL for HD on chromosome 7B (HD_QTL2).

Table 5.2: Table of QTL discovered for number of anthers extruded per floret (AE), days to heading (HD) and plant height (HT).

Trait	Model	LG	Chr	Pos	LOD	% Var	Add Eff	+	CI Begin	CI End
AE	CIM	12	5B	121	6.8	9.1	0.2	T	101	124
AE	MQM	12	5B	121	3.8	9.1	0.2	T	101	124
AE	SIM	12	5B	107	6.1	8.9	0.2	T	103	124
AE	CIM	14	6B	64	6.1	8.8	0.2	F	60	70
AE	MQM	14	6B	64	3.8	8.8	0.2	F	60	70
AE	SIM	14	6B	64	5.6	8.2	0.2	F	52	70
HD	CIM	7	2D	41	40.9	45.0	2.4	F	25	51
HD	MQM	7	2D	41	36.8	45.2	2.4	F	25	51
HD	SIM	7	2D	41	40.9	45.9	2.4	F	25	51
HD	CIM	18	7B	4	4.2	3.5	0.6	T	0.0	22
HD	MQM	18	7B	6	36.8	3.5	0.6	T	0.0	22
HT	CIM	7	2D	41	5.3	7.9	3.1	T	25	51
HT	MQM	7	2D	41	1.2	7.9	3.1	T	25	51
HT	SIM	7	2D	41	5.4	5.9	2.8	T	25	51
HT	SIM	10	4A	162	4.3	4.3	2.2	T	148	181

Abbreviations, LG linkage group, Chr chromosome, Pos position, LOD logarithm of odds, % Var percentage variance explained, Add Eff additive effect, CI begin start position of confidence interval, CI end end position of confidence interval, + positive allele coming from KWS Ferrum (F) or KWS Tempo (T).

5.3.3.1 QTL effects

The KWS Tempo allele of HD/HT_QTL1 (2D) was associated with later flowering and greater plant height (Figure 5.5). HD_QTL2 (7B) was associated with a less pronounced effect on days to heading as compared to HD_QTL1 (Figure 5.5).

For AE_QTL1 (5B) and AE_QTL2 (6B), greater anther extrusion was inherited from KWS Tempo and KWS Ferrum respectively (Figure 5.4 C).

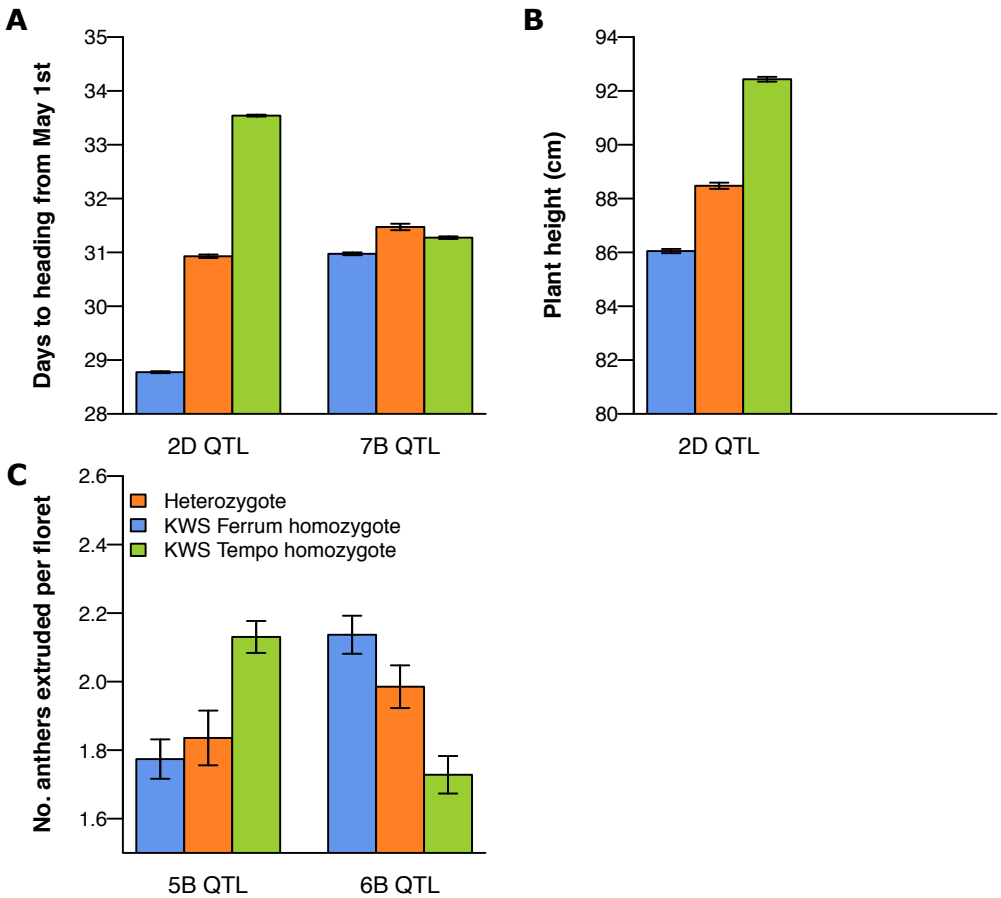

Figure 5.5: Allelic effects for QTL detected; **A)** days to heading; **B)** plant height and **C)** number of anthers extruded per floret.

Means and ±S.E. are presented.

5.3.2.2 LOD profiles and confidence intervals

All three QTL mapping methods detected the identical HD QTL on linkage group 7 (2D); with no differences in confidence intervals or position of the peak marker (Figure 5.6).

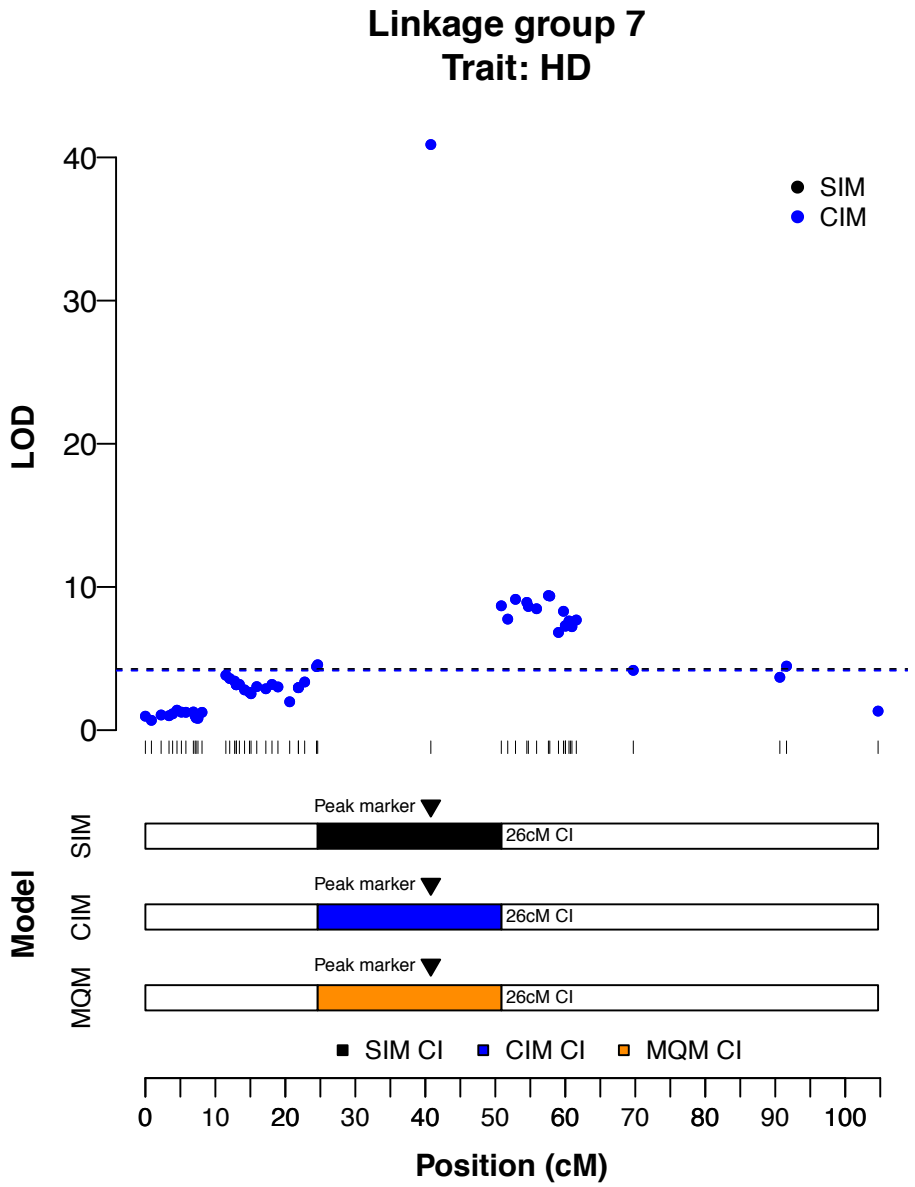

Figure 5.6: LOD profile for trait HD along linkage group 7.

Blue and black dashed lines indicate the declared significance threshold for CIM and SIM and each coloured dot represents a single marker used in linkage mapping. A graphical representation of linkage group 7 is present below the LOD profile, with the position and confidence interval of the putative QTL, as determined by SIM, CIM and MQM models, being highlighted.

The size of the confidence intervals for the HD QTL detected on linkage group 18 (7B) were identical between CIM and MQM; the position of the peak marker however differed (Figure 5.7). The SIM LOD profile was similar to that of CIM but did not pass the significance threshold (Figure 5.7).

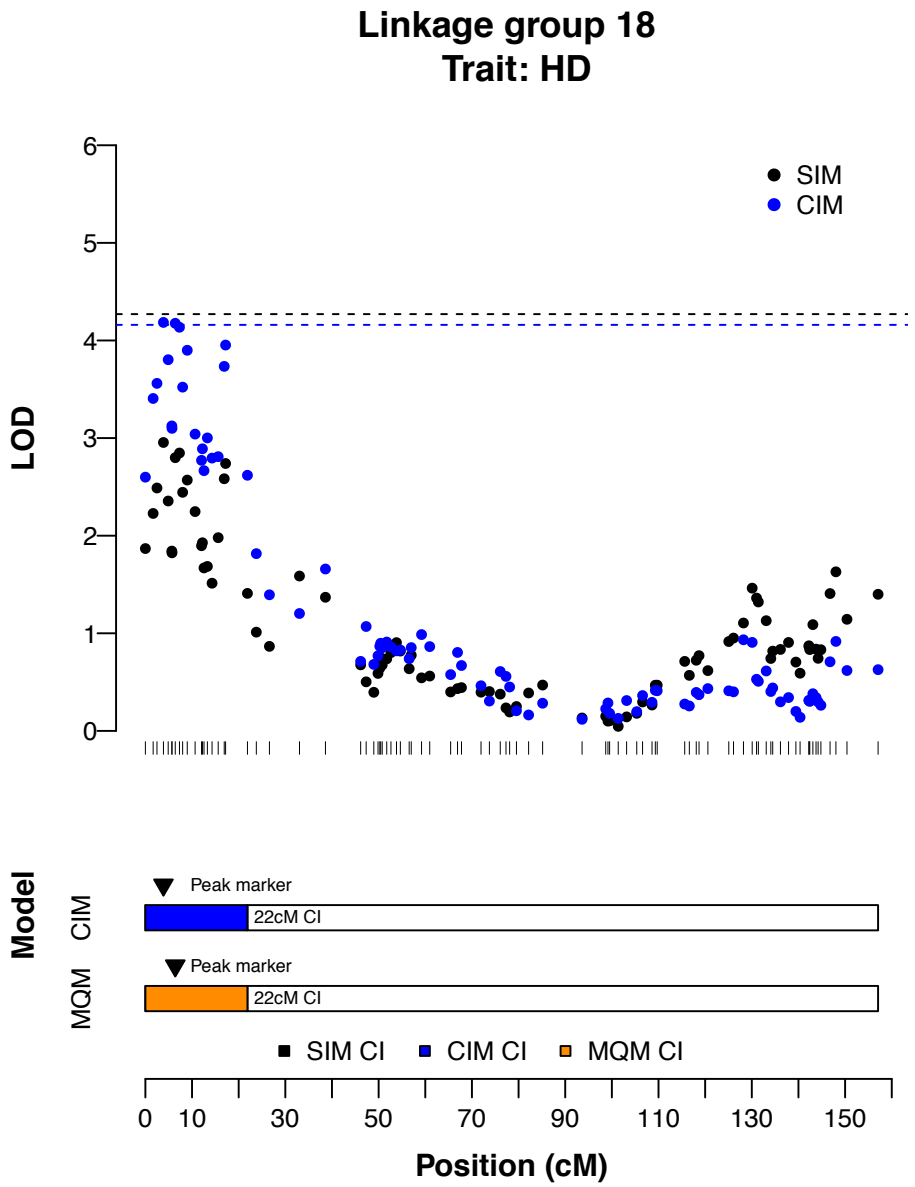

Figure 5.7: LOD profile for trait HD along linkage group 18.

Blue and black dashed lines indicate the declared significance threshold for CIM and SIM and each coloured dot represents a single marker used in linkage mapping. A graphical representation of linkage group 18 is present below the LOD profile, with the position and confidence interval of the putative QTL, as determined by CIM and MQM models being highlighted.

The identical QTL region detected for HD on linkage group 7 (2D) was detected for HT; LOD scores were considerably lower for HT as compared to HD (Figure 5.8).

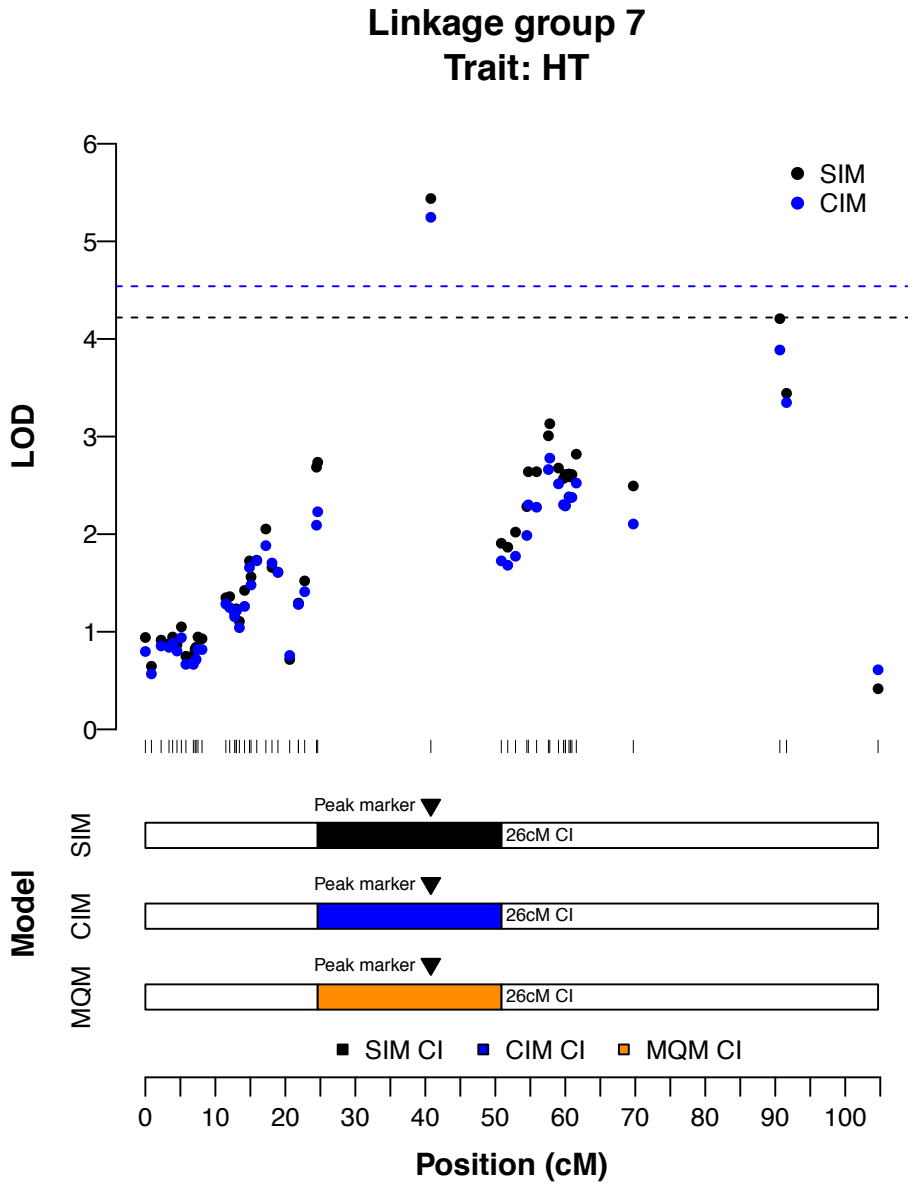

Figure 5.8: LOD profile for trait HT along linkage group 7.

Blue and black dashed lines indicate the declared significance threshold for CIM and SIM and each coloured dot represents a single marker used in linkage mapping. A graphical representation of linkage group 7 is present below the LOD profile, with the position and confidence interval of the putative QTL, as determined by SIM, CIM and MQM models being highlighted.

SIM, CIM and MQM detected the same AE QTL on linkage group 12 (5B); the peak marker was localised closer to the right limit of the confidence interval (CI) with CIM and MQM (Figure 5.9). With both SIM and CIM two LOD score peaks were apparent (Figure 5.9), only the second peak was declared within the QTL CI using the LOD-drop-off method.

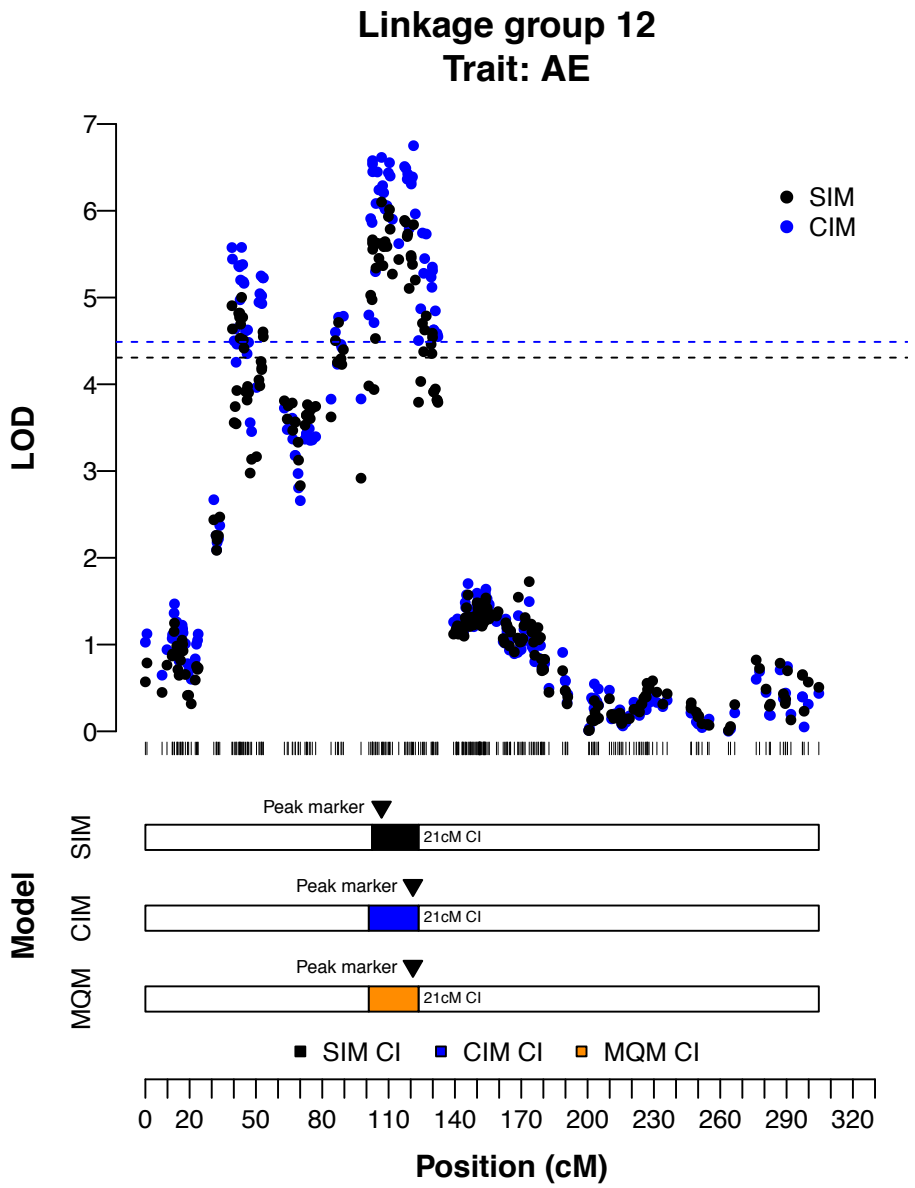

Figure 5.9: LOD profile for trait AE along linkage group 12.

Blue and black dashed lines indicate the declared significance threshold for CIM and SIM and each coloured dot represents a single marker used in linkage mapping. A graphical representation of linkage group 12 is present below the LOD profile, with the position and confidence interval of the putative QTL, as determined by SIM, CIM and MQM models being highlighted.

An additional AE QTL was detected on linkage group 14 (6B); the size of the confidence interval (CI) was narrower than that observed on linkage group 12 (Figure 5.10). CIM and MQM both narrowed the size of the CI and repositioned the peak marker towards the left limit of the CI (Figure 5.10).

Blue and black dashed lines indicate the declared significance threshold for

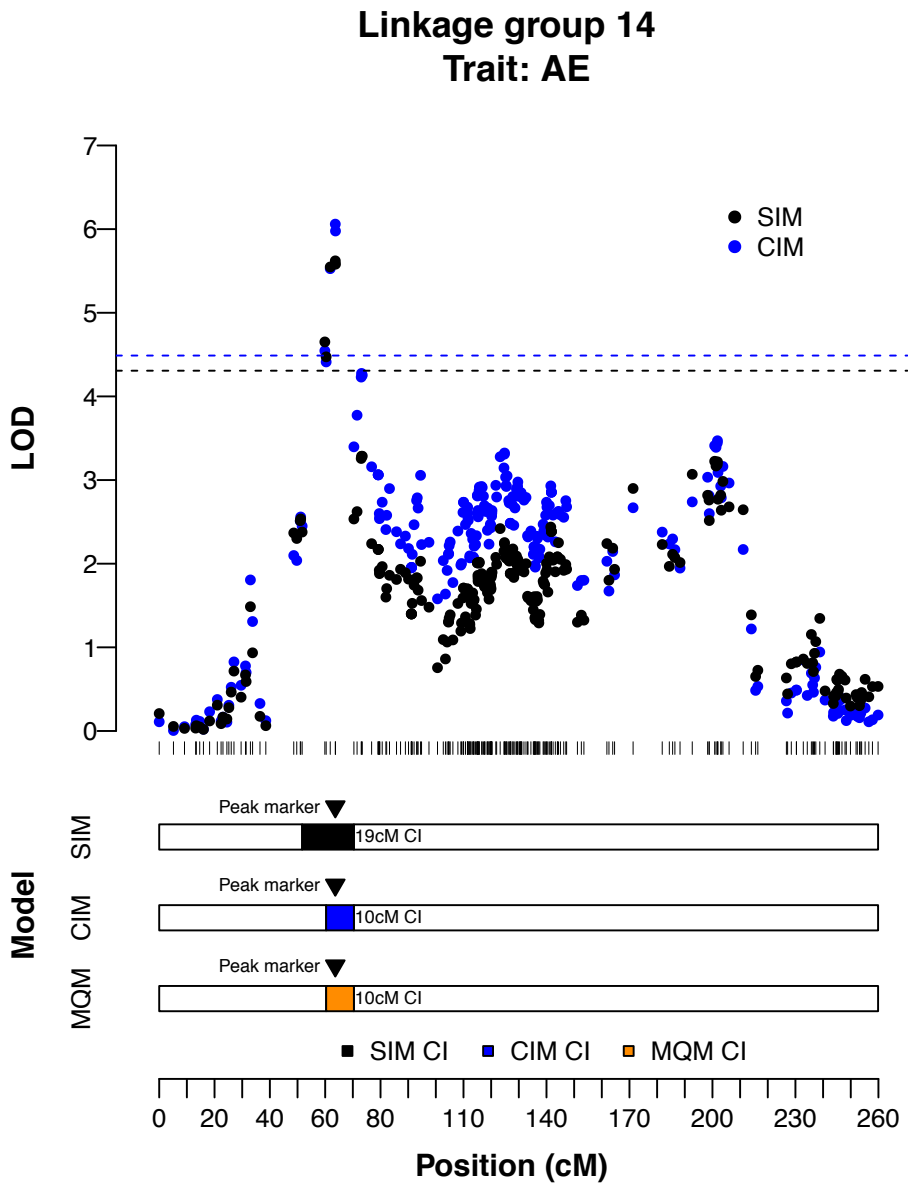

Figure 5.10: LOD profile for trait AE along linkage group 14.

CIM and SIM and each coloured dot represents a single marker used in linkage mapping. A graphical representation of linkage group 14 is present below the LOD profile, with the position and confidence interval of the putative QTL, as determined by SIM, CIM and MQM models being highlighted.

5.3.2.3 Identification of putative anther extrusion genes

The number of genes falling within the 5B AE QTL confidence interval was found to be fifteen times that of the 6B QTL; 744 and 50 respectively (Table 5.3). Among these genes, a large proprotion showed no expression within the spike during the developmental time course of *T. aestivum* cv. Chinese Spring (Table 5.3) (Borrill et al., 2016) and were therefore excluded from subsequent analyses. In total, 247 spike expressed genes were identified within the 5B QTL region, while seven were identified within the 6B QTL region.

Table 5.3: The number of genes falling into different categories during the process of narrowing the list of putative candidate genes within the two anther extrusion QTLs on chromosomes 5B and 6B.

Chr	No. genes within confidence interval	No. genes with expression in spike
5B	744	247
6B	50	7

Abbreviation, Chr Chromosome.

Within the subset of spike expressed genes, 13 on chromosome 5B and two on chromosome 6B were found to have significantly higher expression in the spike as compared to leaves and/or shoots (Table 5.4). These genes were selected as anther extrusion candidates.

Table 5.4: Transcripts of candidate genes within QTL regions on chromosome 5B and 6B showing higher expression in the spike as compared to leaves and/or shoots.

Transcript stable ID	Expression higher in spike than leaves	Expression higher in spike than shoot
TRIAE_CS42_5BL_TGACv1_404111_AA1284370.1	X	X
TRIAE_CS42_5BL_TGACv1_404224_AA1291270.3	X	
TRIAE_CS42_5BL_TGACv1_404360_AA1296920.1	X	
TRIAE_CS42_5BL_TGACv1_404421_AA1299400.2	X	X
TRIAE_CS42_5BL_TGACv1_404530_AA1303080.2	X	X
TRIAE_CS42_5BL_TGACv1_404957_AA1316330.6	X	X
TRIAE_CS42_5BL_TGACv1_405427_AA1327120.3	X	
TRIAE_CS42_5BL_TGACv1_405517_AA1329190.1	X	
TRIAE_CS42_5BL_TGACv1_405693_AA1333220.1	X	X
TRIAE_CS42_5BL_TGACv1_405891_AA1337090.1	X	X
TRIAE_CS42_5BL_TGACv1_406098_AA1340730.1		X
TRIAE_CS42_5BL_TGACv1_406602_AA1347640.2	X	
TRIAE_CS42_5BL_TGACv1_406894_AA1351250.3	X	
TRIAE_CS42_6BS_TGACv1_513188_AA1633570.1	X	X
TRIAE_CS42_6BS_TGACv1_514834_AA1664880.1	X	X

Significant differences have been determined at the $P<0.05$ level. Abbreviations, sp spike; lf leaves; sh shoots; exp expression.

A *T. aestivum* ortholog of the *A. thaliana* genes *NPR5* (*BOP2*) (Source: UniProtKB/Swiss-Prot; Acc: Q9ZVC2) and *NPR6* (*BOP1*) (Source: UniProtKB/Swiss-Prot; Acc: Q9M1I7) was present within the final subset of candidate genes for the 5B QTL (Table 5.5).

An orthologous relationship to previously characterized *A. thaliana* genes was not found for the two candidates identified within the 6B QTL (Table 5.5).

Table 5.5: Putative candidate genes within QTL regions and their orthologous *Arabidopsis thaliana* (AT) genes.

Gene ID	Gene description	AT Gene
TRIAE_CS42_5BL_TGACv1_404111_AA1284370		
TRIAE_CS42_5BL_TGACv1_404224_AA1291270		
TRIAE_CS42_5BL_TGACv1_404360_AA1296920		CXE1,3,5,7, 4,13,15
TRIAE_CS42_5BL_TGACv1_404421_AA1299400		NPR5, NPR6
TRIAE_CS42_5BL_TGACv1_404530_AA1303080		
TRIAE_CS42_5BL_TGACv1_404957_AA1316330	rRNA N-glycosidase	
TRIAE_CS42_5BL_TGACv1_405427_AA1327120		
TRIAE_CS42_5BL_TGACv1_405517_AA1329190	Patatin	PLP1, 3, 4, 5
TRIAE_CS42_5BL_TGACv1_405693_AA1333220		
TRIAE_CS42_5BL_TGACv1_405891_AA1337090		ATHB5,6,16
TRIAE_CS42_5BL_TGACv1_406098_AA1340730		NAK APK1A, B BIK1
TRIAE_CS42_5BL_TGACv1_406602_AA1347640		
TRIAE_CS42_5BL_TGACv1_406894_AA1351250		
TRIAE_CS42_6BS_TGACv1_513188_AA1633570		
TRIAE_CS42_6BS_TGACv1_514834_AA1664880		

5.4 Discussion

5.4.1 Anther extrusion capacity; phenotypic variation within population

KWS Ferrum and KWS Tempo have previously been characterised as high and moderate anther extruders respectively. During the course of the present study, yield plots of each parental variety were grown by KWS UK Ltd, from which anther extrusion was assessed for both. The average number of anthers extruded per floret (AE) was found to be 3.0 (maximum) for KWS Ferrum and and 1.5 for KWS Tempo. A number of F3:F4 lines had an average AE less than that observed for KWS Tempo (Figure 5.2 C), suggesting that multiple anther extrusion loci are segregating within the population and AE enhancing alleles are possessed by both parents.

Prior studies have reported anther extrusion capacity to be correlated with plant height and heading date (Boeven et al., 2016, Beri and Anand, 1971). Within the bi-parental population used in the present study, weak correlations were observed between anther extrusion and these phenological characteristics. It is therefore unlikely that *Rht* or *Ppd/Eps* loci exerted a large influence on anther extrusion capacity within the population.

5.4.2 QTL detection methods

5.4.1.1 Comparison of composite interval mapping (CIM), multiple QTL mapping (MQM) and single interval mapping (SIM)

The strengths of CIM and MQM are demonstrated by the results of plant height (HT) and days to heading (HD) QTL analyses. SIM reported a significant HT QTL on 4A that was not detected by either CIM or MQM. It is likely that this region co-segregates with the QTL on 2D within our population but is not directly associated with phenotypic variation. Conversely, SIM was not able to detect the HD QTL on 7B, possibly due to it being linked in repulsion to the QTL on 2D.

Across traits, significant QTLs detected by CIM were also detected by MQM. MQM LOD scores were lower for most QTL, with the exception of the HD QTL on 7B which was assigned a considerably lower LOD score with CIM. The significance of this QTL may have been underestimated by CIM due to the large effect of the 2D HD QTL.

5.4.1.2 Selection of marker covariates

Within the literature there is no general consensus with respect to appropriate exclusion window sizes to use in the selection of marker covariates for CIM and MQM. Population size and the decay of linkage disequilibrium (LD) with genetic

distance are likely to influence the effect of window size on subsequent analyses.

Our population was characterised by very slow decay of LD. At an inter-marker distance of 20 cM LD was considerably high, r^2 being 0.89. At 60 cM r^2 values were within a more reasonable range, on average 0.57. Comparisons were therefore carried out between analysis excluding marker covariates on the same chromosome, by setting the exclusion window size to 2000 cM, and analyses using window sizes of 20 cM and 60 cM.

For AE, all analyses detected a QTL within the same region of chromosome 5B. The position of the peak marker was shifted with CIM analyses as compared to SIM; to varied extents depending on the exclusion window size used. Confidence intervals were notably smaller for the 20 cM and 60 cM CIM models, compared to SIM and restricted CIM. The size of the exclusion window was found to have no effect on the peak marker or size of the confidence interval reported by MQM.

Marker covariates that are closely linked to a region being tested are unlikely to contribute to a more informative model. MQM is capable of removing these covariates from the maximal model as it employs a backward selection step. For the 5B AE QTL, results using restricted CIM were identical to that of MQM, with respect to position of the peak marker and

the size of the confidence interval. As LD decay is variable along each chromosome, a given window exclusion size may be more appropriate for some chromosomes than others. We therefore concluded that restricted CIM and MQM provided a more robust and less biased approach for our analyses.

5.4.3 Anther extrusion QTL detected

Due to the distribution of phenotypic values observed, it was hypothesized that more than one anther extrusion QTL would be present within the population. This was found to be true, with an anther extrusion QTL being detected on linkage group 12 and another on linkage group 14, which correspond to chromosomes 5B and 6B respectively (Table 5.2). Anther extrusion enhancing alleles were found to be inherited from both parents, with KWS Ferrum and KWS Tempo contributing the positive allele for the 5B and 6B QTLs respectively (Table 5.2).

Within the literature, there are a disproportionate number of investigations into the genetic architecture of anther extrusion, as compared to that of other floral traits. It appears that prior to the re-visiting of hybrid breeding as an avenue for yield increase, interest in the trait was sustained by pathologists aiming to elucidate the relationship between

anther extrusion and the severity of Fusarium head blight (Skinnes et al., 2010).

In the past ten years, a number of association mapping and linkage mapping experiments targeting anther extrusion have been conducted using a variety of marker platforms (Skinnes et al., 2010, Lu et al., 2013, Buerstmayr and Buerstmayr, 2015, Boeven et al., 2016, Muqaddasi et al., 2017, Muqaddasi et al., 2016). The results of these studies suggest that anther extrusion is in fact quantitatively inherited; with a number of marker-trait associations and QTLs being reported across the wheat genome.

The two anther extrusion QTLs detected in the present study each had an additive effect of 0.2 anthers per floret and explained approximately nine percent of the phenotypic variation within the population. These values are comparable to that reported in previous linkage mapping experiments (Skinnes et al., 2010, Lu et al., 2013, Buerstmayr and Buerstmayr, 2015) and add further support for the hypothesis that anther extrusion is governed by a number of loci each having relatively small effects on the trait.

As mentioned previously, both parental lines appear to carry positive and negative AE alleles. Genotypes homozygous for the AE promoting alleles of both QTLs extruded an average of 2.3 anthers per floret as compared to 1.5 anthers for

genotypes with neither allele and 1.9 for genotypes homozygous for just one of the AE promoting alleles. This further demonstrates that the AE enhancing alleles from both QTLs are capable of producing tangible additive phenotypic effects within the population. While each QTL was associated with a modest phenotypic effect, the simultaneous targeting of both loci appears to be a feasible approach towards enhancing anther extrusion capacity within KWS breeding pools.

Assessing the phenotypic effects of loci identified by linkage mapping, and the interaction between loci, in an F4 population is inherently limited due to the relatively high proportion of heterozygous individuals (Takuno et al., 2012). Subsequent to the validation of the QTLs in a second field trial; further insight can be gained by investigating the phenotypic effects of these QTLs in the recombinant inbred lines (RILs) (i.e. the F6/F7 generation) descended from this population (Takuno et al., 2012).

Buerstmayr and Buerstmayr et al. (2015) reported no overlap between AE QTLs detected in their study and those reported in previous linkage mapping studies (Skinnes et al., 2010, Lu et al., 2013), leading the authors to assert that the phenotypic effects of AE loci are largely dependent on genetic background. Further support for this assertion is provided by

the results of AE association mapping studies; while *Rht-D1* was found to be a major AE QTL in the panel investigated by Boeven et al. (2016), the locus was not found to be significantly associated with anther extrusion by Muqaddasi et al. (2016). The work of Muqaddasi et al. (2017) does however provide evidence for robust AE loci; with AE marker-trait associations for a winter and spring wheat panel having been detected 2 cM apart on chromosome 6A.

Anther extrusion QTLs have been reported by previous linkage mapping studies on both chromosomes 5B (Lu et al., 2013, Buerstmayr and Buerstmayr, 2015) and 6B (Buerstmayr and Buerstmayr, 2015); the present study detected AE QTLs on both these chromosomes. However, the use of DArT markers in the aforementioned studies prevents direct comparisons from being made between the localization of these QTLs and those detected in the present study.

A key area for continuation of the the present study's work would entail investigating how robust the phenotypic effects of the described QTLs are across genetic backgrounds. This can be achieved by introgressing QTL regions from each parent into additional varieties or alternatively by generating a multi-parent advanced generation inter-cross (MAGIC) population with KWS Tempo and KWS Ferrum. MAGIC populations have been previously demonstrated to be a

powerful genetic resource in *T. aestivum* (Mackay et al., 2014). This work will help to determine whether population specific effects have been observed in the present study and ultimately how practical it would be to target these loci in additional breeding pools.

While the AE QTLs described in the present study have been identified using a single trial's data, due to the high heritability observed for anther extrusion across locations and across years in Chapter 3, it is likely that these two QTLs would be detected across trials. Conclusively demonstrating the robustness of these loci can be achieved by carrying out linkage mapping with the F5 population derived from the F4 population described in the present chapter. This work may also be of value towards fine mapping, the need for which is described in the following section.

5.4.3 Towards identifying the causative AE loci

Fine mapping of the anther extrusion QTL regions on chromosomes 5B and 6B is pre-requisite for the identification of causative loci within them. Typically this is achieved by LM using a population with a greater frequency of recombination events within the target region (Röder et al., 2008, Cuthbert et al., 2006). Despite not having such a population available for a second field trial (Section 5.2.3) an effort was made to identify putative causative loci within these regions using the

same broad filters applied to the results of association mapping (Chapter 4).

The relatively larger size of the 5B QTL (21 cM) compared to the 6B QTL (10 cM) was reflected in the number of genes overlapping each; approximately fifteen times more genes were found in the 5B region (Table 5.3). Amongst the list of candidates described in Table 5.4, a *T. aestivum* ortholog of the *A. thaliana* genes *NPR5* (*BOP2*) and *NPR6* (*BOP1*) was present. In *A. thaliana* both genes are actively expressed during early reproductive development and are known to govern floral meristem determinacy (Ha et al., 2004). Further to this, Jost et al. (2016) recently reported that an allelic variant of a *BOP1* and *BOP2* ortholog in *Hordeum vulgare* results in the development of a more relaxed inflorescence which would be expected to promote anther extrusion. The putative *T. aestivum BOP1* and *BOP2* ortholog therefore represents a promising candidate for the causative loci of the 5B anther extrusion QTL.

A more refined approach towards narrowing the list of candidate AE loci could have been applied, as proposed in Section 5.2.3. In this instance, RNA-seq analysis would be carried out on the F5 lines, derived from the F4 population, showing recombination events within the 5B and 6B QTL regions. These efforts would have complemented fine-

mapping via linkage mapping in this population; overlaying expression data, phenotypic data and the recombination landscape would have confirmed conclusions drawn from fine-mapping as to the genomic location of the causative locus and provided its probable identity.

5.5 Summary

In the present linkage mapping study, two anther extrusion QTLs were detected within the population investigated; one on chromosome 5B and another on 6B. The two QTLs showed a modest influence on anther extrusion capacity, each explaining approximately ten percent of the phenotypic variation observed. Results indicate fine mapping targeting both QTLs is required, especially in the case of the 5B QTL which spanned a 21 cM distance.

A number of putative candidate genes, including an ortholog of *A. thaliana BOP1* and *BOP2*, have been identified within these QTL regions. Areas for continuation of the work described in the present chapter include fine mapping of each QTL region in order to narrow the list of candidate genes; potentially using F5 lines derived from the F4 population used. Simultaneous to this, the phenotypic effects of candidate genes can be investigated using publically available genomic resources available such as the *T. aestivum* cv. Cadenza TILLING population (Uauy et al., 2009).

CHAPTER 6: PHENOTYPING TILLING LINES WITH MUTATIONS IN CANDIDATE FLORAL TRAIT GENES.

6.1 Introduction

6.1.1 Candidate genes

From the results of association mapping and linkage mapping carried out in Chapter 4 and 5 respectively, candidate floral trait genes were identified which co-localise with detected QTLs (Chapter 4, Table 4.7 and 4.8; Chapter 5, Table 5.3 and 5.4) and show expression within the spike (Borrill et al., 2016).

Among the candidates, some were found, by *in silico* analysis, to be previously characterised in *T. aestivum* and their potential role as influencers of floral traits proposed. Others exhibit an orthologous relationship to genes known to be expressed in floral tissues in other species. Three candidate anther extrusion genes met these criteria, along with a single anther length candidate.

This subset of genes is the focus of the present chapter; with investigations into their influence on floral characteristics having been undertaken through the use of TILLING mutants. The TGACv1 and the corresponding CSS gene models for each candidate was obtained from the EnsemblPlants database (http://plants.ensembl.org/) and are listed in Table 6.1 together with the trait they may be associated with.

While it was not identified as a candidate for the 2A anther extrusion marker trait association (MTA) described in Chapter 4, TRIAE_CS42_2AL_TGACv1_097448_AA0324210 was also included for two reasons. The gene is homoeologus to TRIAE_CS42_2BL_TGACv1_129463_AA0384870, which was selected as the target candidate for the 2B anther extrusion marker-trait association. Additionally, based on the consensus genetic map, the 2A and 2B anther extrusion MTAs appear to localise to homoeologus positions on these chromosomes (Chapter 4, Section 4.3.2.7). Presently, no gene model corresponding to D genome homoeolog has been annotated within the TGACv1 assembly currently available on the EnsemblPlants database (http://plants.ensembl.org/).

Table 6.1: List of putative floral trait genes identified in Chapters 4 and 5 by association and linkage mapping.
The TGACv1, and their corresponding CSS gene models are listed. Abbreviations, Chr Chromosome.

Trait	Chr	TGAC Gene Model	CSS Gene Model
Anther extrusion	2A	TRIAE_CS42_2AL_TGACv1_097448_AA0324210	Traes_2AL_5BA7E2623
Anther extrusion	2B	TRIAE_CS42_2BL_TGACv1_129463_AA0384870	Traes_2BL_FE46FC938
Anther length	4A	TRIAE_CS42_4AL_TGACv1_290275_AA0983490	Traes_4AL_F4F5960D1
Anther extrusion	5B	TRIAE_CS42_5BL_TGACv1_404421_AA1299400	Traes_5BL_4780FA980

It should be emphasised that care must be taken in the interpretation of the results of the present study, as an observable phenotypic effect within a mutant line does not

conclusively prove that the candidate selected is in fact the causative loci within a QTL. However, if these candidates are found to have an influence on floral characteristics in *T. aestivum*, as in other species, these findings are likely to be of value towards the modification of these traits within wheat germplasm.

6.1.1.1 *TaAP2* homoeoloci *and* anther extrusion capacity

Investigations of linkage disequilibrium (LD) subsequent to association mapping (Chapter 4; Section 4.3.4), revealed that the gene model corresponding to *TaAP2-B* localised to a genomic region in high LD with the 2B anther extrusion MTA. Previously Ning et al. (2013b) characterised the homoeologus orthologs of the *Hordeum vulgare* cleistogamy gene *Cly1* (*HvAP2*) in hexaploid wheat (*T. aestivum* cv. Shinchunaga) ; *TaAP2-A*, *TaAP2-B* and *TaAP2-D*, which localise to the long arms of chromosomes 2A, 2B and 2D, respectively.

The *TaAP2* loci were reported as being similar in structure as well as transcription profile as *Cly1*; transcript abundance is highest in the florets, especially within the lodicules, at and immediately prior to the onset of anthesis (Ning et al., 2013b).

In *H. vulgare*, non-cleistogamous (open) flowering has been demonstrated to arise due to microRNA (miR172) guided

cleavage of *Cly1* transcripts; this enables the lodicules to fully develop and swell at the time of anthesis, forcing the palea and lemma open as they do so (Nair et al., 2010). A synonymous nucleotide substitution within the miR172 targeting site of *Cly1* serves to effectively suppress cleavage of its transcripts, resulting in cleistogamous flowering (Nair et al., 2010).

Ning et al. (2013a) re-sequenced wheat *AP2* homoeologs in an attempt to ascertain whether any naturally occurring variation existed. The accessions used in their study were all described as non-cleistogamous (Ning et al., 2013b), however due to the recessive nature of *Cly1* in *H. vulgare*, there still existed the possibility that functional variants were present at the miR172 targeting site of at least one homoeologs.

No functional variants were however detected, even amongst wild relatives, with a high level of gene conservation being reported (Ning et al., 2013a). Provided that the miR172 targeting sites of completely cleistogamous accessions have not yet been sequenced, the possibility that functional variants are present within existing germplasm cannot be definitively ruled out. Additionally, alternative mechanisms may be present in *T. aestivum* that down-regulate *TaAP2* expression.

6.1.1.2 Putative ortholog of Blade-*On-Petiole1/2* (*BOP1/2*) and anther extrusion capacity

Pleiotropic changes in spike architecture are observed in the *Hordeum vulgare laxatum-a* (*lax-a*) mutant (Laurie et al., 1996). Jost et al. (2016) applied map-based cloning and sequencing to identify the causal gene, which was found to be orthologous to the *Arabidopsis thaliana* genes *BOP1* and *BOP2*. Within the the confidence interval of the anther extrusion QTL detected in the present study on chromosome 5B, a putative *T. aestivum* ortholog of *BOP1* and *BOP2* was present (Chapter 5; Section 5.3.2.3).

The *lax-a* mutant produces a more relaxed inflorescence with less spikelets per unit of spike length relative to wild type, due to an elongation of the rachis internodes (Jost et al., 2016). This trait coupled to the florets developing a narrower palea and lemma appear to confer the mutant with the capacity for open flowering despite the conversion of its lodicules into stamenoid structures (Jost et al., 2016). Independently, any of these modifications to spike architecture would be expected to influence anther extrusion capacity; with the former two possibly enhancing the trait.

6.1.1.2 A putative ortholog of *GN* and changes to anther length

Within one of the genomic regions in high LD with the anther length marker-trait association detected on chromosome 4A, a putative ortholog of the *Arabidopsis thaliana* ARF guanine-nucleotide exchange factor *GNOM* (*GN*) was detected (Chapter 4; Section 4.3.4). In *A. thaliana*, *GN* is best known for its functional role in embryogenesis; however, the gene is known to be continually expressed during development, especially in actively dividing and elongating cells (Geldner et al., 2003).

Geldner et al. (2003) described a model for *GNOM*/*GN* action wherein it plays a role in mediating auxin transport during embryogenesis as well as post-embryonic organ development. The expression of the putative *T. aestivum GN* ortholog within the spike together with reports that auxin polar transport plays a critical role in the formation and development of the *A. thaliana* stamen (Cardarelli and Cecchetti, 2014) make it a candidate gene for influencing anther length.

6.1.2 TILLING populations

TILLING, a reverse genetic approach, has garnered significant attention from the wheat research community in the past decade, as it provides an avenue for the modification of gene function without the introduction of foreign DNA (Slade et al., 2005). TILLING populations typically consist of mutagenised individuals which serve as a library of DNA samples that can be screened for single-nucleotide polymorphisms (SNPs) induced by mutagenesis (Slade et al., 2005, Perry et al., 2003, McCallum et al., 2000).

Subsequent to its application as an effective functional genomics tool in model species (Slade et al., 2005, Perry et al., 2003, McCallum et al., 2000), the potential of TILLING as a tool in wheat genetic improvement was recognised. The approach has been demonstrated to be effective in the identification of *waxy* (Slade et al., 2005) *and VERNALIZATION1* (Chen and Dubcovsky, 2012) alleles with modified function.

Currently, two TILLING resources have been made available for the wheat research community via the www.wheat-tilling.com database (Krasileva et al., 2017), a tetraploid and a hexaploid TILLING population (*Triticum durum* cv. Kronos and *Triticum aestivum* cv. Cadenza, respectively). For each of these populations a library of induced SNPs within annotated

gene models are available together with a SIFT score for each variant, therein facilitating the identification of functional variants in target genes.

SIFT (Sorting Intolerant From Intolerant) is an approach based on sequence homology that enables predictions to be made as to the likelihood that substitutions will alter protein function, and by extension modify phenotype (Ng and Henikoff, 2003). SIFT assumes that key amino acids, required for proper functioning of protein domains, will be conserved within a protein family. The SIFT score is derived by obtaining an alignment of closely related proteins with the protein under investigation, and subsequently generating a score based on the amino acid differences. Variations in these key amino acids are weighed more heavily. Although opinions vary within the literature, SIFT scores less than 0.1 are typically described as deleterious (Ng and Henikoff, 2003).

Hexaploid *T.* aestivum appears to possess a greater tolerance for mutations, as compared to diploid species such as *A. thaliana,* attributable to the functional redundancy observed between homoeologus genes in the three genomes (Slade et al., 2005). While, Slade et al. (2005) reported apparently normal phenotypes in the majority of plants within the mutation library investigated, they emphasised that a pre-requisite for the application of TILLING lines in crop

improvement is carrying out backcrossing cycles to eliminate non-target background mutations.

The influence of these non-target mutations together with the functional redundancy described between genomes can either lead to masked phenotypes or phenotypes not linked to the target mutation. Backcrossing coupled to the generation of double and triple mutants is necessary for proof of gene function (Chen and Dubcovsky, 2012). The aim of the present chapter is therefore not to describe the role of the candidates in the determination of floral characteristics but rather to describe phenotypic variation in their corresponding TILLING lines. The results of the present study can provide a starting point for identifying targets for further investigation.

6.2 Methods

6.2.1 Accessing expression profile data for candidate genes

For each candidate, the TGACv1 gene model was queried within http://www.wheat-expression.com (Borrill et al., 2016) in order to obtain expression profiles from the developmental time-course of *T. aestivum* cv. Chinese Spring dataset. Transcript abundance (transcripts per kilobase million, tpm) was retrieved for five tissue types; grain, leaves, roots, spike and stem at the developmental stages shown in Table 6.2.

Table 6.2: Tissue types and developmental stages for which expression data is available within the developmental time course of Chinese Spring (http://www.wheat-expression.com database (Borrill et al., 2016)).

Stage	Growth phase	Grain	Leaves	Roots	Spike	Stem
1 cm spike	Vegetative					
Seedling stage	Vegetative					
Three leaf stage	Vegetative					
Two node stage	Vegetative					
Tillering stage	Vegetative					
2 dpa	Reproductive					
30 dpa	Reproductive					
14 dpa	Reproductive					
Anthesis	Reproductive					
Flag leaf stage	Reproductive					

Abbreviations, dpa days post anthesis. Coloured boxes (green) indicate developmental stages for which expression data is available for a given tissue.

6.2.2 Plant material

6.2.2.1 Identification of target mutations within a *T. aestivum* cv. Cadenza TILLING population

The CSS gene model corresponding to each candidate (Table 6.1) was used to search the www.wheat-tilling.com database (Krasileva et al., 2017) for available mutations and the lines in which they were present. Lines were selected for subsequent phenotypic analysis provided that they met the following criteria; 1) they were characterised as homozygous for the target mutation and 2) they possessed either missense variants (SIFT score < 0.1) or variants predicted to produce a premature stop codon. In instances where only a single line met these criteria for a candidate, a splice region/intron variant was also included for comparison.

Using the annotated TGACv1 gene models the position of target mutations within each candidate gene was determined. The translated protein product was also retrieved, accessed via EnsemblPlants (http://plants.ensembl.org/), in order to predict the effects each mutation had on the corresponding protein.

The mutant status of lines was confirmed by KASP genotyping, using genome specific primers targeting each mutation. Genomic DNA extraction from leaf tissue and genotyping was performed by KWS.

6.2.2.2 Glasshouse trial and phenotyping

Selected mutant lines were grown under controlled conditions as described in Section 2.2.3.1, together with wild type Cadenza, with each line represented by six biological replicates. Lines were phenotyped for either anther length or anther extrusion, dependant on their target mutation, as described in Sections 2.2.3.1.1 and 2.2.3.1.2 respectively. Additional phenotypic data was collected for each replicate; the number of fully formed spikelets and spike length i.e. the length from the first to the last rachis internode.

6.2.3 Statistical analysis of phenotypic data

Within the R environment, pairwise t-tests were conducted for each mutant line together with wild type in order to test for significant differences in each of the assessed traits. Significant differences were declared using threshold of a $P<0.05$.

6.3 Results

6.3.1 Expression profiles of candidate genes

6.3.1.1 The *TaAP2* homoeoloci; anther extrusion candidates

The gene TRIAE_CS42_2AL_TGACv1_097448_AA0324210.1, corresponding to *TaAP2-A*, is transcribed throughout spike development in Chinese Spring. Transcript abundance at anthesis was three times that observed at the two node stage (Figure 6.1). No tissues showed expression comparable to the spike at anthesis (Figure 6.1).

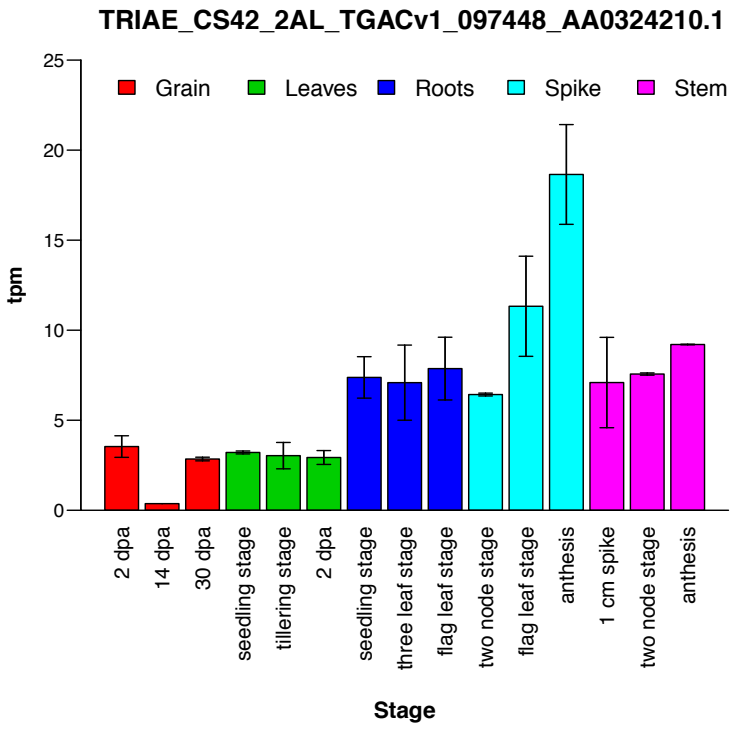

Figure 6.1: Expression profile of the candidate anther extrusion gene 2AL_TGACv1_097448_AA0324210.1 during the development of Chinese Spring (Borrill et al., 2016).
Means and +/- S.E. are presented for five tissues types; grain, leaves, roots, spike and stem. Abbreviations, TPM transcripts per kilobase million, dpa days post anthesis.

TRIAE_CS42_2BL_TGACv1_129463_AA0384870.1 (*TaAP2-B*) showed a similar transcription profile as was observed for its 2A homoeolog (Figure 6.1). Within the spike, an increase in transcript abundance accompanied the transition from vegetative to reproductive growth (Figure 6.2). Transcript abundance at anthesis was lower than that observed for *TaAP2-A*; 13.97 versus 18.64 tpm (Figure 6.1 and 6.2).

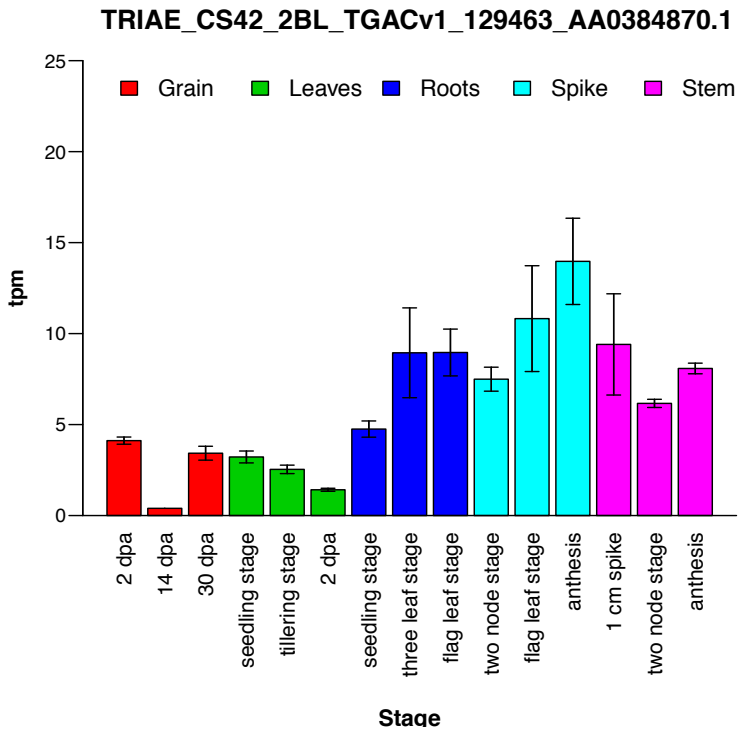

Figure 6.2: Expression profile of the candidate anther extrusion gene 2BL_TGACv1_129463_AA0384870.1 during the development of Chinese Spring (Borrill et al., 2016).

Means and +/- S.E. are presented for five tissues types; grain, leaves, roots, spike and stem. Abbreviations, TPM transcripts per kilobase million, dpa days post anthesis.

6.3.1.2 Putative ortholog of *Blade-On-Petiole1/2* (*BOP1/2*); an anther extrusion candidate

Transcript abundance was found to be highest within the spike for TRIAE_CS42_5BL_TGACv1_404421_AA1299400.2, a putative ortholog of *A. thaliana NPR5/6* (*BOP1/2*) (Figure 6.3). Within the spike, a down-regulation of expression accompanies the transition from the flag leaf to anthesis stage (Figure 6.3).

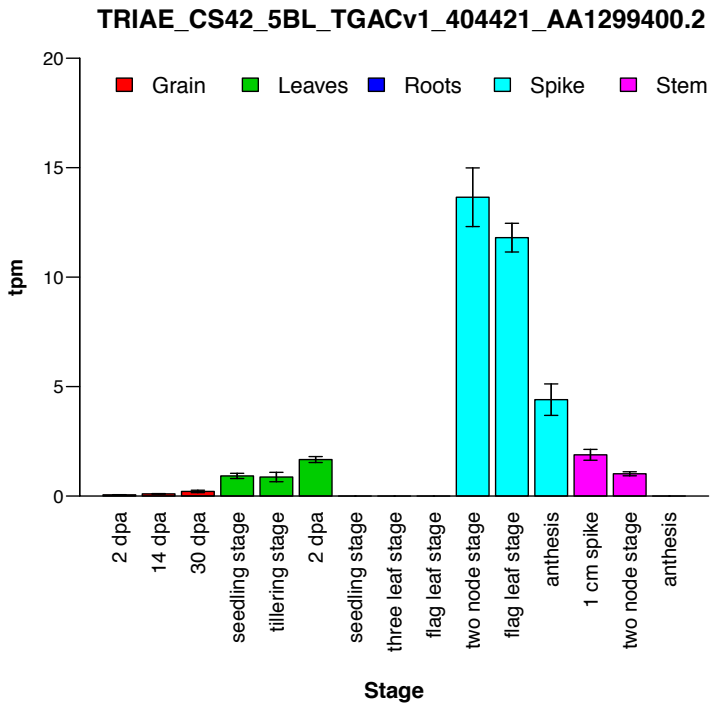

Figure 6.3: Expression profile of the candidate anther extrusion gene 5BL_TGACv1_404421_AA1299400.2 during the development of Chinese Spring (Borrill et al., 2016).
Means and +/- S.E. are presented for five tissues types; grain, leaves, roots, spike and stem. Abbreviations, TPM transcripts per kilobase million, dpa days post anthesis.

6.3.1.3 Putative ortholog of *GN*; an anther length candidate

TRIAE_CS42_4AL_TGACv1_290275_AA0983490.1, a putative ortholog of *A. thaliana GN/GNL1*, shows the highest expression within the spike during early developmental stages (two node and flag leaf stages). A reduction in transcription accompanies the transition from the flag leaf to anthesis stage (Figure 6.4).

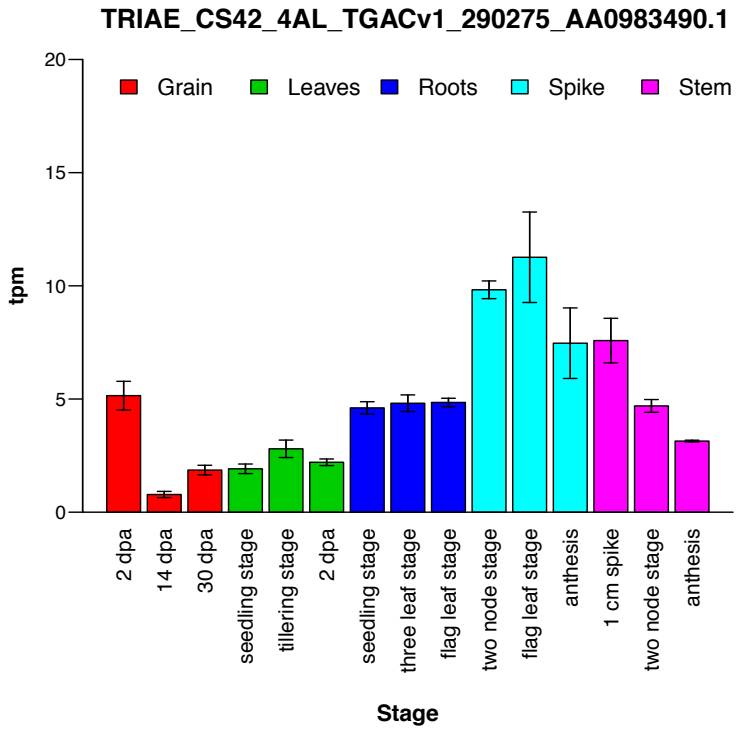

Figure 6.4: Expression profile of the candidate anther length gene 4AL_TGACv1_290275_AA0983490.1 during the development of Chinese Spring (Borrill et al., 2016).
Means and +/- S.E. are presented for five tissues types; grain, leaves, roots, spike and stem. Abbreviations, TPM transcripts per kilobase million, dpa days post anthesis.

6.3.2 Selected mutant lines and target mutations

In total, ten TILLING Cadenza lines were selected for phenotypic characterisation; 0747, 1556, 0064, 1806, 1778, 1759, 0755, 1458, 1004, 0179 (Table 6.3). The position of their corresponding target mutations within candidate genes is described subsequently.

Table 6.3: List of mutations targeted in the present chapter.
The *Triticum aestivum* cv. Cadenza line in which each mutation is presented is listed together with the variant type and its corresponding SIFT score.

Code	CSS Gene Model	Line	Variant	SIFT
2A Mutation1	Traes_2AL_5BA7E2623	0747	Splice region variant & intron variant	Na
2A Mutation2	Traes_2AL_5BA7E2623	1556	Missense variant & splice region variant	0.05
2B Mutation1	Traes_2BL_FE46FC938	0064	Splice region variant & intron variant	Na
2B Mutation2	Traes_2BL_FE46FC938	1806	Missense variant & splice region variant	0.01
4A Mutation1	Traes_4AL_F4F5960D1	1778	Stop gained	Na
4A Mutation2	Traes_4AL_F4F5960D1	1759	Stop gained	Na
4A Mutation3	Traes_4AL_F4F5960D1	0755	Stop gained	Na
5B Mutation1	Traes_5BL_4780FA980	1458	Missense variant	0.02
5B Mutation2	Traes_5BL_4780FA980	1004	Missense variant	0.00
5B Mutation3	Traes_5BL_4780FA980	0179	Missense variant	0.03

Data extracted from the www.wheat-tilling.com database.

2A_Mutation1 (Cadenza 0747) is present within the ninth intron of the TRIAE_CS42_2AL_TGACv1_097448_AA0324210 gene model; resulting in a guanine residue being subsisted by adenine (Figure 6.5 A). 2A_Mutation2 (Cadenza 1556) is present within the seventh exon; the mutation is missense resulting in a cytosine residue being substituted by thymine (Figure 6.5 A). Within the translated protein, 2A_Mutation2 results in an amino acid substitution outside of the two annotated AP2/ERF domains (Figure 6.5 B).

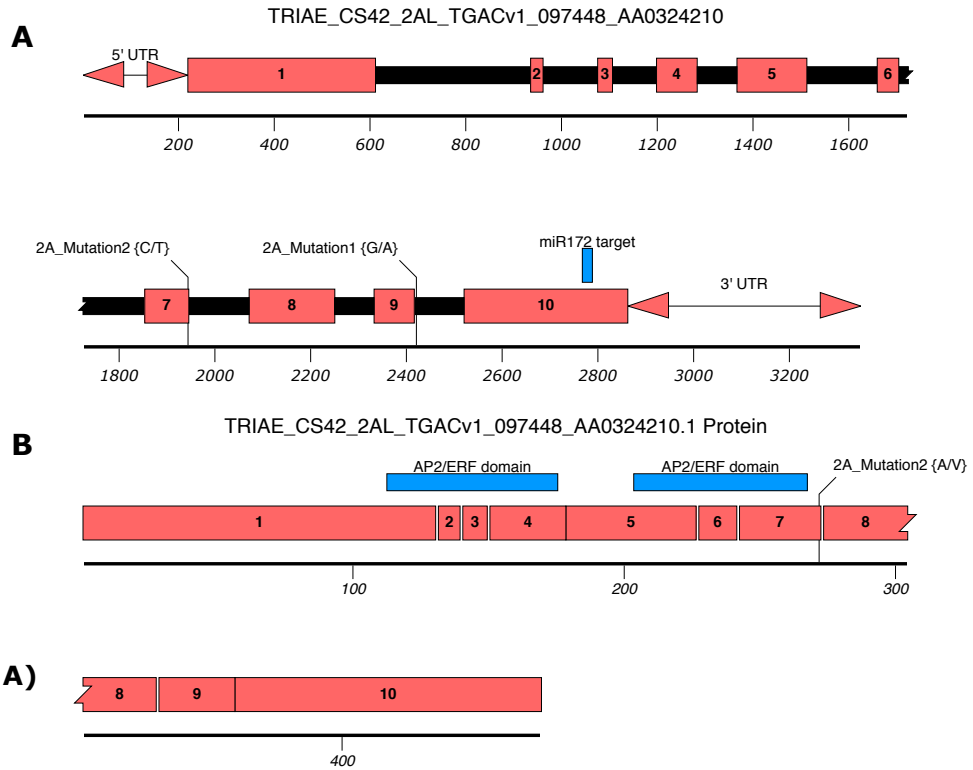

Figure 6.5: **A)** Gene model of candidate anther extrusion gene 2AL_TGACv1_097448_AA0324210. **B)** Corresponding translated protein model.

Red and black rectangles have been used to represent the exon and intron sequences respectively, while the 5' and 3' UTR are represented by double headed arrows. **B)** The amino acid sequences coded for by each exon are represented by numbered red rectangles while blue rectangles above the sequence indicate annotated protein domains; residue overlap splice sites are represented by gaps.

2B_Mutation1 (Cadenza 0064) is present within the fifth intron of the TRIAE_CS42_2BL_TGACv1_129463_AA0384870 gene model, resulting in a cytosine residue being substituted by thymine (Figure 6.6 A). 2B_Mutation2 (Cadenza 1806) is present within the fourth exon; the mutation is missense, resulting in a guanine residue being subsisted by adenine (Figure 6.6 A). Within the translated protein, 2B_Mutation2 results in an amino acid substitution within an annotated AP2/ERF domain (Figure 6.6 B)

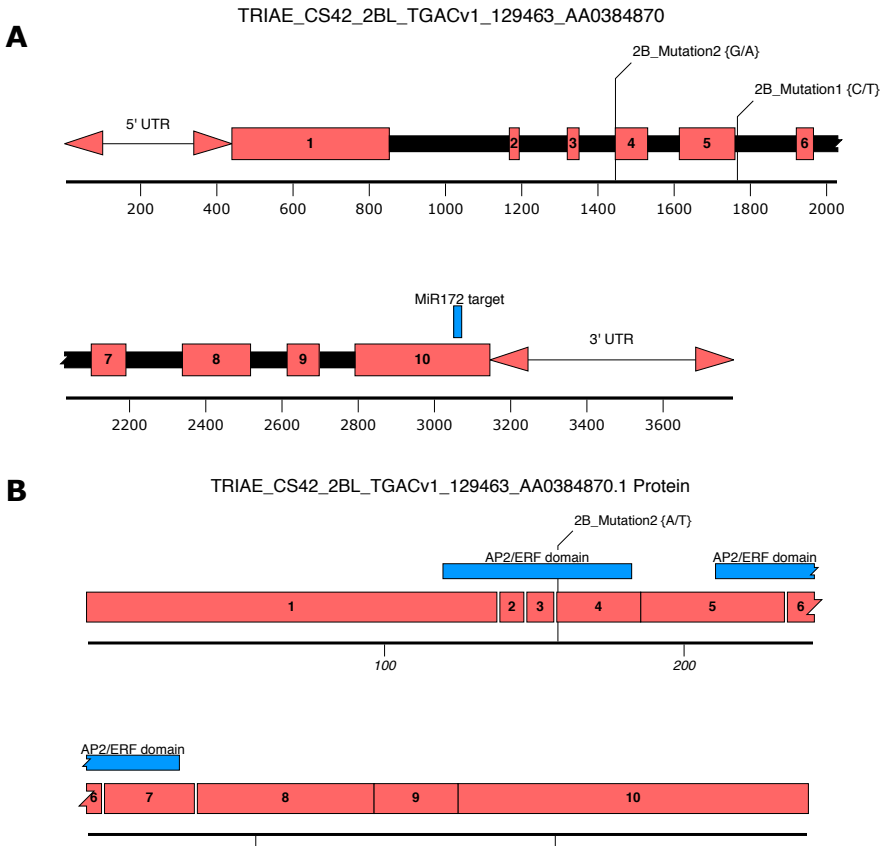

Figure 6.6: **A)** Gene model of candidate anther extrusion gene 2BL_TGACv1_129463_AA0384870. **B)** Corresponding translated protein model.

A) Red and black rectangles have been used to represent the exon and intron sequences respectively, while the 5' and 3' UTR are represented by double headed arrows. **B)** The amino acid sequences coded for by each exon are represented by numbered red rectangles while blue rectangles above the sequence indicate annotated protein domains; residue overlap splice sites are represented by gaps.

All three target mutations, 4A_Mutation1 (Cadenza 1778), 4A_Mutation2 (Cadenza 1759) and 4A_Mutation3 (Cadenza 0755) are present within the second intron of the gene model for TRIAE_CS42_4AL_TGACv1_290275_AA0983490 (Figure 6.7 A). 4A_Mutation2 and 3 both result in a substitution of a cytosine residue by thymine; 4A_Mutation1 results in a guanine residue being subsisted by adenine (Figure 6.7 A).

All three mutations are predicted to result in truncated proteins being translated due to the introduction of premature stop codons (Figure 6.7 B). The proteins resulting from 4A_Mutation1 and 3 are predicted to have direct effects on function of the annotated Sec7 domain (Figure 6.7 B).

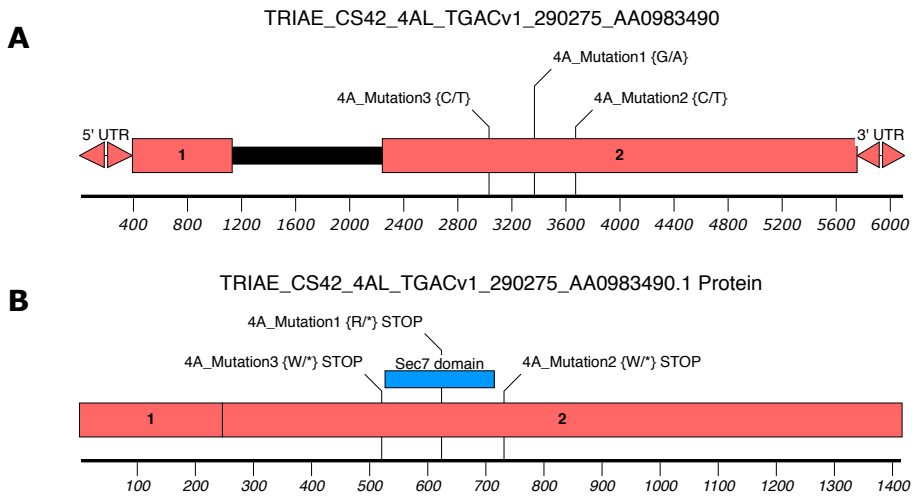

Figure 6.7: Gene model of candidate anther length gene 4AL_TGACv1_290275_AA0983490. **B)** Corresponding translated protein model.
A) Red and black rectangles have been used to represent the exon and intron sequences respectively, while the 5' and 3' UTR are represented by double headed arrows. **B)** The amino acid sequences coded for by each exon are represented by numbered red rectangles while blue rectangles above the sequence indicate annotated protein domains; residue overlap splice sites are represented by gaps.

5B_Mutation2 (Cadenza 1004) is present within the first exon of the TRIAE_CS42_5BL_TGACv1_404421_AA1299400 gene model (Figure 6.8 A). 5B_Mutation1 (Cadenza 1458) and 5B_Mutation3 (Cadenza 0179) are both present within the second exon. All three mutations are missense in nature and lead to the substitution of a guanine residue by adenine (Figure 6.8 A). Within the translated protein, 5B_Mutation2 results in an amino acid substitution within the annotated BTB/POZ domain (Figure 6.8 B). 5B_Mutation1 and 3 both result in amino acid substitutions outside any of the annotated protein domains (Figure 6.8 B).

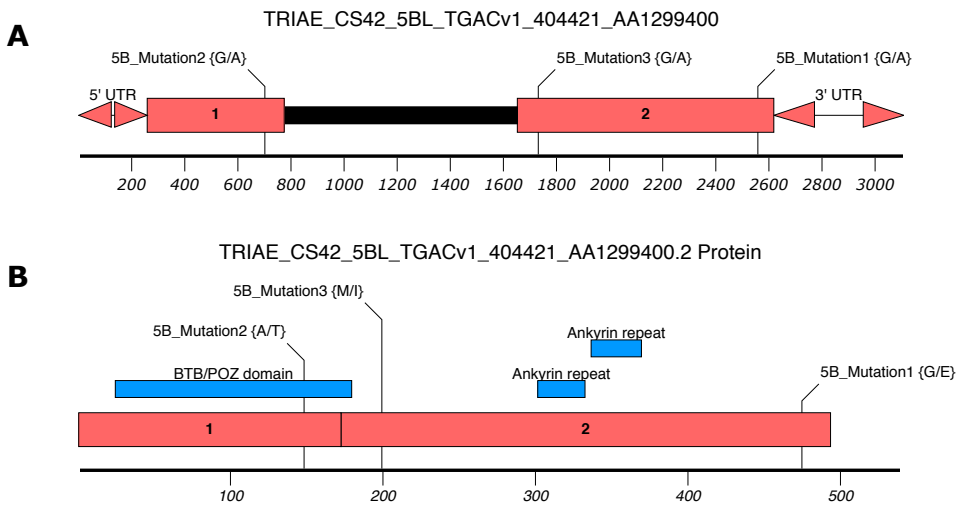

Figure 6.8: Gene model of candidate anther extrusion gene 5BL_TGACv1_404421_AA1299400. **B)** Corresponding translated protein model.
A) Red and black rectangles have been used to represent the exon and intron sequences respectively, while the 5' and 3' UTR are represented by double headed arrows. **B)** The amino acid sequences coded for by each exon are represented by numbered red rectangles while blue rectangles above the sequence indicate annotated protein domains; residue overlap splice sites are represented by gaps.

6.3.3 Confirmation of mutations in TILLING lines

Together with wild-type *Triticum aestivum* cv. Cadenza and Chinese-Spring, biological replicates of each mutant line were genotyped using the KASP platform using genome specific primers targeting each of the mutations described in Table 6.3. Within each KASP fluorescence plot, provided by KWS UK Ltd., there a clear separation of individuals into two distinct clusters, as opposed to three which would be indicative of lines heterozygous for the target mutations.

In each instance, the two wild-type samples were present within the larger of the two clusters; indicating that individuals within the smaller cluster were homozygous for the target mutation. The smaller clusters were each composed of six individuals; the six biological replicates of each line anticipated to be homozygous for the target mutation.

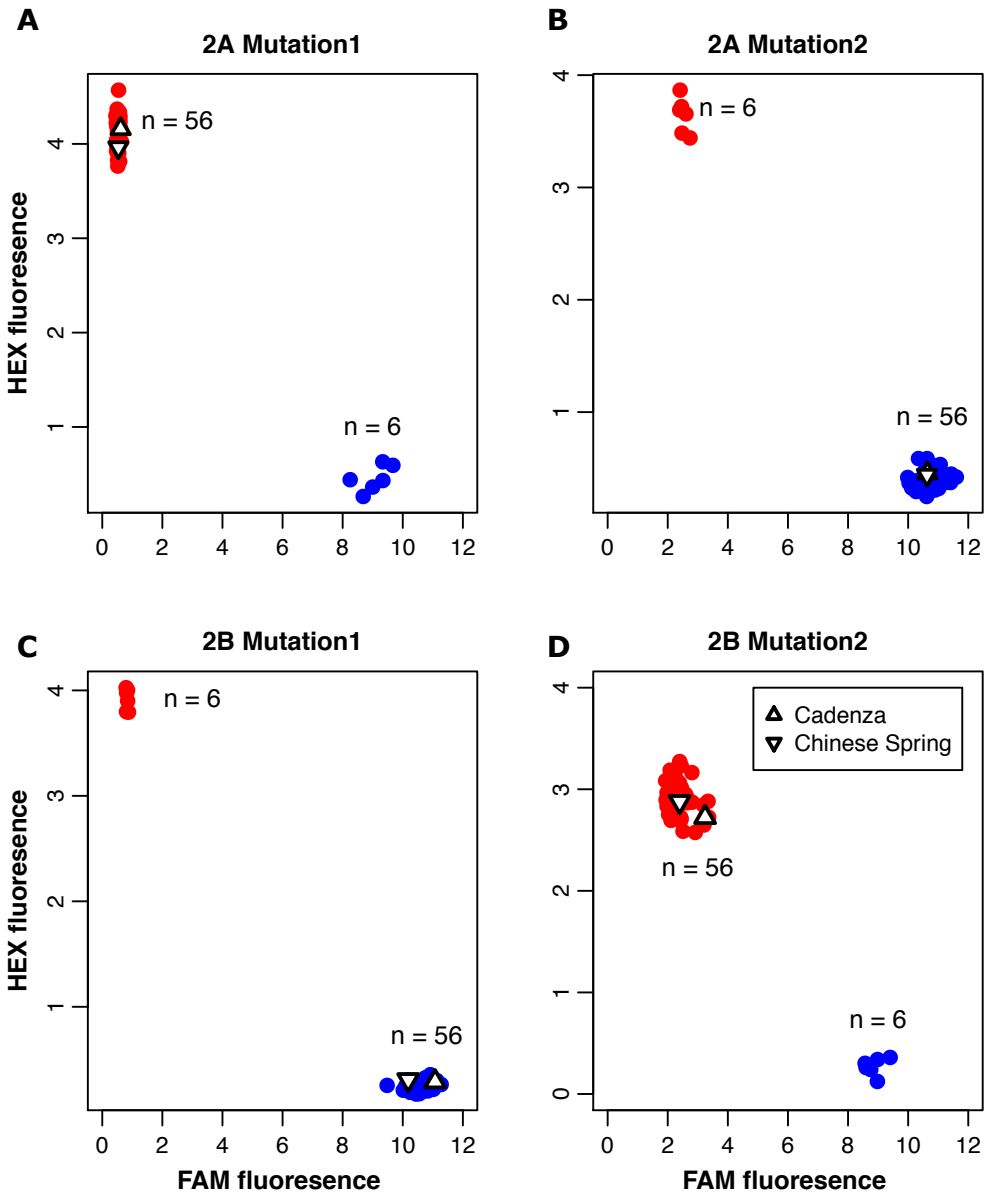

Figure 6.9: Fluorescence plots generated from KASP marker genotyping using primers targeting the following mutations; **A)** 2A_Mutation1 **B)** 2A_Mutation2 **C)** 2B_Mutation1 and **D)** 2B_Mutation2.

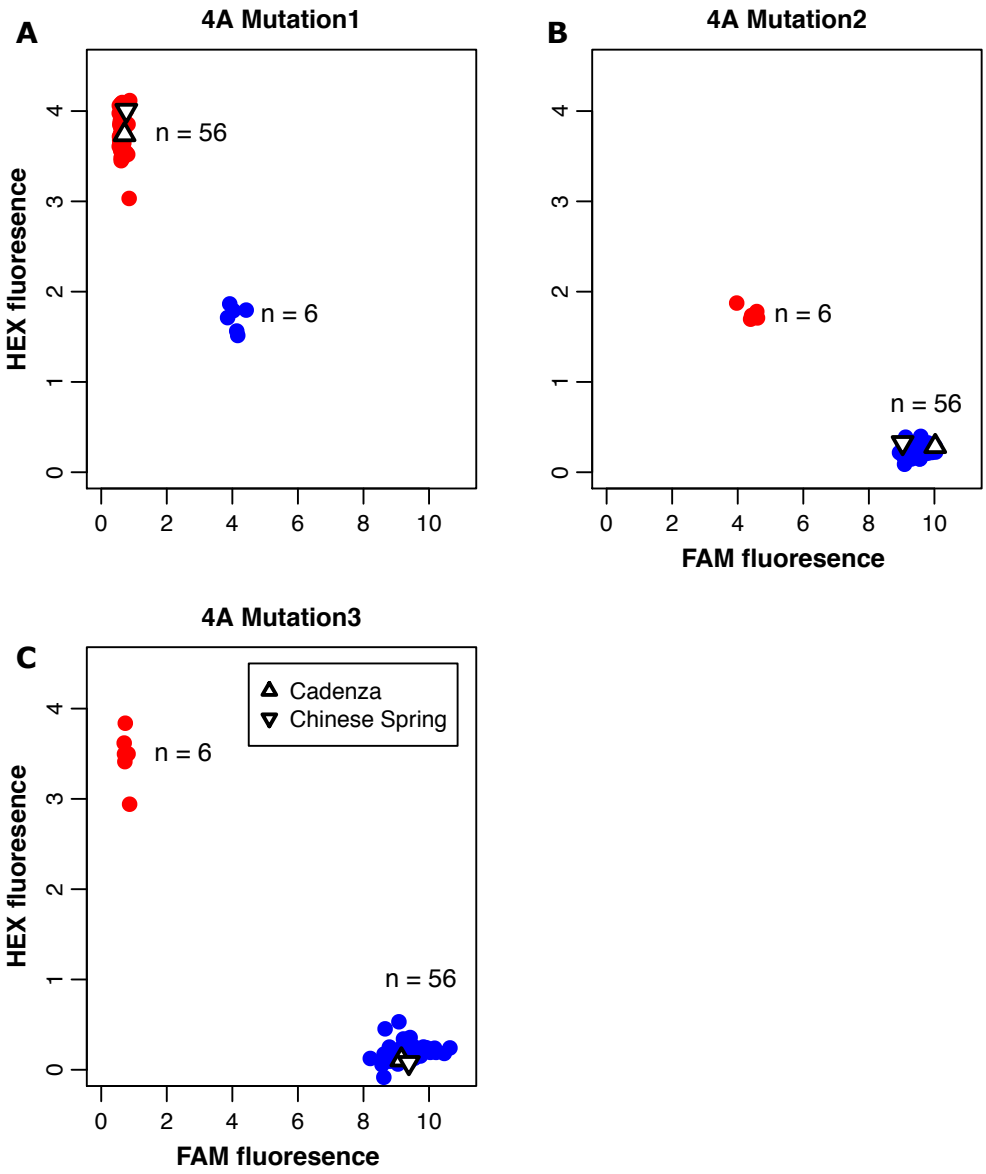

Figure 6.10: Fluorescence plots generated from KASP marker genotyping using primers targeting the following mutations; **A)** 4A_Mutation1 **B)**4A_Mutation2 **C)** 4A_Mutation3.

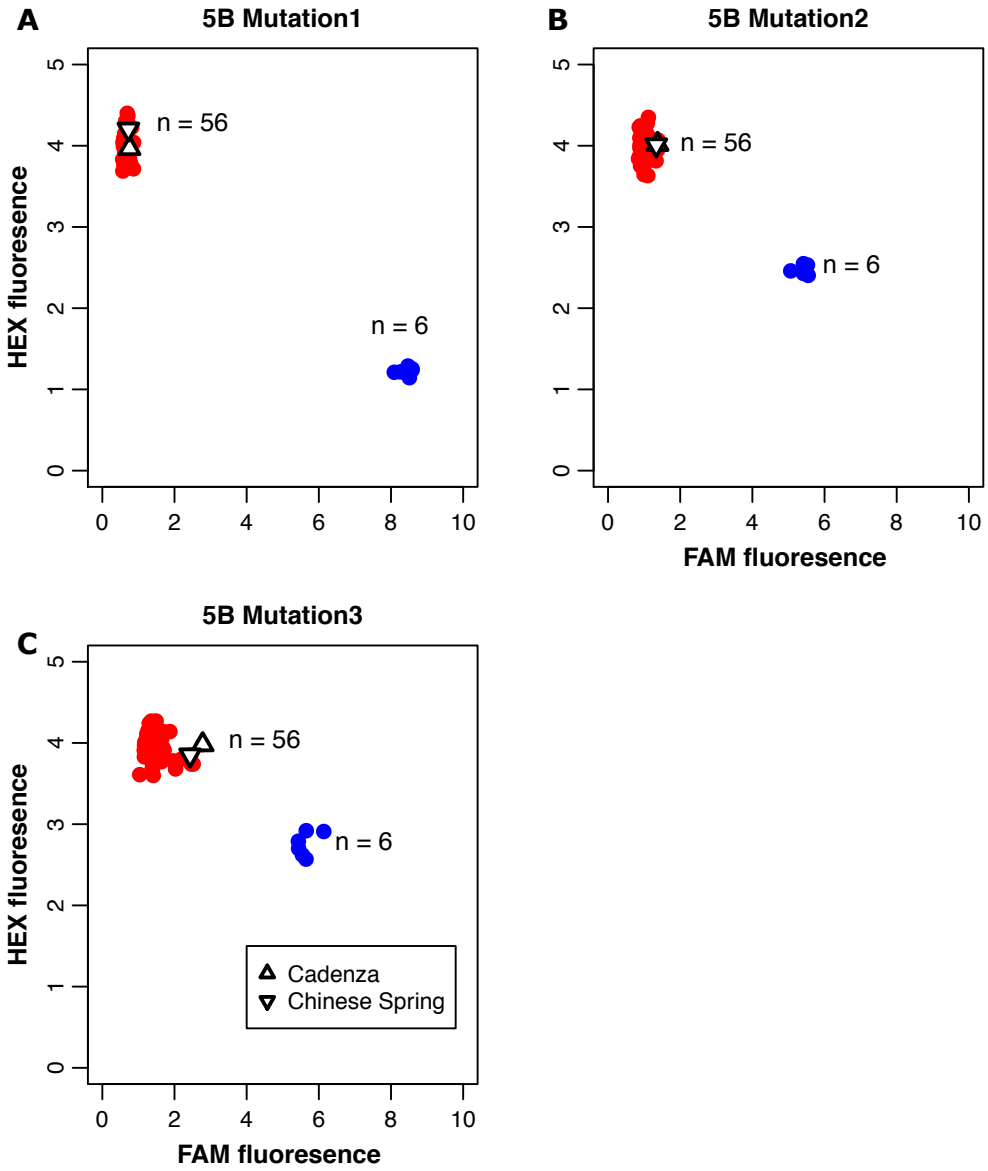

Figure 6.11: Fluorescence plots generated from KASP marker genotyping using primers targeting the following mutations; **A)** 5B_Mutation1 **B)** 5B_Mutation2 **C)** 5B_Mutation3.

6.3.4 Phenotypic results

6.3.4.1 Anther extrusion

Amongst the mutant lines anticipated to exhibit an altered anther extrusion capacity, lines 2B_Mutation2, 5B_Mutation1 and 5B_Mutation2 extruded a significantly higher number of anthers per floret as compared to wild type (Figure 6.12 A; Table 6.4 and 6.6).

The average number of anthers extruded per floret by the 2B_Mutation2 line was 3.5 times higher than that of wild type; 1.38 versus 0.40 respectively (Figure 6.12 A). For 5B_Mutation1 and 5B_Mutation2 differences were of a smaller magnitude; 0.79 and 0.92 anthers extruded per floret respectively.

Relative to wild type, no significant differences were detected with respect to spike length or the number of fully formed spikelets for any of the mutant lines (Figure 6.12 B and C; Table 6.4 and 6.6).

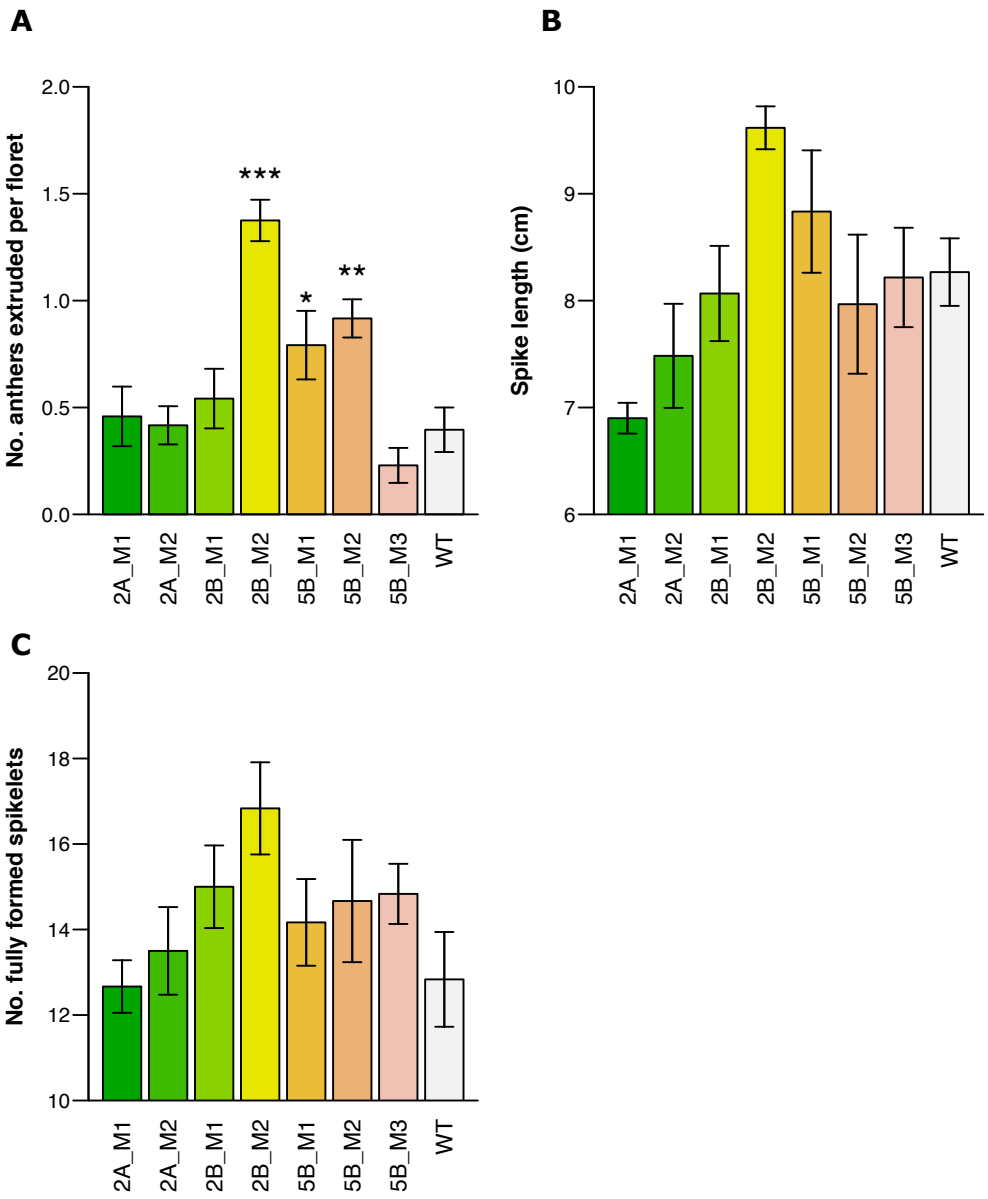

Figure 6.12: Mean phenotypic values of mutant lines and wild type for the following traits; **A)** the number of anthers extruded per floret, **B)** spike length and **C)** the number of fully formed spikelets.

Means and +/- S.E. presented. ***, **, * significantly different from wild type at the 0.001, 0.01 and 0.05 levels of probability.

6.3.4.2 Anther length

All three mutant lines, possessing target mutations in the candidate anther length gene were observed to have significantly shorter anthers relative to wild type (Figure 6.13 A; Table 6.5). The mean anther length of lines 4A_Mutation1, 4A_Mutation2 and 4A_Mutation3 was found to be 2.93, 2.97 and 2.83mm respectively, as compared to 3.18mm for wild type (Figure 6.13 A).

For the 4A_Mutation1 line a reduction in anther length was concurrent with a significant reduction in spike length relative to wild type; 8.2 cm versus 9.5 cm respectively. (Figure 6.13 B). The average number of fully formed spikelets was observed to be significantly greater for the 4A_Mutation2 line as compared to wild type; 17.50 versus 13.17 respectively (Figure 6.13 C). No significant differences were detected in either spike length or the number of fully formed spikelets for of the 4A_Mutation3 line relative to wild type.

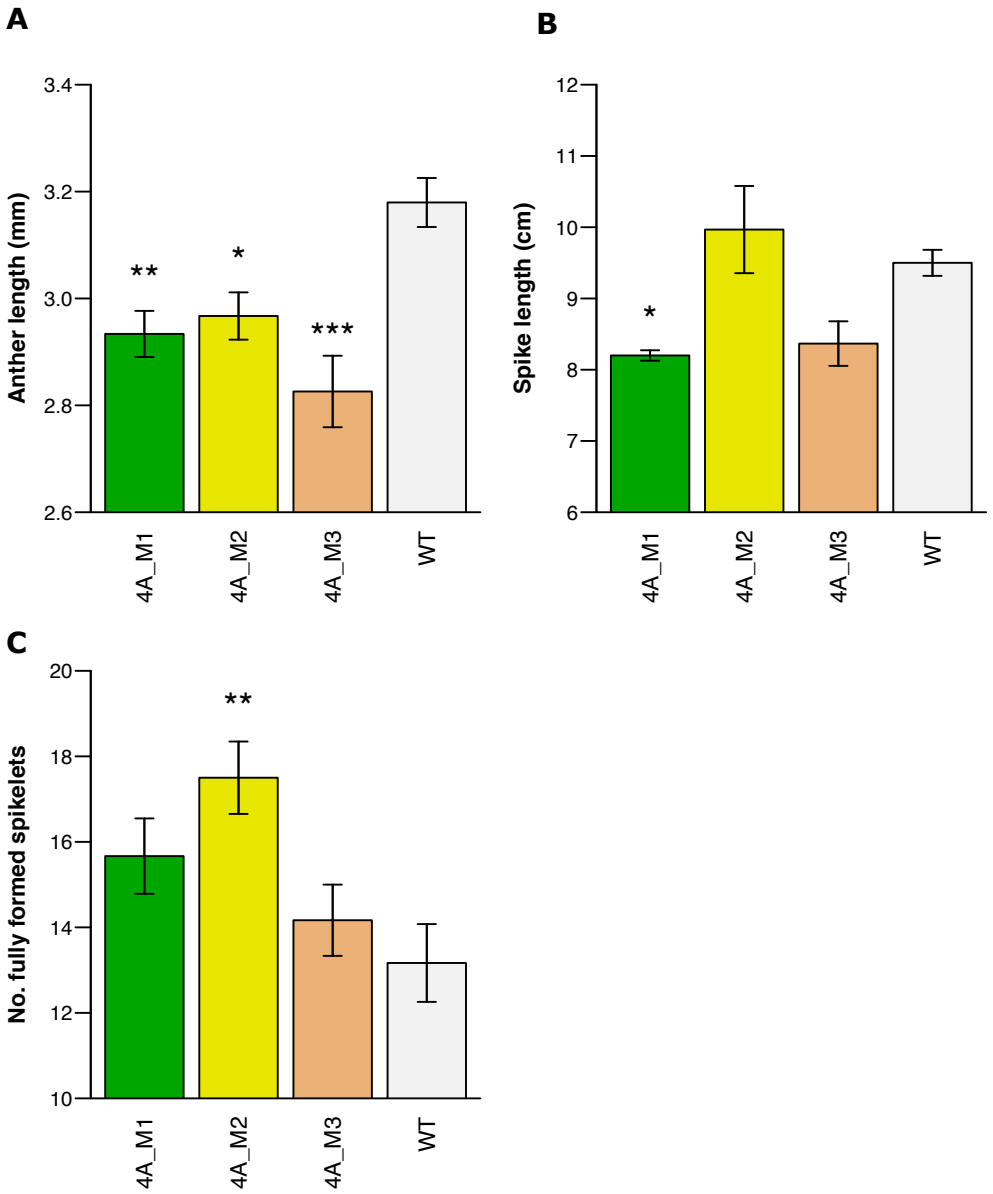

Figure 6.13: Mean phenotypic values of mutant lines and wild type for the following traits; **A)** anther length, **B)** spike length and **C)** the number of fully formed spikelets.

Means and +/- S.E. presented. ***, **, * significantly different from wild type at the 0.001, 0.01 and 0.05 levels of probability.

6.3.4.3 Summary of statistical analysis of phenotypic data

Table 6.4: Results of pairwise t-tests comparing the number of anthers extruded per floret (AE), spike length (SL) and the number of fully formed spikelets (FS) observed for each 2A and 2B mutant line, relative to wild type.
Table shows *P* values from each test with significant differences highlighted in blue.

Trait	2A Mutation1	2A Mutation2	2B Mutation1	2B Mutation2
AE	0.76	0.90	0.48	0.00
SL	0.11	0.40	0.84	0.11
FS	0.91	0.79	0.44	0.12

P values < 0.05 were deemed significantly different from wild type.

Table 6.5: Results of pairwise t-tests comparing anther length (AL), spike length (SL) and the number of fully formed spikelets (FS) observed for each 4A mutant line, relative to wild type.
Table shows *P* values from each test with significant differences highlighted in blue.

Trait	4A Mutation1	4A Mutation2	4A Mutation3
AL	0.01	0.02	0.001
SL	0.04	0.44	0.06
FS	0.11	0.01	0.43

P values < 0.05 were deemed significantly different from wild type.

Table 6.6: Results of pairwise t-tests comparing the number of anthers extruded per floret (AE), spike length (SL) and the number of fully formed spikelets (FS) observed for each 5B mutant line, relative to wild type.
Table shows *P* values from each test with significant differences highlighted in blue.

Trait	5B Mutation1	5B Mutation2	5B Mutation3
AE	0.04	0.01	0.42
SL	0.49	0.77	0.94
FS	0.59	0.45	0.44

P values < 0.05 were deemed significantly different from wild type.

5.3 Discussion

5.3.1 Expression profiles of candidate genes

All four candidate genes selected for investigation in the present chapter were previously known to be expressed within the spike (Chapter 4, Table 4.7 and 4.8; Chapter 5, Table 5.3 and 5.4). The expression profile of these candidates within the spike throughout the development of Chinese Spring adds further support for their putative role in influencing spike characteristics (Figure 6.1, 6.2, 6.3 and 6.4).

5.3.1.1 *TaAP2* homoeoloci

The genes TRIAE_CS42_2AL_TGACv1_097448_AA0324210 and TRIAE_CS42_2BL_TGACv1_129463_AA0384870, which correspond to *TaAP2-A* and *TaAP2-B* respectively, exhibited expression patterns in Chinese Spring that mirror those reported for the variety Shinchunaga (Ning et al., 2013b); with transcript abundance being highest within the spike at/around anthesis. While comparable levels of expression were observed between the two homoeoloci in Shinchunaga, in Chinese Spring transcript abundance of *TaAP2-B* was relatively lower throughout development (Figure 6.1 and 6.2). Although not at comparable levels, to that observed during anthesis in the spike, both homoeoloci were expressed in all

other tissue types (Figure 6.1 and 6.2). This is similar to what is observed with *A. thaliana AP2*, which is expressed within the siliques, leaves and root and shoot apex but at enhanced levels in the inflorescence meristem (Liu et al., 2012). In addition to floral development, the *AP2* family of genes is known to govern a number of developmental processes in *O. sativa* for example root formation and somatic embryogenesis (Rashid et al., 2012).

5.3.1.2 *BOP1/2* putative ortholog

The expression profile of the putative ortholog of *BOP1* and *BOP2* TRIAE_CS42_5BL_TGACv1_404421_AA1299400, adds further support for this gene as a spike architecture candidate. Expression occurred predominantly within the spike during its early development (Figure 6.3); with no other tissue types showing comparable levels of expression. Similarly in *A. thaliana*, BOP1/2 exhibits expression predominantly within the flower (Liu et al., 2012) and is known to be confined to young floral buds (Ha et al., 2004).

5.3.1.3 *GN* putative ortholog

While expression of the *GN* putative ortholog, TRIAE_CS42_4AL_TGACv1_290275_AA0983490 was not confined to the spike tissue (Figure 6.4), the downregulation of gene expression that accompanied the transition from the

flag leaf stage to anthesis adds support for the role of the candidate in early spike development.

No comparable levels of expression, to that observed in the developing spike, was seen in any other tissue type. While *A. thaliana GN*, exhibits expression throughout the plant (Liu et al., 2012), expression is confined to actively dividing cells (Geldner et al., 2003) which may help to explain why the *T. aestivum* candidate ortholog shows greater transcript abundance in the young spike.

5.3.2 Mutations in candidate anther extrusion genes

5.3.2.1 *TaAP2* homoeoloci

The target mutations 2A_Mutation2 and 2B_Muation2 are both non-synonymous variants, assigned SIFT scores of 0.1 and 0.01 respectively (Table 6.3). The differences in SIFT scores appear to be due to the localization of these mutations within *TaAP2-A* and *TaAP2-B* (Figure 6.5 and 6.6). 2B_Mutation2 leads to an amino acid substitution within the first AP2/ERF domain of the TaAP2-B protein (Figure 6.6 B); the change associated with 2A_Mutation2 occurs outside of the AP2/ERF domains of the TaAP2-A protein (Figure 6.5B).

In *A. thaliana*, *AP2* suppresses the expression of floral homeotic genes, such as *AGMAOUS*, and *ethylene-responsive*

factor (*ERF*) genes (Ogawa et al., 2007). It is the AP2/ERF domain that facilitates the transcriptional regulation of these genes via DNA recognition enabled by a β–sheet structure (Riechmann et al., 2000). The localization of 2B_Mutation2 within an AP2/ERF domain would therefore be anticipated to compromise the function of *TaAP2-B* and provides an explanation for the altered anther extrusion capacity observed for the 2B_Mutation2 line (Figure 6.12 A).

The target mutation, 2B_Mutation2 was not present within or proximal to the miR172 targeting site described by Ning et al. (2013b) localizing within the fourth exon as opposed to the tenth exon (Figure 6.6 A). Hence, it is unlikely that the altered phenotype observed in this line is attributable to changes in the capacity of miR172 to post-transcriptionally regulate *TaAP2-B*.

In *H. vulgare*, Nair et al. (2010) proposed that the cleavage of *AP2* transcripts by mi172 during lodicule formation reduces transcript abundance to levels conducive to proper lodicule development. This in turn facilitates the opening of florets at anthesis and ultimately the extrusion of anthers. From this, a model can be proposed to explain the altered phenotype of 2B_Muation2. Rather than a timed down-regulation of *TaAP2* translation by MiR172, 2B_Muation2 is predicted to result in a

reduction of functional TaAP2 proteins throughout the development of the mutant line.

In addition to the regulation of floral homeotic genes in *A. thaliana*, *AP2* has also been demonstrated to play a role in the establishment of floral meristem identity and appears to be involved in the regulation of non-floral developmental processes (Jofuku et al., 1994, Shannon and Meeks-Wagner, 1993). It would therefore be expected that a number of developmental processes apart from lodicule development, would have been affected. Rather, the line was observed to exhibit a normal phenotype with reference to overall growth and development.

Due to the functional redundancy observed between homoeologus genes in *T. aestivum* (Slade et al., 2005), having a functional copy of *TaAP2-A* in the 2B_Mutation2 may be sufficient for the normal progression of other developmental processes such as the establishment of floral meristem identify. The overall reduction in TaAP2 protein in the 2B_Mutation2 line, relative to wild type, may have promoted the proper development of the lodicules throughout the spike, in turn leading to enhanced anther extrusion capacity.

The absence of an altered anther extrusion phenotype in the 2B_Mutation1 and 2A_Mutation1 lines suggest that the

associated splice site variants within *TaAP2-A* and *TaAP2-B* have a minimal effect on the translation efficiency of their corresponding mRNAs.

An apparent increase in spike length relative to wild type, was also observed for the 2B_Muation2 line. While this difference was not found to be significantly different (Table 6.4), no overlap was observed between the S.E ranges of 2B_Mutation2 and wild type (Figure 6.12 B). In *H. vulgare*, *HvAP2* in addition to governing the development of lodicules (Nair et al., 2010), has been reported to influence inflorescence architecture (Houston et al., 2013). Houston et al. (2013) reported that lines possessing the cleistogamous alleles of *HvAP2*, *cly1.b* and *cly1.c*, were characterised by the development of shorter rachis. This may provide an explanation for the increase in spike length observed for 2B_Mutation2, having a lower amount of TaAP2 translated would be anticipated to have the opposite effect of that observed in these cleistogamous barley varieties; the development of longer rachis/spikes.

5.3.2.2 *BOP1/2* putative ortholog

The gene described in the present study as a putative ortholog of *A. thaliana BOP1* and *BOP2,* has to date not been characterised in *Triticum aestivum* but shows a 95.2% alignment with a *H. vulgare* putative ortholog of *BOP1*/*2*

(Protein ID="BAJ90297.1") which has been described as the causal gene within the *lax-a* mutant locus (Jost et al., 2016). The expression profile of this candidate in *T. aestivum* (Figure 6.3) adds support for a role in spike development; with a down-regulation in expression coinciding with the transition into reproductive development.

A SIFT score less than 0.05 was assigned to all three target mutations of this candidate investigated (Table 6.3). The line possessing 5B_Mutation3, which showed the highest SIFT score, showed no alteration in anther extrusion capacity (Figure 6.12A). A significant increase in the number of anthers extruded per floret was observed for lines 5B_Mutation1 and 5B_Mutation2 (Figure 6.12A). Only 5B_Mutation2 (SIFT score 0.00) led to an amino acid change within an annotated protein domain, the BTB/POZ domain, the line possessing this mutation showed a marginal increase in anther extrusion capacity relative to the 5B_Mutation1 line (Figure 6.12 A).

BOP1 and BOP2 contain BTB/POZ and ankyrin repeat domains which are known to be involved in protein-protein interactions (Aravind and Koonin, 1999, Cao et al., 1997). BOP proteins have been found to bind to transcription factors that regulate inflorescence development in a number of species; including the A. *thaliana* TGA factor *PAN* (Hepworth et al., 2005);

MADS-box *A. thaliana* factors *AP1* and *AGL24* (Xu et al., 2010); *Solanum lycopersicum* factor *TERMINATING FLOWER* (*TMF*) (Xu et al., 2016) and *Oryza sativa* MADS-box factors *RAP1A* and *OsMADS1* (Kyozuka et al., 2000).

RAP1A an ortholog of *A. thaliana AP1*, is expressed early in floral development of *O. sativa*, subsequent to which expression is confined to the developing palea, lemma and lodicules (Kyozuka et al., 2000). Interactions between the *H. vulgare BOP1/2* ortholog and a gene orthologous to *O. sativa RAP1A*, may explain the conversion of lodicules into stamenoid structures and the development of a narrower palea and lemma in the *lax-a* mutant.

A single H. *vulgare* and two homoeologus *T. aestivum* genes show an orthologous relationship to *RAP1A* within the EnsemblPlants database (http://plants.ensembl.org/). The gene TRIAE_CS42_2AL_TGACv1_092934_AA0267480 and its homoeolog TRIAE_CS42_2BL_TGACv1_130188_AA0405830 are both expressed within the developing spike (expression data accessed via the http://www.wheat-expression.com database (Borrill et al., 2016)). A reduction in transcript abundance of each gene coincides with the transition from vegetative to reproductive development, as observed for the *T. aestivum BOP1/2* putative ortholog described in the present chapter. These two homoeologs warrant further investigation

towards elucidating the functional role of the proposed *T. aestivum BOP1/2* ortholog and improving our understanding of the determinants of inflorescence architecture in *T. aestivum*.

A relaxed inflorescence phenotype, as reported for the *lax-a H*. *vulgare* mutant*,* was not observed in any of the mutants investigated; including the 5B_Mutation2 line in which a mutation is present within the BTB/POZ domain of the translated protein. Significant differences relative to wild type were not detected for either the number of fully formed spikelets or spike length (Figure 6.12 B and C and Table 6.6).

The increased anther extrusion capacity, relative to wild type, observed for 5B_Mutation1 and 5B_Mutation2 (Figure 6.12A) may be attributable to the development of a narrower palea and lemma, which leads to the characteristic gaping floret appearance of *lax-a H. vulgare*. This trait was not assessed within the present study however and should be revisited in future work.

5.3.3 Mutations in candidate anther length gene the *T. aestivum GN* putative ortholog

Unlike the other two candidate genes investigated in the present study, no prior investigations have elucidated the functional roles that *GN* orthologs play in either *T. aestivum*

or *H. vulgare*. In *A. thaliana*, *GN* is involved in the activation of auxin response factors (ARFs) and the basipetal transport of auxin (Mayer et al., 1993, Steinmann et al., 1999, Shevell et al., 1994).

ARFs activate primary or early auxin-responsive genes, by binding to auxin response elements within their promoters (Guilfoyle et al., 1998), which are known to play roles in the development of the *H. vulgare* anther (Oshino et al., 2011, Sakata et al., 2010). The activation of ARFs results from guanine nucleotide exchange facilitated by the Sec7 domain of *GN* (Shevell et al., 1994).

The putative *T. aestivum GN* ortholog investigated in the present study as an anther length candidate codes for a protein with an annotated Sec7 domain (Figure 6.7 B). All three target mutations result in the formation of premature stop codons within the second exon of the gene (Figure 6.7 A). A Sec7 domain is predicted to be absent within the truncated protein associated with 4A_Mutation3 (Figure 6.7 B), the line possessing this mutation exhibited the greatest reduction in anther length relative to wild type (Figure 6.13 A). While the truncation position associated with 4A_Mutation1 overlaps the Sec7 domain (Figure 6.7B), a comparable reduction in anther length was not observed in its corresponding line (Figure 6.13A). The truncated protein

resulting from 4A_Mutation2 is predicted to contain the entire Sec7 domain (Figure 6.7 B), which may explain the relatively longer anthers of this line as compared to the other two mutants (Figure 6.13 A).

Assessments of spike length in the three mutant line suggest that the *GN* candidate plays an additional role in the determination of spike length. The 4A_Mutation1 and 4A_Mutation3 lines showed a notable reduction in spike length, the difference relative to wild type was marginally below the significance threshold for 4A_Mutation3 (P= 0.06) (Figure 6.13B and Table 6.5). The 4A_Mutation2 showed comparable spike lengths to that of wild type (Figure 6.13B). This mirroring of the pattern observed for anther length, suggests that it is unlikely that reductions in spike length are attributable to non-target background mutations. As to whether these changes are pleiotropic or shorter anthers result from the development of a smaller inflorescence requires more detailed phenotypic characterisation of the mutant lines.

Among the four candidates, the mutations described within the putative *GN* ortholog are least likely to be of direct relevance to breeding programmes as the corresponding mutant lines all show reductions, as opposed to increases, in anther length. There is the possibility that variants of this *GN*

ortholog with compromised protein function already exist within breeding pools and impose a constraint on anther length. An interesting area for continuation of this research would be to identify such variants in existing germplasm; this effort can be aided by the publishing of draft genome sequences for additional varieties. Subsequently, expression analyses can be carried out with the aim of determining the impact of mutations on gene expression and the levels of functional redundancy between homoeologs.

5.3.4 Applicability of allelic variation associated with overexpression of candidate loci towards trait improvement

The TILLING mutants used in the present study possess mutations in the candidate floral trait loci that are anticipated to lead to complete loss of gene function or reduced gene expression. However, past work in barley has demonstrated that overexpression of anther development genes can result in large phenotypic effects.

Fernandez et al. (2014) reported that transgenic lines overexpressing the *H. vulgare* transcription factor *MS1* exhibited male sterility; the same was observed for lines in which the gene was silenced. Similarly, Murray et al. (2003)

reported that increased expression of the anther development gene *HvGAMYB* was associated with a progressive reduction in anther length among the transgenic lines they generated. These results emphasize that there may be an upper threshold of gene expression, above which function can be compromised.

For the candidate anther length gene described in the present study, the loss of function TILLING mutants all produced shorter anthers than wild type. It would therefore be logical to assume that overexpression of the gene would result in the development of longer anthers. Further work is however required to ascertain the level of gene expression at which the desirable phenotype can be obtained without compromising gene function.

In the case of the *AP2* and *BOP1/2* homologs, from theory reduced gene expression would be expected to produce the desired phenotype; higher anther extrusion. This was in fact observed for the TILLING lines investigated; wherein mutations anticipated to lead to reduced gene expression showed a higher anther extrusion capacity. As both *AP2* and *BOP1/2* are known to play key roles in the development of floral meristem, there may be a lower threshold of expression required for the other components of the wheat ear; this therefore warrants investigation.

Further insight into these areas can be obtained by conducting targeted gene expression studies using KWS Tempo, KWS Ferrum and lines from the AM panel showing contrasting floral phenotypes. For each of the candidate floral trait loci, gene models are available on the Ensembl plants database (http://plants.ensembl.org/) that would facilitate the development of primers for use reverse transcription PCR (RT-PCR) and quantitative reverse transcription PCR (qRT-PCR) as done by Ning et al. (Ning et al., 2013b). This would not only provide evidence that the candidate loci identified are causal for the observed phenotypes but will also provide some insight into levels of gene expression in existing germplasm that results in tangible phenotypic effects without compromised ear development.

5.3 Summary

Further work is required to conclusively demonstrate that the modifications to anther extrusion capacity and anther length observed in mutant lines are in fact related to the target mutations as opposed to background mutations. This would entail carrying out at least four backcrossing cycles (Slade et al., 2005) to generate near isogenic lines that carry each of the target mutations, while simultaneously tracking the inheritance of anther extrusion in each cycle. Alternatively, bulk segregant analysis can be carried out using populations derived from a single cross between each mutant line and wild type Cadenza.

For each candidate gene, we have however observed phenotypic changes in at least one of their corresponding mutant lines. In the case of the 2B and 5B anther extrusion candidates, near-isogenic lines generated from mutant lines with enhanced anther extrusion capacity are likely to be of direct value to improving the trait in breeding programmes.

Further, our understanding of the functional roles these genes play and the levels of functional redundancy between genomes can be improved through the generation of double or triple mutants and expression analyses. This would be pre-requisite to the effective exploitation of these resources.

CHAPTER 7: GENERAL DISCUSSION

7.1 Introduction

An improved understanding of floral biology is anticipated to be of benefit towards improving the efficiency of *T. aestivum* hybrid breeding programmes and may also play a role in the development of more stress tolerant varieties. However, due to difficulties in phenotyping for floral traits coupled to reduced interest in hybrid breeding by the commercial sector, optimization of floral characteristics remains an underutilized avenue for yield improvement (Langer et al., 2014a).

In recent years there has been an increase in the amount of genetic resources available to wheat researchers. The International Wheat Genome Sequencing Consortium has recently published a number of draft assemblies for the genome of *T. aestivum* cv. Chinese Spring (Bierman and Botha, 2017). Coupled to this there has been a large improvement in the accessibility of single nucleotide polymorphism (SNP) genotyping platforms in wheat; with relatively cheap SNP arrays being available, such as the Infinium iSelect 90,000 SNP (Wang et al., 2014) array used in the present study.

These improvements in genetic resources coupled with the renewed interest in developing phenotyping approaches for floral traits in wheat (Langer et al., 2014a), now provide a

research environment which is conducive to dissecting the physiological and genetic factors that govern floral characteristics and investigating their role as determinants of grain yield.

7.1.1 Applying floral ideotypes to hybrid and stress tolerance breeding

Despite *T. aestivum* being characterized as an obligate inbreeder (Lukac et al., 2012, Langer et al., 2014a), within existing germplasm a number of characteristics anticipated for out-crossers are observed; including but not limited to, a high extent of anther extrusion (Beri and Anand, 1971) and the production of relatively large quantities of pollen (Beri and Anand, 1971).

In past attempts at hybrid breeding in *T. aestivum*, lines possessing these characteristics have been used as male parental lines. However, breeders were unable to develop cost-effective systems as the ratio of male to female parental lines required to achieve an adequate spread and quantity of pollen was still found to be too high (Langer et al., 2014a).

This demonstrates that the development of efficient hybrid wheat breeding requires a targeted and systematic approach that aims to incorporate all traits expected to influence out-crossing potential within male parental pools. A male ideotype would shed high quantities of pollen external to the florets,

which is released over a long period of time so as to cover the entire period of female receptivity. The present study therefore targeted traits anticipated to contribute towards achieving these goals.

Longer anthers, established as a proxy for high pollen production (Beri and Anand, 1971), and a high capacity for anther extrusion, which ensures that pollen is released into the environment (Beri and Anand, 1971) are key pre-requisites for outcrossing. However, an improved outcrossing potential would not translate into more efficient seed production in instances where the flowering of male parental lines and female parental lines are not synchronized. Thus additional selection criteria are required, specifically the timing of anthesis initiation, overall anthesis duration as well as the proportion of ears at anthesis on a daily basis.

Recent investigations into the floral characteristics of wheat (Boeven et al., 2016, Muqaddasi et al., 2017, Muqaddasi et al., 2016, Langer et al., 2014a) have predominantly discussed the relevance of findings towards hybrid wheat breeding. It should be emphasized that germplasm, genomic and phenotyping resources of relevance to hybrid breeding can also be of value in the development of lines with improved stress tolerance. For instance, in wheat severe reductions in pollen fertility, and ultimately seed set, are known to occur

under conditions of heat stress (Saini et al., 1984), water deficit (Dorion et al., 1996) and suboptimal irradiance (Demotes-Mainard et al., 1995).

In the literature there already exists some insight as to what traits can be targeted. Work in rice has demonstrated that longer anthers can help to mitigate against abiotic stress induced pollen sterility (Saito et al., 2001, Suzuki, 1981). Higher pollen production has also been demonstrated to provide consistent yields in commercial rye production across seasons, irrespective of climatic conditions (KWS, 2017). Further, Lukac et al. (2012) and Demotes-Mainard et al. (1995) proposed that pollen shedding external to florets can help to mitigate against pollen sterility in earlier flowering florets via intra-ear fertilization. There is however a gap in the understanding of relationships between additional floral traits and abiotic stress tolerance in wheat. Investigations into this area can help to identify additional target traits that can be used to develop ideotypes appropriate for specific environments.

The work carried out in the present study confirms that genotypic variation exists in wheat for two phenotypic characteristics that are anticipated to be of value to improving abiotic stress tolerance; the quantity of pollen produced and the synchronicity of floret development/flowering within

individual plots (AUFPC). A third key trait, anther dehiscence, was not directly assessed within the germplasm used in the present study, pollen trap assessments however showed that a comparable anther extrusion capacity did not necessarily translate into a similar number of pollen grains being shed external to the wheat ear. This observation suggests that there may in fact be variation for the extent of anther dehiscence in wheat germplasm and warrants further investigation. For each of these traits, their value to improving hybrid seed production has been discussed extensively; genomic resources developed such as genetic markers and mutant germplasm, regardless of their intended use, will be applicable to stress tolerance and hybrid breeding in wheat.

Amongst the traits described, anther extrusion has by far received the most attention by the scientific community in the recent past (Langer et al., 2014a, Boeven et al., 2016, Lu et al., 2013, Skinnes et al., 2010, Muqaddasi et al., 2017, Muqaddasi et al., 2016); possibly due to ease of phenotyping. Also, anther extrusion is of interest to pathologists due to its relationship with *Fusarium* infestation.

Anther extrusion is known to be a function of two biological processes; filament extension and floret opening via expansion of the lodicules (De Vries, 1971, Peterson, 1965).

In grass species, an influx of potassium ions into filaments and lodicules initiates these processes. In wheat it appears that this influx of potassium into the lodicules and filaments is controlled by two independent mechanisms (Heslop-Harrison and Heslop-Harrison, 1996). Additionally, work in barley has shown that jasmonic acid and its analogues have an antagonistic relationship with anther extrusion (Honda et al., 2006), while auxins appear to promote floret opening (Honda et al., 2005).

Together, these findings suggest that anther extrusion in wheat is influenced by a number of genes, possibly acting independently of one another, and may help to explain the large number of anther extrusion QTLs reported in recent years (Langer et al., 2014a, Boeven et al., 2016, Lu et al., 2013, Skinnes et al., 2010, Muqaddasi et al., 2017, Muqaddasi et al., 2016). Amongst these candidate loci, some are likely to have more pronounced phenotypic effects, making them ideal targets for trait improvement.

Anther extrusion is indeed a primary determinant of the proportion of pollen grains shed external to the ear (Beri and Anand, 1971). However, if advances are to be made towards the development of efficient male ideotypes, comparable advances must be made for other important floral traits. This

requires the knowledge and capacity to phenotype for these traits and identify the genetic factors that govern them.

7.2 The feasibility of breeding for floral ideotypes in *Triticum aestivum*

7.2.1 Capacity to phenotype floral traits

A number of floral phenotyping strategies were successfully applied in the characterization of European *T. aestivum* germplasm (Chapter 2 and 3), hybrid male parental lines (Chapter 2) and mutant lines from a TILLING population (Chapter 6). For each trait, the strategy employed enabled the detection of significant genotypic variation for traits of interest. This was achieved with an adequate level of accuracy, as shown by the strong correlations of phenotypic values between trials, in both glasshouse and field conditions, as well as the detection of modest to high trait heritability.

Methods describing assessment of anthesis duration, anthesis pattern (area under the flowering progress curve), anther extrusion, anther length and the number of pollen grains per anther in Chapter 2, were all low-tech in nature and did not require a significant investment towards their development. Further, while experience with these methods improves the speed and accuracy of phenotyping; it is anticipated that

these methods can be successfully applied with little prior training.

The key difference between the phenotyping methods described in the present study was with respect to how time demanding and labour intensive they are. These factors are the key determinants of how feasible it would be to apply a given method towards the screening of a large number of lines, whether within breeding programmes or QTL studies. Approaches applied towards characterising anthesis duration, anthesis profile (area under the flowering progress curve) and the number of pollen grains per anther were constrained by being both labour intensive and time demanding. These limitations were further exacerbated by the narrow time window in which they needed to be applied, limiting the number of traits that can be simultaneously screened for, as well as the number of lines that can be assessed.

There certainly exists a need for the development of higher throughput approaches for assessing these traits. This would be of direct benefit to plant breeders and to researchers aiming to dissect the genetic factors that influence variation in these traits. The advent of a wide array of crop imaging platforms (Li et al., 2014) and even pollen sampling technologies (Heidmann et al., 2016) in recent years suggest that these goals will be realized in the near future.

The adoption of newly developed methods will be largely determined by the extent of investment required into relevant technologies/apparatus. It is likely that low-tech approaches to phenotyping floral traits that are labour/time intensive will still be of practical use in breeding floral ideotypes, provided they are applied at an appropriate stage within the breeding cycle.

With reference to developing male parental lines for use in hybrid breeding programmes, traits that are less labour/time-intensive to assess, can be targeted in the initial selection stage. Anther extrusion capacity, which is known to be pre-requisite for outcrossing, can be assessed at this stage using the phenotyping method described in the present study. Promising lines can be taken forward into a second round of selection and assessed for anther length and the extent of pollen shed external to the spike as done in the present study. While the pollen trap method was time demanding, counting the number of pollen grains adhered to the surface of traps can be carried out after flowering has ceased and has the added benefit of being able to provide insight into anthesis pattern and duration. Pollen traps can be applied together with the pollen mass method described by Langer et al. (2014a).

In the final selection steps, a more detailed characterization of anthesis patterns (AUFPC) and duration will be required in order to ensure that male parental lines with an anthesis profile capable of covering the entire period of female receptivity. This can be achieved using the approaches described in the present study, which were found to be the most time and labour intensive methods (Chapter 2). At this point within the breeding cycle, molecular markers tagging causative loci will be of the most relevance if marker-assisted selection (MAS) proves to be more cost-effective and accurate than phenotypic selection.

7.2.2 Extent of variation for floral traits in adapted germplasm

That there exists variation for floral characteristics in *T. aestivum* is not a recent finding; a number of studies over the past 50 years have reported genotypic variation for pollen production, anthesis pattern, anther extrusion and pollen shedding external to the wheat ear (Cahn, 1925, Heyne and Livers, 1967, Kherde et al., 1967, Wilson, 1968, Johnson and Schmidt, 1968, D'SouzA, 1970, Athwal and Kimber, 1970, Beri and Anand, 1971, De Vries, 1971, De Vries, 1972, De Vries, 1973, Khan et al., 1973, De Vries, 1974, Jôst et al., 1976). It is only within the recent past however that some insight has been gained with respect to the prevalence of

characteristics that promote out-crossing in European adapted germplasm.

Large scale field studies using European material, have reported considerable genotypic variation for traits such as anther extrusion (Muqaddasi et al., 2017, Muqaddasi et al., 2016, Langer et al., 2014a, Boeven et al., 2016), anthesis duration (Langer et al., 2014a), anther length (Langer et al., 2014a, Boeven et al., 2016) and pollen shedding external to the spike (Langer et al., 2014a).

The results of the present study are generally in line with these reports, with a wide distribution of phenotypic values being observed for anther extrusion, anther length, anthesis duration, pollen shedding external to the spike and anthesis profile. Additionally, the linkage mapping study carried out in Chapter 5 revealed anther extrusion promoting alleles in KWS Tempo and KWS Ferrum, two recently released commercial varieties.

These findings add support for the feasibility of developing floral ideotypes for the European market, whether for use as parental lines in hybrid breeding or as an approach towards improving abiotic and biotic stress tolerance. Desirable traits, such as a high capacity for anther extrusion and the production of large quantities of pollen can be incorporated into breeding pools from material that is already adapted to

local environments with little or no linkage drag being anticipated.

As a consequence, sourcing floral characteristics from wild germ-plasm or closely related species may be viewed as a less attractive approach towards trait improvement by breeders. It should however be emphasized that variation for a number of additional traits that can enhance outcrossing potential have yet to be focused on in wheat, such as pollen longevity and pollen aerodynamic properties. Genetic resources that enable exploitation of variation in closely related species, are or are becoming publicly available to the research community and the commercial sector such as rye chromosome addition lines (Driscoll and Sears, 1971) and amphidiploids from hybrids of wheat and other species (Nemeth et al., 2015).

7.2.3 The robustness of floral phenotypes and the feasibility of selection for these traits

While the feasibility of selecting for floral traits is supported by the large amount of phenotypic variation present within germplasm, the robustness of phenotypes over time would inevitability determine the success of breeding programmes and may place a constraint on the exchange of germplasm between environments.

In the initial investigations of the present study (Chapter 2), a strong correlation between the phenotypic values obtained in each glasshouse trial was observed for the floral traits investigated. Similar results were obtained from the field trials conducted in 2014 and 2015 (Chapter 3). The high trait heritability of anther extrusion, anther length and anthesis duration phenotypes is further supported by the capacity to detect marker trait associations in the present study (Chapter 4), as association mapping detects robust loci capable of exhibiting phenotypic effects across genetic backgrounds and environments.

These observations all support the validity of selecting for floral traits and suggest that phenotypes are consistent across environments. The modest heritability observed for anthesis duration and the area under the flowering progress curve should not be viewed as indicative of these traits being too environmentally sensitive for effective selection to be achieved, as MAS and GS as well as multi-location trials can aid in breeding for these traits.

Line breeding and hybrid breeding schemes in wheat typically require five to six years for the completion of one breeding cycle (Longin et al., 2014). The difficulty of stacking multiple floral traits into individual varieties, some of which appear sensitive to environmental conditions, should not be

underestimated and may in fact require a greater investment in time. This can effectively compromise the capacity of breeders to respond to changing climates and consumer needs. It should be therefore emphasized that avenues for improved breeding efficiency through the use of molecular resources need to be explored if these goals are to be met.

7.3 Genetic resources made available for plant breeders and researchers

7.3.1 Marker trait associations and quantitative trait loci; towards marker assisted selection for floral traits

Association mapping (AM) (Chapter 4) and linkage mapping (LM) (Chapter 5) provide a starting point for the development of marker assisted selection (MAS) programmes targeting floral traits. Genomic regions associated with tangible effects on anther extrusion, anther length, and anthesis duration were detected, each explaining ten percent or more of the phenotypic variation within the respective populations. AM detected common alleles within the association panel that promote desirable floral characteristics across genetic backgrounds. While the results of LM revealed alleles present in current commercial germplasm capable of enhancing

anther extrusion. Using both approaches ideal targets for MAS have been identified; it is anticipated that selection for these identified loci will not be associated with linkage drag.

The identification of reliable markers for use in marker-assisted selection, that are consistently inherited together with the desired phenotype, requires fine mapping of these genomic regions be conducted. In the case of AM this can be achieved through the enrichment of SNPs within each region together with the use of larger association mapping panels with a faster rate of linkage disequilibrium decay (Atwell et al., 2010). While increased LM precision can be achieved through the use of populations with a higher number of recombination events within the described QTLs regions (Röder et al., 2008, Cuthbert et al., 2006); for instance, the F5 population derived from the F4 population utilized in the present study. These approaches can help to provide markers useful in MAS and identify the causative loci within each region.

Marker-trait association and QTL regions found to be associated with anther extrusion here have not been reported in prior studies (Muqaddasi et al., 2017, Muqaddasi et al., 2016, Boeven et al., 2016, Lu et al., 2013, Skinnes et al., 2010). This observation supports the assertion that anther extrusion is a polygenic trait under the influence of a number

of loci and the assumption that a number of independent biological processes influence the expression of the trait (Chapter 7, Section 7.1.1). It should be emphasized that some loci appear to have more pronounced phenotypic effects, making them ideal targets for trait improvement. The decision as to which loci should be targeted in MAS can also be guided by further research into the robustness of loci within individual breeding pools.

7.3.2 Induced mutation lines with altered floral characteristics

Due to the limitations associated with analysis of the TILLING mutants described in Chapter 6, the results do not conclusively demonstrate that targeted genes are in fact the causative loci of the marker-trait associations (Chapter 4) and QTLs (Chapter 5) identified. The development of near-isogenic lines possessing the target mutations, as well conducting fine mapping as described in Chapter 7, Section 7.3.1, could provide the necessary evidence.

Three mutant lines exhibiting increased anther extrusion relative to wild type were characterized; one possessing a missense variant within *TaAP2-B* and two possessing missense variants in a putative *T. aestivum BOP1/2* ortholog. These lines may be of immediate value to breeding programmes aiming to improve anther extrusion assuming it

can be confirmed that observed phenotypes arise due to target mutations as opposed to background mutations. This can be achieved through bulk segregant analysis and/or the development of near-isogenic lines.

The mutants described as exhibiting shorter anthers relative to wild type, would be of little to no practical use to breeding programmes, as longer anthers are desirable in improving stress tolerance and out-crossing potential. Rather these mutants, possessing a premature stop codon within a putative *T. aestivum GN* ortholog, may be of value towards elucidating the genetic factors that govern anther length in wheat. There also exists the possibility that variants of this loci may be present in modern germplasm that impose a constraint on anther length.

In the present study, only phenotypic characterization of TILLING lines was carried out. Our understanding of the impact of mutations on gene expression and the levels of functional redundancy between homoeologs can be improved by carrying out expression analyses for target loci, and their homoeologs, in mutant lines.

For each target mutation in the candidate anther extrusion genes, homoeolog specific primers, used to confirm the mutant status of each line, are available to facilitate their introgression into breeding pools. If the candidate loci are not

responsible for observed phenotype, bulk segregant analysis can be used to identify the causative mutation, as described in Chapter 6.

7.4 Areas of follow up research

The main areas for continuation of the work carried out in the present study are:

1. **Phenotyping**: The development of higher throughput phenotyping approaches towards assessing anthesis duration and anthesis pattern, as well as the number of pollen grains produced per anther.
2. **Fine mapping**: Identifying causative loci, especially for traits that are labour intensive to assess, by fine mapping of genomic regions possessing floral trait marker-trait associations and QTLs.
3. **Candidate gene investigations:** Development of near-isogenic lines, double and triple mutants from the TILLING lines used in the present study, as well as expression analyses of mutant line; which will enable confirmation of candidate gene effects and elucidating the extent of functional redundancy between genomes.
4. **Exploring additional traits**: Investigations into additional floral traits not explored in the present study, including but not limited to: pollen sterility, pollen longevity and pollen aerodynamic properties. Emphasis should be placed on the extent of genotypic variation for these traits and the role each plays as a determinant of out-crossing potential.

7.5 Summary

1. The accuracy and time/labour demands of a number of phenotyping approaches for floral traits was investigated. Methods appropriate for use in breeding programmes and research programmes have been developed and are already being deployed in a commercial breeding programme.
2. A high level of variation for floral characteristics was found within a collection of European germplasm, composed primarily of lines released within the last 30 years. Phenotypes were found to be robust across years and environments, pre-requisites for efficient breeding and selection.
3. Genomic regions associated with phenotypic variation in floral characteristics were identified by both association mapping and linkage mapping. In each instance these regions explained a modest proportion of phenotypic variance within the populations investigated, at similar levels to that reported in the literature.
4. A number of TILLING lines possessing mutations in candidate floral trait genes were characterized; lines exhibited modifications to floral characteristics relative to wild type were identified.

REFERENCES

Akel, W., Thorwarth, P., Mirdita, V., Weissman, E. A., Liu, G., Würschum, T. & Longin, C. F. H. (2018). Can spelt wheat be used as heterotic group for hybrid wheat breeding? Theoretical and Applied Genetics: 1-12.

Akhunov, E. D., Akhunova, A. R., Anderson, O. D., Anderson, J. A., Blake, N., Clegg, M. T., Coleman-Derr, D., Conley, E. J., Crossman, C. C. & Deal, K. R. (2010). Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. BMC Genomics **11**: 1.

Andersen, J. R., Schrag, T., Melchinger, A. E., Zein, I. & Lübberstedt, T. (2005). Validation of *Dwarf8* polymorphisms associated with flowering time in elite European inbred lines of maize (*Zea mays* L.). Theoretical and Applied Genetics **111**: 206-217.

Aravind, L. & Koonin, E. V. (1999). Fold prediction and evolutionary analysis of the POZ domain: structural and evolutionary relationship with the potassium channel tetramerization domain. Journal of Molecular Biology **285**: 1353-1361.

Arends, D., Prins, P., Broman, K. W. & Jansen, R. C. (2014). Tutorial-Multiple-QTL mapping (MQM) analysis for R/qtl.

Athwal, R. & Kimber, G. (1970). Anther size and pollen longevity in wheat/rye addition lines. Wheat Information Service: 30-32.

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., et al. (2010). Genome-wide association study of 107 phenotypes in a common set of *Arabidopsis thaliana* inbred lines. Nature **465**: 627-631.

Bansept, P. (2013). A genome-wide association study of resistance to the yellow rust pathogen (Puccinia striiformis f. sp. tritici) in elite UK wheat germplasm. University of East Anglia.

Barnard, C. (1955). Histogenesis of the inflorescence and flower of *Triticum aestivum* L. Australian Journal of Botany **3**: 1-20.

Bentley, A. R., Scutari, M., Gosman, N., Faure, S., Bedford, F., Howell, P., Cockram, J., Rose, G. A., Barber, T. & Irigoyen, J. (2014). Applying association mapping and genomic selection to the dissection of key traits in elite European wheat. Theoretical and applied genetics **127**: 2619-2633.

Beri, S. & Anand, S. (1971). Factors affecting pollen shedding capacity in wheat. Euphytica **20**: 327-332.

Bierman, A. & Botha, A.-M. (2017). A review of genome sequencing in the largest cereal genome, *Triticum aestivum* L. Agricultural Sciences **8**: 194.

Boeven, P. H., Longin, C. F. H., Leiser, W. L., Kollers, S., Ebmeyer, E. & Würschum, T. (2016). Genetic architecture of male floral traits required for hybrid wheat breeding. Theoretical and Applied Genetics: 2343-2357.

Borrill, P., Ramirez-Gonzalez, R. & Uauy, C. (2016). expVIP: a customisable RNA-seq data analysis and visualisation platform opens up gene expression analysis. Plant Physiology **176**: 1667.

Brachi, B., Morris, G. P. & Borevitz, J. O. (2011). Genome-wide association studies in plants: the missing heritability is in the field. Genome Biology **12**: 1.

Breseghello, F. & Sorrells, M. E. (2006). Mapping of kernel size and milling quality in wheat cultivars. Genetics **172**: 1165-1177.

Brown, A. H. D. (1989). The use of plant genetic resources. Cambridge University Press.

Buerstmayr, M. & Buerstmayr, H. (2015). Comparative mapping of quantitative trait loci for *Fusarium* head blight resistance and anther retention in the winter wheat population Capo× Arina. Theoretical and Applied Genetics **128**: 1519-1530.

Cahn, E. (1925). Study of fertility in certain varieties of common wheat with respect to anther length and amount of pollen in parents and offspring. North Dakota Agricultural College.

Cao, H., Glazebrook, J., Clarke, J. D., Volko, S. & Dong, X. (1997). The *Arabidopsis NPR1* gene that controls systemic

acquired resistance encodes a novel protein containing ankyrin repeats. Cell **88**: 57-63.

Cardarelli, M. & Cecchetti, V. (2014). Auxin polar transport in stamen formation and development: how many actors? Frontiers in Plant Science **5**: 333.

Chao, S., Dubcovsky, J., Dvorak, J., Luo, M.-C., Baenziger, S. P., Matnyazov, R., Clark, D. R., Talbert, L. E., Anderson, J. A. & Dreisigacker, S. (2010). Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). BMC Genomics **11**: 1.

Chao, S., Zhang, W., Dubcovsky, J. & Sorrells, M. (2007). Evaluation of genetic diversity and genome-wide linkage disequilibrium among US Wheat germplasm representing different market classes. Crop Science **47**: 1018-1030.

Chen, A. & Dubcovsky, J. (2012). Wheat TILLING mutants show that the vernalization gene *VRN1* down-regulates the flowering repressor *VRN2* in leaves but is not essential for flowering. PLoS Genetics **8**: 31-34.

Cheng, H., Qin, L., Lee, S., Fu, X., Richards, D. E., Cao, D., Luo, D., Harberd, N. P. & Peng, J. (2004). Gibberellin regulates *Arabidopsis* floral development via suppression of DELLA protein function. Development **131**: 1055.

Civelek, M. & Lusis, A. J. (2014). Systems genetics approaches to understand complex traits. Nature reviews: Genetics **15**: 34-48.

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

Cpm (2017). Hybrid wheat – A new dimension of genetic scrutiny [Online]. Available: http://www.cpm-magazine.co.uk/2017/06/21/hybrid-wheat-new-dimension-genetic-scrutiny/ [Accessed 22/08/2017 2017].

Crossett, A., Lauter, N. & Love, T. M. (2010). An empirical method for establishing positional confidence intervals tailored for composite interval mapping of QTL. PLoS One **5**: e9039.

Cuthbert, P. A., Somers, D. J., Thomas, J., Cloutier, S. & Brulé-Babel, A. (2006). Fine mapping *Fhb1*, a major gene

controlling *Fusarium* head blight resistance in bread wheat (*Triticum aestivum* L.). Theoretical and Applied Genetics **112**: 1465-1472.

D'souza, L. (1970). Studies on the suitability of wheat as a pollen donor for cross pollination, compared with rye, Triticale and Secalotricum. Zeitschrift fur Pflanzenzuchtung **63**: 246-269.

Dafni, A. & Firmage, D. (2000). Pollen viability and longevity: practical, ecological and evolutionary implications. Springer.

Das, S., Krishnan, P., Nayak, M. & Ramakrishnan, B. (2014). High temperature stress effects on pollen of rice (*Oryza sativa* L.) genotypes. Environmental and Experimental Botany **101**: 36-46.

De Vries, A. P. (1971). Flowering biology of wheat, particularly in view of hybrid seed production—a review. Euphytica **20**: 152-170.

De Vries, A. P. (1972). Some aspects of cross-pollination in wheat (*Triticum aestivum* L.) 1. Pollen concentration in the field as influenced by variety, diurnal pattern, weather conditions and level as compared to the height of the pollen donor. Euphytica **21**: 185-203.

De Vries, A. P. (1973). Some aspects of cross-pollination in wheat (*Triticum aestivum* L.) 2. Anther extrusion and ear and plant flowering pattern and duration. Euphytica **22**: 445-456.

De Vries, A. P. (1974). Some aspects of cross-pollination in wheat (*Triticum aestivum* L.). 3. Anther length and number of pollen grains per anther. Euphytica **23**: 11-19.

Demotes-Mainard, S., Doussinault, G. & Meynard, J. (1995). Effects of low radiation and low temperature at meiosis on pollen viability and grain set in wheat. Agronomie **15**: 357-365.

Dodig, D., Zoric, M., Kobiljski, B., Savic, J., Kandic, V., Quarrie, S. & Barnes, J. (2012). Genetic and association mapping study of wheat agronomic traits under contrasting water regimes. International Journal of Molecular Sciences **13**: 6167-6188.

Dorion, S., Lalonde, S. & Saini, H. S. (1996). Induction of male sterility in wheat by meiotic-stage water deficit Is

preceded by a decline in invertase activity and changes in carbohydrate metabolism in anthers. Plant Physiology **111**: 137-145.

Driscoll, C. J. & Sears, E. R. (1971). Individual addition of the chromosomes of 'Imperial' rye to wheat. Agronomy Abstracts.

Fernández Gómez, J. & Wilson, Z. A. (2014). A barley PHD finger transcription factor that confers male sterility by affecting tapetal development. Plant Biotechnology Journal **12**: 765-777.

Fritz, S. E. & Lukaszewski, A. J. (1989). Pollen longevity in wheat, rye and triticale. Plant Breeding **102**: 31-34.

Gardner, K. A., Wittern, L. M. & Mackay, I. J. (2016). A highly recombined, high-density, eight-founder wheat MAGIC map reveals extensive segregation distortion and genomic locations of introgression segments. Plant Biotechnology Journal **14**: 1406-1417.

Geldner, N., Richter, S., Vieten, A., Marquardt, S., Torres-Ruiz, R. A., Mayer, U. & Jürgens, G. (2003). Partial loss-of-function alleles reveal a role for *GNOM* in auxin transport-related, post-embryonic development of *Arabidopsis*. Development **131**: 389.

Gilmore, E. (1964). Suggested Method of Using Reciprocal Recurrent Selection in Some Naturally Self-Pollinated Species. Crop Science **4**: 323-325.

Gómez-Cadenas, A., Zentella, R., Walker-Simmons, M. K. & Ho, T.-H. D. (2001). Gibberellin/Abscisic Acid Antagonism in Barley Aleurone Cells: Site of Action of the Protein Kinase *PKABA1* in Relation to Gibberellin Signaling Molecules. The Plant Cell **13**: 667.

Griffiths, S., Simmonds, J., Leverington, M., et al. (2009). Meta-QTL analysis of the genetic control of ear emergence in elite European winter wheat germplasm. Theoretical and Applied Genetics **119**: 383-395.

Guilfoyle, T. J., Ulmasov, T. & Hagen, G. (1998). The *ARF* family of transcription factors and their role in plant hormone-responsive transcription. Cellular and Molecular Life Sciences **54**: 619-627.

Guo, Z., Chen, D. & Schnurbusch, T. (2015). Variance components, heritability and correlation analysis of anther and ovary size during the floral development of bread wheat. Journal of experimental botany **66**: 117.

Ha, C. M., Jun, J. H., Nam, H. G. & Fletcher, J. C. (2004). *BLADE-ON-PETIOLE1* encodes a BTB/POZ domain protein required for leaf morphogenesis in *Arabidopsis* thaliana. Plant and Cell Physiology **45**: 1361-1370.

Haseneyer, G., Ravel, C., Dardevet, M., Balfourier, F., Sourdille, P., Charmet, G., Brunel, D., Sauer, S., Geiger, H. H. & Graner, A. (2008). High level of conservation between genes coding for the *GAMYB* transcription factor in barley (*Hordeum vulgare* L.) and bread wheat (*Triticum aestivum* L.) collections. Theoretical and applied genetics **117**: 321-331.

Heidmann, I., Schade-Kampmann, G., Lambalk, J., Ottiger, M. & Di Berardino, M. (2016). Impedance flow cytometry: a novel technique in pollen analysis. PloS One **11**: e0165531.

Hepworth, S. R., Zhang, Y., Mckim, S., Li, X. & Haughn, G. W. (2005). BLADE-ON-PETIOLE dependent signaling controls leaf and floral patterning in *Arabidopsis*. The Plant Cell **17**: 1434-1448.

Heslop-Harrison, Y. & Heslop-Harrison, J. S. (1996). Lodicule function and filament extension in the grasses: potassium ion movement and tissue specialization. Annals of Botany **77**: 573-582.

Heyne, E. & Livers, R. (1967). Wheat-pollen studies. Wheat News Newsletter **14**: 64-66.

Honda, I., Seto, H., Turuspekov, Y., Watanabe, Y. & Yoshida, S. (2006). Inhibitory effects of jasmonic acid and its analogues on barley (*Hordeum vulgare* L.) anther extrusion. Plant Growth Regulation **48**: 201-206.

Honda, I., Turuspekov, Y., Komatsuda, T. & Watanabe, Y. (2005). Morphological and physiological analysis of cleistogamy in barley (*Hordeum vulgare*). Physiologia Plantarum **124**: 524-531.

Houston, K., Mckim, S. M., Comadran, J., et al. (2013). Variation in the interaction between alleles of *HvAPETALA2* and microRNA172 determines the density of grains on the

barley inflorescence. Proceedings of the National Academy of Sciences **110**: 16675-16680.

Ingvarsson, P. K. & Street, N. R. (2011). Association genetics of complex traits in plants. New Phytologist **189**: 909-922.

Jannink, J.-L. (2007). Identifying quantitative trait locus by genetic background interactions in association studies. Genetics **176**: 553-561.

Jansen, R. C. & Stam, P. (1994). High resolution of quantitative traits into multiple loci via interval mapping. Genetics **136**: 1447-1455.

Jeger, M. & Viljanen-Rollinson, S. (2001). The use of the area under the disease-progress curve (AUDPC) to assess quantitative disease resistance in crop cultivars. Theoretical and Applied Genetics **102**: 32-40.

Jofuku, K. D., Den Boer, B. G., Van Montagu, M. & Okamuro, J. K. (1994). Control of *Arabidopsis* flower and seed development by the homeotic gene *APETALA2*. The Plant Cell **6**: 1211-1225.

Johnson, V. A. & Schmidt, J. W. (1968). Hybrid wheat. Advances in Agronomy **20**: 199-233.

Jôst, M., Glatki Jost, M. & Milohnic, J. (1976). Variability in anther extrusion and pollen production of 51 different wheat cultivars or lines (*Triticum aestivum* L.). Agriculturae Conspectus Scientificus **38**: 113-121.

Jost, M., Taketa, S., Mascher, M., et al. (2016). A homolog of *Blade-On-Petiole 1* and *2* (*BOP1/2*) controls internode length and homeotic changes of the barley inflorescence. Plant Physiology **171**: 1113-1127.

Jumbo-Lucioni, P., Ayroles, J. F., Chambers, M. M., Jordan, K. W., Leips, J., Mackay, T. F. C. & De Luca, M. (2010). Systems genetics analysis of body weight and energy metabolism traits in Drosophila melanogaster. BMC Genomics **11**: 297.

Kamran, A., Iqbal, M. & Spaner, D. (2014). Flowering time in wheat (*Triticum aestivum* L.): a key factor for global adaptability. Euphytica **197**: 1-26.

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S.-Y., Freimer, N. B., Sabatti, C. & Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. Nature Genetics **42**: 348-354.

Khan, M. N., Heyne, E. & Arp, A. L. (1973). Pollen distribution and the seedset on *Triticum aestivum* L. Crop Science **13**: 223-226.

Kherde, M. K., Atkins, I., Merkle, O. G. & Porter, K. (1967). Cross pollination studies with male sterile wheats of three cytoplasms, seed size on F1 plants, and seed and anther size of 45 pollinators. Crop Science **7**: 389-394.

Kirby, E. (2002). Botany of the wheat plant. Bread Wheat. Improvement and Production. Food and Agriculture Organization of the United Nation. Rome: 19-37.

Kirby, E. & Appleyard, M. (1987). Development and structure of the wheat plant. Springer.

Kloth, K. J., Thoen, M. P. M., Bouwmeester, H. J., Jongsma, M. A. & Dicke, M. (2012). Association mapping of plant resistance to insects. Trends in Plant Science **17**: 311-319.

Komaki, M. K. & Tsunewaki, K. (1981). Genetical studies on the difference of anther length among common wheat cultivars. Euphytica **30**: 45-53.

Krasileva, K. V., Vasquez-Gross, H. A., Howell, T., et al. (2017). Uncovering hidden variation in polyploid wheat. Proceedings of the National Academy of Sciences **114**: 913-921.

Kws (2017). Hybrid Rye Growing Guide [Online]. Available: https://www.kws-uk.com/aw/Products-TopMenu/Energy-Crops/Hybrid-Rye-Growing-Guide/~frtr/ [Accessed 22/08/2017 2017].

Kyozuka, J., Kobayashi, T., Morita, M. & Shimamoto, K. (2000). Spatially and temporally regulated expression of rice MADS box genes with similarity to *Arabidopsis* class A, B and C genes. Plant and Cell Physiology **41**: 710-718.

Lander, E. S. & Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121**: 185-199.

Langer, S. M., Longin, C. F. H. & Würschum, T. (2014a). Phenotypic evaluation of floral and flowering traits with relevance for hybrid breeding in wheat (*Triticum aestivum* L.). Plant Breeding **133**: 433-441.

Langer, S. M., Longin, C. F. H. & Würschum, T. (2014b). Flowering time control in European winter wheat. Frontiers in Plant Science **5**.

Larsson, S. J., Lipka, A. E. & Buckler, E. S. (2013). Lessons from *Dwarf8* on the strengths and weaknesses of structured association mapping. PLoS Genetics **9**: e1003246.

Laurie, D. A., Pratchett, N., Allen, R. L. & Hantke, S. S. (1996). RFLP mapping of the barley homeotic mutant lax-a. Theoretical and Applied Genetics **93**: 81-85.

Li, L., Zhang, Q. & Huang, D. (2014). A review of imaging techniques for plant phenotyping. Sensors **14**: 20078-20111.

Licker, R., Johnston, M., Foley, J. A., Barford, C., Kucharik, C. J., Monfreda, C. & Ramankutty, N. (2010). Mind the gap: how do climate and agricultural management explain the 'yield gap'of croplands around the world? Global Ecology and Biogeography **19**: 769-782.

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., Gore, M. A., Buckler, E. S. & Zhang, Z. (2012). GAPIT: genome association and prediction integrated tool. Bioinformatics **28**: 2397-2399.

Liu, J., Jung, C., Xu, J., Wang, H., Deng, S., Bernad, L., Arenas-Huertero, C. & Chua, N.-H. (2012). Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. The Plant Cell **24**: 4333-4345.

Liu, J., Liao, D., Oane, R., Estenor, L., Yang, X., Li, Z. & Bennett, J. (2006). Genetic variation in the sensitivity of anther dehiscence to drought stress in rice. Field Crops Research **97**: 87-100.

Löffler, M., Schön, C.-C. & Miedaner, T. (2009). Revealing the genetic architecture of FHB resistance in hexaploid wheat (*Triticum aestivum* L.) by QTL meta-analysis. Molecular Breeding **23**: 473-488.

Loiselle, B. A., Sork, V. L., Nason, J. & Graham, C. (1995). Spatial genetic structure of a tropical understory

shrub, *Psychotria officinalis* (Rubiaceae). American Journal of Botany **82**: 1420-1425.

Longin, C. F. H., Gowda, M., Mühleisen, J., Ebmeyer, E., Kazman, E., Schachschneider, R., Schacht, J., Kirchhoff, M., Zhao, Y. & Reif, J. C. (2013). Hybrid wheat: quantitative genetic parameters and consequences for the design of breeding programs. Theoretical and Applied Genetics **126**: 2791-2801.

Longin, C. F. H., Reif, J. C. & Würschum, T. (2014). Long-term perspective of hybrid versus line breeding in wheat based on quantitative genetic theory. Theoretical and Applied Genetics **127**: 1635-1641.

Lopes, M. S., Dreisigacker, S., Peña, R. J., Sukumaran, S. & Reynolds, M. P. (2015). Genetic characterization of the wheat association mapping initiative (WAMI) panel for dissection of complex traits in spring wheat. Theoretical and Applied Genetics **128**: 453-464.

Lu, Q., Lillemo, M., Skinnes, H., He, X., Shi, J., Ji, F., Dong, Y. & Bjørnstad, Å. (2013). Anther extrusion and plant height are associated with Type I resistance to *Fusarium* head blight in bread wheat line 'Shanghai-3/Catbird'. Theoretical and Applied Genetics **126**: 317-334.

Lukac, M., Gooding, M. J., Griffiths, S. & Jones, H. E. (2012). Asynchronous flowering and within-plant flowering diversity in wheat and the implications for crop resilience to heat. Annals of Botany **109**: 843-850.

Ma, J., Li, H. B., Zhang, C. Y., Yang, X. M., Liu, Y. X., Yan, G. J. & Liu, C. J. (2010). Identification and validation of a major QTL conferring crown rot resistance in hexaploid wheat. Theoretical and Applied Genetics **120**: 1119-1128.

Mackay, I. J., Bansept-Basler, P., Barber, T., et al. (2014). An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: creation, properties, and validation. G3: Genes|Genomes|Genetics **4**: 1603-1610.

Mangin, B., Goffinet, B. & Rebai, A. (1994). Constructing confidence intervals for QTL location. Genetics **138**: 1301-1308.

Marchini, J. & Marchini, M. J. (2013). Package 'popgen'.

Mayer, U., Buttner, G. & Jurgens, G. (1993). Apical-basal pattern formation in the *Arabidopsis* embryo: studies on the role of the *GNOM* gene. Development **117**: 149-162.

Mccallum, C. M., Comai, L., Greene, E. A. & Henikoff, S. (2000). Targeting induced locallesions in genomes (TILLING) for plant functional genomics. Plant Physiology **123**: 439-442.

Melchinger, A. E., Utz, H. F. & Schön, C. C. (1998). Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics **149**: 383-403.

Meuwissen, T. H., Hayes, B. J. & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. Genetics **157**: 1819-1829.

Muqaddasi, Q. H., Brassac, J., Börner, A., Pillen, K. & Röder, M. S. (2017). Genetic architecture of anther extrusion in spring and winter wheat. Frontiers in Plant Science **8**: 754-763.

Muqaddasi, Q. H., Lohwasser, U., Nagel, M., Borner, A., Pillen, K. & Roder, M. S. (2016). Genome-wide association mapping of anther extrusion in hexaploid spring wheat. PLoS One **11**: e0155494.

Murray, F., Kalla, R., Jacobsen, J. & Gubler, F. (2003). A role for HvGAMYB in anther development. The Plant Journal **33**: 481-491.

Myles, S., Peiffer, J., Brown, P. J., Ersoz, E. S., Zhang, Z., Costich, D. E. & Buckler, E. S. (2009). Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell **21**: 2194-2202.

Nair, S. K., Wang, N., Turuspekov, Y., Pourkheirandish, M., Sinsuwongwat, S., Chen, G., Sameri, M., Tagiri, A., Honda, I. & Watanabe, Y. (2010). Cleistogamous flowering in barley arises from the suppression of microRNA-guided HvAP2 mRNA cleavage. Proceedings of the National Academy of Sciences **107**: 490-495.

Nakaya, A. & Isobe, S. N. (2012). Will genomic selection be a practical method for plant breeding? Annals of Botany **110**: 1303-1316.

Nemeth, C., Yang, C., Kasprzak, P., Hubbart, S., Scholefield, D., Mehra, S., Skipper, E., King, I. & King, J. (2015). Generation of amphidiploids from hybrids of wheat and related species from the genera *Aegilops*, *Secale*, *Thinopyrum*, and *Triticum* as a source of genetic variation for wheat improvement. Genome **58**: 71-79.

Ng, P. C. & Henikoff, S. (2003). SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Research **31**: 3812-3814.

Nguyen, V., Fleury, D., Timmins, A., Laga, H., Hayden, M., Mather, D. & Okada, T. (2015). Addition of rye chromosome 4R to wheat increases anther length and pollen grain number. Theoretical and Applied Genetics **128**: 953-964.

Ning, S., Wang, N., Sakuma, S., Pourkheirandish, M., Koba, T. & Komatsuda, T. (2013a). Variation in the wheat *AP2* homoeologs, the genes underlying lodicule development. Breeding Science **63**: 255-266.

Ning, S., Wang, N., Sakuma, S., Pourkheirandish, M., Wu, J., Matsumoto, T., Koba, T. & Komatsuda, T. (2013b). Structure, transcription and post-transcriptional regulation of the bread wheat orthologs of the barley cleistogamy gene *Cly1*. Theoretical and Applied Genetics **126**: 1273-1283.

Ogawa, T., Uchimiya, H. & Kawai-Yamada, M. (2007). Mutual regulation of *Arabidopsis thaliana* ethylene-responsive element binding protein and a plant floral homeotic gene, *APETALA2*. Annals of Botany **99**: 239-244.

Oshino, T., Miura, S., Kikuchi, S., Hamada, K., Yano, K., Watanabe, M. & Higashitani, A. (2011). Auxin depletion in barley plants under high-temperature conditions represses DNA proliferation in organelles and nuclei via transcriptional alterations. Plant, Cell & Environment **34**: 284-290.

Pandey, S. & Gardner, C. (1992). Recurrent selection for population, variety, and hybrid improvement in tropical maize. Elsevier.

Pánková, K., Milec, Z., Simmonds, J., Leverington-Waite, M., Fish, L. & Snape, J. W. (2008). Genetic mapping of a new flowering time gene on chromosome 3B of wheat. Euphytica **164**: 779-787.

Perry, J. A., Wang, T. L., Welham, T. J., Gardner, S., Pike, J. M., Yoshida, S. & Parniske, M. (2003). A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume *Lotus japonicus*. Plant Physiology **131**: 866-871.

Peterson, R. F. (1965). Wheat: botany, cultivation, and utilization. L. Hill Books.

Platt, A., Vilhjálmsson, B. J. & Nordborg, M. (2010). Conditions under which genome-wide association studies will be positively misleading. Genetics **186**: 1045-1052.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics **38**: 904-909.

Pritchard, J. K. & Rosenberg, N. A. (1999). Use of unlinked genetic markers to detect population stratification in association studies. The American Journal of Human Genetics **65**: 220-228.

Rang, Z., Jagadish, S., Zhou, Q., Craufurd, P. & Heuer, S. (2011). Effect of high temperature and water stress on pollen germination and spikelet fertility in rice. Environmental and Experimental Botany **70**: 58-65.

Rashid, M., Guangyuan, H., Guangxiao, Y., Hussain, J. & Xu, Y. (2012). *AP2*/*ERF* transcription factor in rice: genome-wide canvas and syntenic relationships between monocots and eudicots. Evolutionary Bioinformatics Online **8**: 321-355.

Rebetzke, G. J., Rattey, A. R., Farquhar, G. D., Richards, R. A. & Condon, A. T. G. (2013). Genomic regions for canopy temperature and their genetic association with stomatal conductance and grain yield in wheat. Functional Plant Biology **40**: 14-33.

Reif, J. C., Gowda, M., Maurer, H. P., Longin, C., Korzun, V., Ebmeyer, E., Bothe, R., Pietsch, C. & Würschum, T. (2011). Association mapping for quality traits in soft winter wheat. Theoretical and Applied Genetics **122**: 961-970.

Reynolds, M., Foulkes, J., Furbank, R., Griffiths, S., King, J., Murchie, E., Parry, M. & Slafer, G. (2012). Achieving yield gains in wheat. Plant, Cell & Environment **35**: 1799-1823.

Reynolds, M., Manes, Y., Izanloo, A. & Langridge, P. (2009). Phenotyping approaches for physiological breeding and gene discovery in wheat. Annals of Applied Biology **155**: 309-320.

Riechmann, J. L., Heard, J., Martin, G., et al. (2000). *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. Science **290**: 2105-2110.

Röder, M. S., Huang, X.-Q. & Börner, A. (2008). Fine mapping of the region on wheat chromosome 7D controlling grain weight. Functional & Integrative Genomics **8**: 79-86.

Saini, H., Sedgley, M. & Aspinall, D. (1984). Development anatomy in wheat of male sterility induced by heat stress, water deficit or abscisic acid. Functional Plant Biology **11**: 243-253.

Saito, K., Miura, K., Nagano, K., Hayano-Saito, Y., Araki, H. & Kato, A. (2001). Identification of two closely linked quantitative trait loci for cold tolerance on chromosome 4 of rice and their association with anther length. Theoretical and Applied Genetics **103**: 862-868.

Sakata, T., Oshino, T., Miura, S., Tomabechi, M., Tsunaga, Y., Higashitani, N., Miyazawa, Y., Takahashi, H., Watanabe, M. & Higashitani, A. (2010). Auxins reverse plant male sterility caused by high temperatures. Proceedings of the National Academy of Sciences **107**: 8569-8574.

Seedworld (2015). Syngenta Increases Focus to Develop Hybrid Wheat [Online]. Available: http://seedworld.com/syngenta-increases-focus-to-develop-hybrid-wheat/ [Accessed 22/08/2017 2017].

Shannon, S. & Meeks-Wagner, D. R. (1993). Genetic interactions that regulate inflorescence development in *Arabidopsis*. The Plant Cell **5**: 639-655.

Shevell, D. E., Leu, W.-M., Gillmor, C. S., Xia, G., Feldmann, K. A. & Chua, N.-H. (1994). *EMB30* is essential for normal cell division, cell expansion, and cell adhesion in *Arabidopsis* and encodes a protein that has similarity to *Sec7*. Cell **77**: 1051-1062.

Skinnes, H., Semagn, K., Tarkegne, Y., Marøy, A. & Bjørnstad, Å. (2010). The inheritance of anther extrusion in hexaploid wheat and its relationship to *Fusarium* head blight

resistance and deoxynivalenol content. Plant Breeding **129**: 149-155.

Slade, A. J., Fuerstenberg, S. I., Loeffler, D., Steine, M. N. & Facciotti, D. (2005). A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nature Biotechnology **23**: 75.

Steinmann, T., Geldner, N., Grebe, M., Mangold, S., Jackson, C. L., Paris, S., Gälweiler, L., Palme, K. & Jürgens, G. (1999). Coordinated polar localization of auxin efflux carrier *PIN1* by *GNOM ARF GEF*. Science **286**: 316-318.

Storey, J. D. (2002). A direct approach to false discovery rates. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **64**: 479-498.

Stram, D. O. & Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. Biometrics **50**: 1171-1177.

Sukumaran, S., Lopes, M. S., Dreisigacker, S., Dixon, L. E., Zikhali, M., Griffiths, S., Zheng, B., Chapman, S. & Reynolds, M. P. (2016). Identification of earliness per se flowering time locus in spring wheat through a genome-wide association study. Crop Science **56**: 2962-2672.

Suzuki, S. (1981). Cold tolerance in rice plants with special reference to the floral characters: I. varietal differences in anther and stigma lengths and the effects of planting densities on these characters. Japanese Journal of Breeding **31**: 57-64.

Takuno, S., Terauchi, R. & Innan, H. (2012). The Power of QTL Mapping with RILs. PLoS One **7**: e46545.

Tester, M. & Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. Science **327**: 818-822.

Thornsberry, J. M., Goodman, M. M., Doebley, J., Kresovich, S., Nielsen, D. & Buckler, E. S. (2001). Dwarf8 polymorphisms associate with variation in flowering time. Nature Genetics **28**: 286-289.

Tóth, B., Galiba, G., Fehér, E., Sutka, J. & Snape, J. W. (2003). Mapping genes affecting flowering time and frost resistance on chromosome 5B of wheat. Theoretical and Applied Genetics **107**: 509-514.

Uauy, C., Paraiso, F., Colasuonno, P., Tran, R. K., Tsai, H., Berardi, S., Comai, L. & Dubcovsky, J. (2009). A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. BMC Plant Biology **9**: 115.

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., Maccaferri, M., Salvi, S., Milner, S. G. & Cattivelli, L. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnology Journal **12**: 787-796.

Whitford, R., Fleury, D., Reif, J. C., Garcia, M., Okada, T., Korzun, V. & Langridge, P. (2013). Hybrid breeding in wheat: technologies to improve hybrid wheat seed production. Journal of Experimental Botany **64**: 5411-5428.

Williams, R. F. (1974). The shoot apex and leaf growth: a study in quantitative biology. Cambridge University Press.

Wilson, J. A. (1968). Problems in hybrid wheat breeding. Euphytica **17**: 13-33.

Würschum, T., Langer, S. M., Longin, C. F. H., Korzun, V., Akhunov, E., Ebmeyer, E., Schachschneider, R., Schacht, J., Kazman, E. & Reif, J. C. (2013). Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. Theoretical and Applied Genetics **126**: 1477-1486.

Xu, C., Park, S. J., Van Eck, J. & Lippman, Z. B. (2016). Control of inflorescence architecture in tomato by *BTB/POZ* transcriptional regulators. Genes & Development **30**: 2048-2061.

Xu, M., Hu, T., Mckim, S. M., Murmu, J., Haughn, G. W. & Hepworth, S. R. (2010). *Arabidopsis BLADE-ON-PETIOLE1* and *2* promote floral meristem fate and determinacy in a previously undefined pathway targeting *APETALA1* and *AGAMOUS-LIKE24*. The Plant Journal **63**: 974-989.

Xu, Y. & Crouch, J. H. (2008). Marker-assisted selection in plant breeding: from publications to practice. Crop Science **48**: 391-407.

Yin, X., Stam, P., Kropff, M. J. & Schapendonk, A. H. (2003). Crop modeling, QTL mapping, and their complementary role in plant breeding. Agronomy Journal **95**: 90-98.

Yu, J. & Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. Current Opinion in Biotechnology **17**: 155-160.

Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., Mcmullen, M. D., Gaut, B. S., Nielsen, D. M. & Holland, J. B. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics **38**: 203-208.

Zanke, C., Ling, J., Plieske, J., et al. (2014). Genetic architecture of main effect QTL for heading date in European winter wheat. Frontiers in Plant Science **5**: 1-12.

Zeng, Z.-B. (1994). Precision mapping of quantitative trait loci. Genetics **136**: 1457-1468.

Zhao, K., Aranzana, M. J., Kim, S., Lister, C., Shindo, C., Tang, C., Toomajian, C., Zheng, H., Dean, C. & Marjoram, P. (2007). An *Arabidopsis* example of association mapping in structured samples. PLoS Genetics **3**: e4.

Zhao, L., Zhu, Z., Tian, X., Kobayasi, K., Hasegawa, T., Zhang, Y., Chen, Z., Wang, C. & Matsui, T. (2016). Inheritance analysis of anther dehiscence as a trait for the heat tolerance at flowering in japonica hybrid rice (*Oryza sativa* L.). Euphytica **211**: 311-320.

Zhao, Y., Mette, M. F., Gowda, M., Longin, C. F. H. & Reif, J. C. (2014). Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. Heredity **112**: 638.

Zheng, B., Chenu, K., Fernanda Dreccer, M. & Chapman, S. C. (2012). Breeding for the future: what are the potential impacts of future frost and heat events on sowing and flowering time requirements for Australian bread wheat (*Triticum aestivium*) varieties? Global Change Biology **18**: 2899-2914.

APPENDIX 1: Daily environmental variables during 2014, 2015 and 2016 field trials.

Date	Maximum air temperature (°C)			Minimum air temperature (°C)			Rainfall (mm)		
	2014	2015	2016	2014	2015	2016	2014	2015	2016
1 May	13.7	12.4	15.9	7.3	0.4	-1.2	8	0	0
2 May	10.2	13.1	17.7	1.9	-0.9	5.3	0	0	2.2
3 May	13	19.8	14.4	-0.5	8.7	4.6	0	9.2	0
4 May	16.9	18.1	18	4.7	10.6	2.6	0	4.4	0
5 May	18.2	15.4	19.7	4.4	11.1	5.7	0	3	0
6 May	17.6	14	21.3	10.4	8.5	6.8	0.8	2.8	0
7 May	15.3	15.7	23	9.6	8.2	10.7	0	0.6	0
8 May	14.6	15.5	25.9	-11.7	6.7	11.5	11.2	0	0
9 May	16.3	16	22.4	10.2	6.5	13.9	0	0	0
10 May	17.2	19	15.6	9.4	6.8	14	4.4	0	6.6
11 May	13.1	21	17.7	8.9	11.2	11.6	0.2	0	2.2
12 May	14.7	16.4	22.9	5.6	8.2	9.6	1.2	0.6	0
13 May	14.9	17.2	13.9	6.2	5.3	7.7	5.6	0	0
14 May	17.3	10.4	12	5.9	3.5	1.8	0.2	9.4	0
15 May	20.4	14.7	16.4	6.1	6.8	0.4	0	0	0
16 May	21	17.1	17.6	8.2	6.7	5.2	0	0	0

	Maximum air temperature (°C)			Minimum air temperature (°C)			Rainfall (mm)		
Date	**2014**	**2015**	**2016**	**2014**	**2015**	**2016**	**2014**	**2015**	**2016**
17 May	23.1	15.4	17.5	11.2	6.2	5.2	0	0	0
18 May	23.4	15	14	11	6.3	10.6	0	6.6	3.8
19 May	23.9	12.5	17.4	10.3	5.4	9.8	0	2	0
20 May	21.1	14.7	18.2	11.6	4.8	12.5	1	0	0
21 May	18.5	18.9	18.8	9.3	6	12.5	2.2	0	0
22 May	17.7	20.3	18.8	11.4	8	10.4	8.2	0	1
23 May	17.4	15.8	18	9.8	7.8	5.6	0	0	0.4
24 May	16.8	17.7	16	8.5	7.4	2.9	18.2	0	0
25 May	17.9	14.1	11.4	8.4	6.7	7.3	0	0	0.4
26 May	16.9	17.5	18.3	5.9	7.2	8.3	2.2	0	1
27 May	12.5	18.2	20	9.5	5.7	5.7	29.9	0	0
28 May	14.9	16.9	18.8	9.7	9.8	5.9	9.8	0	0
29 May	18	13.1	17.6	11.5	6.3	6	0	2.8	0
30 May	14.8	16.5	15.5	6.6	6.3	10.5	0	0	0
31 May	17.6	16.4	11.4	3.6	7.3	10	0	3.8	23.8
1 June	20.6	16	11.3	8.8	5.9	9.6	0	0	2.8
2 June	19.9	18.7	10.9	10.8	10.9	8.5	2.8	0.4	0.2
3 June	20.4	19.3	11.5	11.5	6.7	8.1	6.8	0	0.4
4 June	14.4	20.8	18.4	6.7	4.1	10.7	1.8	0	0.2

	Maximum air temperature (°C)			Minimum air temperature (°C)			Rainfall (mm)		
Date	**2014**	**2015**	**2016**	**2014**	**2015**	**2016**	**2014**	**2015**	**2016**
5 June	18.1	23.5	21.2	8.1	9.8	8.4	0	0.4	0
6 June	20.6	18.1	23.8	7	7.5	6.6	0	0	0
7 June	21.9	19.2	24.4	12.5	7.2	7.7	2.2	0	1.4
8 June	24.3	17.1	23	10.8	5	10.1	0	0	5
9 June	25.6	14	21.3	11.6	3.9	9.6	2.2	0	0.4
10 June	22.6	16.9	22.9	11.7	6.8	11.2	0	0	0
11 June	21	22.1	21.8	9.9	3.3	12.2	0	0	0
12 June	24.6	25.3	19.4	8.7	8.4	13.8	0	0	2.2
13 June	24.5	15.6	17.7	10.2	13.4	12.9	1.2	5.8	2.4
14 June	19.4	15.6	18.8	11.5	10.4	11.8	1.4	0.4	7
15 June	17	17.5	21.2	10.7	8.8	10.3	0	0	0.6
16 June	15.8	22.5	19.7	11.3	7.6	9.7	0	0	0.2
17 June	18.9	23.2	17	11.2	11.4	9.8	0	0.8	7
18 June	18.8	19.8	15.1	8	9.7	10.2	0.4	0	0.6
19 June	17.4	18.2	21	7.8	9.2	11.8	0	0	0.8
20 June	21.8	17.8	22	7.8	9.6	14	0	0.6	17.2
21 June	23.2	19.5	20.5	10.4	13.3	13.1	0	0	0
22 June	24.4	17.5	21.7	8.3	10.7	14.3	0	3.6	0.2
23 June	23.9	18	23.4	12.5	10.2	15.6	17.2	0	19.8

	Maximum air temperature (°C)			Minimum air temperature (°C)			Rainfall (mm)		
Date	**2014**	**2015**	**2016**	**2014**	**2015**	**2016**	**2014**	**2015**	**2016**
24 June	22.5	23.1	22	10.3	10.6	11.9	0.4	0	1.2
25 June	19.6	24.6	19.6	6.9	11.7	10.7	0	0	2
26 June	20.1	25.7	20.6	8.7	12.7	11.3	1.2	0	1.2
27 June	20.9	23.3	20.3	11.5	12.1	10.7	5.6	0	0.4
28 June	19.4	19.5	20.5	9.1	13.8	7.5	1.6	1.4	3.6
29 June	16.8	24.5	16.3	9.2	11.7	9.9	9	0	0.4
30 June	20.6	27.9	20.5	8.8	9.8	12.3	0	0	0.6
1 July	22.4	33.3	18.6	7.4	16.3	11.2	0	0	0
2 July	23.4	27	19	6.9	13	8.5	0	0.2	0
3 July	25.9	26.3	21.1	12.4	8.1	9.5	0	0.4	0
4 July	26.8	27.4	22	12.6	15	7.9	0	7.2	0
5 July	20	22.3	20.5	15.8	11.8	9.4	2.6	1.4	0
6 July	21.5	22.4	20.7	11.5	10.9	8	1.2	0	0
7 July	22.3	23.1	22.9	9.7	12.1	13.2	0	2.2	0
8 July	21.1	18.9	23.6	8.3	8.9	14.1	11.4	0.8	0.2
9 July	22	21.3	23.7	11.4	7.9	14.4	0.2	0	0
10 July	15.4	26.2	23.5	12.8	10.3	14.8	5.8	0	0
11 July	17.8	24.7	21.5	12	11.7	13.2	15.6	0	0
12 July	24.9	21.9	17.8	13.2	14.3	11.1	2	5	24

	Maximum air temperature (°C)			Minimum air temperature (°C)			Rainfall (mm)		
Date	**2014**	**2015**	**2016**	**2014**	**2015**	**2016**	**2014**	**2015**	**2016**
13 July	22.1	20	20.1	11.7	15.2	10.2	4.8	2.2	8.6
14 July	23.5	20.4	20.8	11.2	16.1	9.8	0	1	0
15 July	23.8	22.1	19.6	13.2	14.1	9.6	0	0.8	0
16 July	25.4	23.3	25.2	12.6	13.6	15.8	0	9	0
17 July	27.2	24.9	27.4	15.3	13.4	15.5	0	3	0
18 July	31.4	22.1	28.9	15.8	10.6	12.6	4.2	0	0
19 July	28	23.5	31.4	18.1	9.4	14.1	0.4	1	0
20 July	26	24.4	30.1	16	8.1	15.9	0	0	0
21 July	25.2	25.1	24.6	14.5	12.3	14.6	0.2	0	0
22 July	24.4	22.1	25.3	14.4	10.1	14.4	0	0.2	3
23 July	28.2	21.4	27.4	14.8	8.6	13.6	0	0	0
24 July	27.9	16.8	25.7	12.9	11.9	13.7	0	34.9	0
25 July	26.4	19.5	23.1	12.8	10.3	13.5	8.8	2.3	0
26 July	28.5	15.4	22.3	14.9	10.2	12.2	0	5.8	0
27 July	24.8	18.9	25.2	17.1	14.2	15.4	0	1	1.2
28 July	21.5	17.9	21.8	14	11.2	12.1	1	2.4	0
29 July	26.7	19.2	22.3	12.7	9.2	14.8	0	0.2	5.6
30 July	24.8	17.2	23.1	11.3	9	10.3	0	2.4	0
31 July	25.6	21	20.1	12.3	6.2	9.8	0	0	0

APPENDIX 2: List of varieties included in association mapping panel.

Country codes; CHE Switzerland, DEU Germany, DNK Denmark, FRA France, GBR United Kingdom, NLD Netherlands, SWE Sweden.

Variety name	Country of registration	Year of registration
AARDVARK	GBR	1995
ALCHEMY	GBR	2007
ALSACE	FRA	2007
AMAROK	NA	NA
AQUILA	GBR	NA
ARINA	CHE	1981
ARISTOCRAT	GBR	1992
ARK	DEU	1997
ARMINDA	NLD	1976
ARRIVA	GBR	1997
ASAGAI	GBR	2005
ATLANTA	GBR	2005
AWARD	DEU	2000
AXIAL	FRA	1989
AXONA	NLD	1983
BELUGA	GBR	2010

Variety name	Country of registration	Year of registration
BENEDICT	GBR	2006
BERSEE	FRA	1936
BILBO	GBR	1972
BOSTON	NLD	2001
BOUNTY	GBR	1979
BRIGAND	GBR	1979
BROMPTON	GBR	2005
BRYDEN	GBR	1997
CADENZA	GBR	1992
CAMP REMY	FRA	1980
CANTERBURY	GBR	1999
CAPNOR	FRA	2001
CARSTENS V	DEU	1921
CHARDONNAY	GBR	2002
CLOVE	GBR	1992
CONQUEROR	GBR	2005
CONVOY	GBR	1999
COPAIN	FRA	1977
CYBER	GBR	1997
DEBEN	GBR	2000
DENMAN	GBR	2010

Variety name	Country of registration	Year of registration
DICKINS	GBR	1995
DIRECTOR	GBR	2005
EKLA	FRA	1988
FRELON	GBR	2001
GLADIATOR	GBR	2005
GOODWILL	GBR	1999
GRAFTON	GBR	2012
GRAVITAS	GBR	2010
HEDGEHOG	GBR	1976
HEINES PEKO	DEU	1946
HEREFORD	DNK	2007
HOBBIT	GBR	1977
HOLSTER	GBR	1995
HUDSON	GBR	1993
HURLEY	GBR	2005
INVICTA	GBR	2009
IONA	GBR	1977
ISENGRAIN	FRA	1997
ISIDOR	FRA	2002
JACADI	FRA	1997
JB DIEGO	DEU	2007

Variety name	Country of registration	Year of registration
JENA	FRA	NA
KWS HORIZON	GBR	2010
KWS SANTIAGO	GBR	2010
KWS STERLING	GBR	2010
KWS TARGET	GBR	2010
LANGDALE	GBR	2010
LONGBOW	GBR	1980
LORRAINE	FRA	1998
MARIS MARKSMAN	GBR	1976
MEXICO	DEU	NA
MITHRAS	GBR	1980
MONTY	GBR	2007
NEWHAVEN	GBR	1992
NORD DESPREZ	FRA	1945
OAKLEY	GBR	2008
ODYSSEY	GBR	2006
OPTION	GBR	2001
ORBIT	GBR	2010
PANORAMA	GBR	2007
PASTICHE	GBR	1988

Variety name	Country of registration	Year of registration
PENNANT	GBR	2002
POSIT	GBR	1998
PREDATOR	GBR	2004
PROPHET	GBR	1992
RAGLAN	GBR	2001
RAINBOW	GBR	2010
RAMPART	GBR	1998
RENARD	GBR	1983
RIALTO	GBR	1993
ROBIGUS	GBR	2005
ROSETTE	GBR	1995
SABRE	GBR	1982
SANCERRE	FRA	2000
SANTANA	GBR	2009
SARSEN	GBR	1987
SCOUT	GBR	2008
SENTRY	GBR	1977
SHAMROCK	FRA	1998
SHOGUN	GBR	2007
SLADE	GBR	2003
SLEJPNER	SWE	1986

Variety name	Country of registration	Year of registration
SOLDIER	GBR	1988
SOLEIL	FRA	1985
SQUADRON	GBR	1985
STIGG	GBR	2010
TILBURI	FRA	1995
TORFRIDA	GBR	1992
TRAWLER	GBR	NA
VERDON	FRA	2000
WARRIOR	GBR	1993
WICKHAM	GBR	1996
XI19	NLD	2002
ZEBEDEE	GBR	2000