Understanding the genetic and physiological control of pre-harvest sprouting and pre-maturity amylase in UK wheat

Oluwaseyi Adebola Shorinola

A thesis submitted to the University of East Anglia for the degree of Doctor of Philosophy

John Innes Centre

Norwich

September 2015

General Abstract

Pre-harvest sprouting (PHS) and Pre-maturity amylase (PMA) are physiological defects in wheat grains that reduce their end-use quality. PHS is the precocious germination of grains before harvest while PMA is the accumulation of α-amylase in grains. Both traits are quantitative in their expression and are strongly influenced by the environment. In this project, I studied six Quantitative Trait Loci (QTL) located on chromosomes 1A, 2D, 3A (2 loci), 4A and 7B, which confer resistance to PHS or PMA in UK wheat varieties. The aims of this project were to validate and characterise the effects of these QTL, as well as to fine-map the QTL with the most significant effect.

To achieve these aims, isogenic materials were developed to independently study these QTL effects. Physiological characterisation of these QTL showed that they exert their effects by affecting the dynamic of dormancy loss in grains, albeit at different stages of grain development and maturation. We also show that temperature during grain development and germination affect the expression of these QTL effects.

In addition to the characterisation above, I also undertook the fine-mapping and positional cloning of the 4A QTL (named *Phs*), as this QTL showed the highest effect on PHS resistance of all the QTL studied. I took advantage of recent advances in wheat genomics, high-throughput genotyping and the syntenic relationship between wheat and other grasses, to delimit *Phs* to a less than 0.2 cM interval. Furthermore, examination of the physical map of this interval identified 17 genes with varied biological functions. High-resolution fine-mapping of the 0.2 cM interval in three independent and diverse populations further delimited *Phs* to a 10 kb genomic interval. Finally, I report on the comparative sequence analysis around this critical interval, and show the presence of some genomic lesions that could be critical for the *Phs* effect.

Acknowledgments

As an African adage says: “A person is a person because of other persons”. I, therefore, could not have completed this Ph.D. programme without the help, support and guidance of the people around me. I will, first of all, like to thank my supervisor, Dr Cristobal Uauy, for his advices and support throughout the course of this Ph.D. project and especially for his belief in me. I am also grateful to my secondary supervisor, Prof. Alison Smith, for her contributions to this project and towards my personal development.

I acknowledge everyone with whom I worked on this project. I am grateful to Dr John Flintham, who was a member of my supervisory committee. I particularly enjoyed our discussions on the project and his company in the office. Worthy of special mention is Nicholas Bird who was the Research Assistant on this project and who developed many of the material used in this project. I am particularly thankful to him for training me when I joined the Uauy lab. I will also like to thank all the members of the Uauy lab – past and present - for the memorable moments I had in the lab.

This project benefited from productive collaborations with industrial and academic partners. I thank our breeding partners - Peter Jack (RAGT Seeds, UK), Simon Berry (Limagrain UK LTD), Claire Fremann (KWS UK LTD) and Tina Henriksson (SW Seeds). I will also like to thank Barbora Klocová, Dr Miroslav Valárik and Prof. Jaroslav Doležel for their help in assessing the physical map of the critical interval of the 4AL QTL.

I acknowledge the support, comfort, company and friendship of all my friends in the UK. I am particularly grateful to Olumide and Jacob, as well as the members of my local church in Norwich, for keeping me in the realm of sanity especially when the going got tough. I will always remember the fun times we had together.

Words cannot express my gratitude to my family. I am particularly thankful to my brothers Jide and Kayode Shorinola, for their sacrifices that brought me to the UK and for their support during my studies. Lastly, to Him to whom is the essence of all I am and do, I give my thanks.

Table of Contents

List of Figures

List of Tables

Abbreviations

ABA	Abscisic Acid
BAC	Bacterial Artificial Chromosome
BLAST	Basic Local Alignment Search Tool
bp	Base Pair
cDNA	Complementary DNA
CER	Controlled Environment Room
CSS	Chinese Spring Survey Sequence
DH	Doubled Haploid Line
dH_2O	Distilled Water
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleic Triphosphates
DPH	Days Post Harvest
EDTA	Ethylenediaminetetraacetic Acid
g	Acceleration Due To Gravity
GA	Gibberellic Acid
GH	Glass House
GI	Germination Index
HAI	Hour After Imbibition
HFN	Hagberg Falling Number
IWGSC	International Wheat Genome Sequencing Consortium
kb	Kilo Base Pair
Mb	Mega Base Pair
mRNA	Messenger RNA
NABIM	National Association Of British And Irish Millers
NCBI	National Centre For Biotechnology Information
NGS	Next-Generation Sequencing
NIL	Near Isogenic Line
PCR	Polymerase Chain Reaction
PHS	Pre-Harvest Sprouting
PMA	Pre-Maturity Amylase
qPCR	Quantitative Polymerase Chain Reaction
QTL	Quantitative Trait Loci
RCF	Relative Centrifugal Force
RIL	Recombinant Inbred Line
RNA	Ribonucleic Acid
RNAi	RNA Interference
RNAseq	RNA Sequencing
SEM	Standard Error Of The Mean
SNP	Single Nucleotide Polymorphism
SSR	Simple Sequence Repeat
TGW	Thousand Grain Weight
TILLING	Target Induced Local Lesion In Genomes
TPM	Transcript Per Million
UK	United Kingdom
WGS	Whole Genome Sequencing
WGS	Whole Genome Sequencing

1 Chapter One: General Introduction

1.1 Wheat: the food of ancient civilisation

Wild or domesticated cereals belonging to the genus *Triticum L*. are collectively known as wheat. These species form part of the *Triticeae* tribe and some of its cultivated members (Emmer and Einkorn wheat) are often referred to as the universal cereals of the ancient world agriculture (Nevo, 2002, Zohary D, 2000). Together with barley and some wild peas, these species form the founder crops responsible for the development of agriculture in Neolithic settlement located in the Fertile Crescent belt of the Near East - which is the present-day South East Turkey ad Northern Syria (Zohary, 1999, Smith, 1998, Dubcovsky and Dvorak, 2007).

The advent of agriculture in this era resulted in a Neolithic revolution that imparted significantly on ancient human civilisation (Allen, 1997). This agricultural revolution led to the adoption of a more sedentary lifestyle in place of the nomadic hunter–gather existence. In light of these historic contributions, wheat can be considered as a major food of ancient civilisation.

Triticum wheat includes diploid, tetraploid and hexaploid species. Although different taxonomic classifications exist for the *Triticum* genus (Goncharov, 2011), *Triticum* wheat can be generally divided into three sections. Sect. *Monococcon* consists of the diploid *Triticum monococcum L*. (AA genome) and *T. urartu Tumanian ex Gandilyan* (AA genome). The tetraploid wheat - *T. turgidum L*. (AABB genome) *and T. timopheevii (Zhuk.) Zhuk* (AAGG genome) - are in the Sect *Dicoccoidea.* While the hexaploid wheat - *T. zhukovskyi Menabde & Ericz* (AAAAGG) and *T. aestivum L*. (AABBDD) - belong to the Sect. *Triticum* (Goncharov, 2011, Matsuoka, 2011). In addition, many of these species have subspecies. Table 1.1 presents the Van Slageren (Slageren, 1994) taxonomic classification as well as the common name of some of these subspecies (Matsuoka, 2011). Due to the close relationship of some members of the *Aegilops* genus (the *Sitopsis* section) to the *Triticum* genus, some classifications combine the *Triticum* and *Aegilops* genus together to form a common *Triticum* genus (Goncharov, 2011, Bowden, 1959).

The species and subspecies of the *Triticum* genus have diverse morphology that in many instances are reflected in their spikes structures (Figure 1.1). Some of the species have been domesticated and have found different uses in many different cultures around the world and over several millenniums. The most notable of these are *T. aestivum* (bread wheat) and *T. turgidum* (durum wheat) which are used for bread making and pasta, respectively.

Table1.1: Van Slageren Taxonomic classification of the *Triticum* genus as presented by Goncharov (2011)

Section	Species and subspecies	Genome constitution	Common names
Monococcon	*Triticum monococcum L.*		
	subsp. aegilopoides	AA	Wild einkorn
	subsp. monococcum		Cultivated einkorn
	Triticum urartu Tumanian ex Gandilyan	AA	
Dicoccoidea	*Triticum turgidum L*		
	subsp. dicoccoides		Wild emmer
	subsp. dicoccom		Cultivated emmer
	subsp. durum	AABB	Durum/macaroni wheat
	subsp. polonicum L.		Polish wheat
	subsp. turanicum		Khorassan wheat
	subsp. turgidum		Rivet wheat
	subsp. carthlicum		Persian wheat
	subsp. paleocolchicum		Georgian wheat
	Triticum timopheevii (Zhuk.) Zhuk.		
	subsp. armeniacum	AAGG	Wild timopheevii
	subsp. timopheevii		Cultivated timopheevii
Triticum	*Triticum aestivum L.*		Common wheat
	subsp. aestivum		Bread wheat
	subsp. compactum	AABBDD	Club wheat
	subsp. sphaerococcum		Indian dwarf wheat
	subsp. macha		
	subsp. spelta (L.)		Spelt
	Triticum zhukovskyi Menabde & Ericz.	AAAAGG	

1.2 Modern bread wheat is a true allohexaploid

Polyploidy is an attribute of having more than two set of chromosomes (Comai, 2005) and can arise either through whole genome duplication (Autopoplyploidy) or the hydridisation of chromosomes from different species (Allohexaploidy). Modern bread wheat (*T. aestivum* subsp *aestivum*) is an allohexaploid resulting from a hybridisation event between the diploid *Aegilops tauschii* (DD) and the tetraploid T. *turgidum* subsp *dicoccom* (AABB, cultivated emmer). Cultivated Emmer is a domesticated form of the wild *T. turgidum* subsp *dicoccoides* (wild Emmer) which itself is an allotetraploid derived from a hybridisation event between *T. urartu* (AA) and an unknown member of the *Sitopsis* family that donated the B genome (Haider, 2013).

While, in some instances, there are disadvantages associated with polyploidy, polyploidy contributes to the ecological adaption of many species including wheat (Meimberg et al., 2009, Borrill et al., 2015a, Dubcovsky and Dvorak, 2007). Due to its recent origin, there is limited diversity in the modern bread wheat genome. However the genome plasticity afforded by its hexaploid nature is allowing it to generate new diversity rapidly (Talbert et al., 1998, Dubcovsky and Dvorak, 2007). In addition, due to its ability to cross freely with it tetraploid progenitor, new variations have been introduced into the modern bread wheat genome from its progenitor species. However, this is only limited to the A and B genomes as the diploid D genome progenitor does not cross freely with the hexaploid wheat (Akhunov et al., 2010).

Figure 1.1: Diversity in the *Triticum* genus: Spike morphology of the ancestral wheat species as presented by (Eversole et al., 2014). Photo courtesy of Susanne Stamp, Ernst Merz/Eth Zurich

1.3 Wheat domestication: From the Fertile Crescent to the rest of the world

Several traits distinguish the domesticated forms of wheat from their wild forms. Most of the domestication events in wheat occurred in the wild tetraploid wheat (Peng, 2011). Some of the major domestication events include the transition from the shattering, hulled and tough-glumed forms to the non-shattering, soft-glumed and free-threshing forms (Tzarfati et al., 2013).

Some of the domestication genes responsible for these domestication traits have been cloned or mapped to chromosome locations (Matsuoka, 2011, Dubcovsky and Dvorak, 2007). These include *Br-A1* and *Br-B1* on chromosome 3, whose mutated forms gave rise to non-brittle shatter-resistance spikes (Watanabe et al., 2006). The *Tg* locus on the long arm of chromosome 2B is responsible for the non-tenacious glume structure of cultivated wheat (Faris et al., 2014). The *Q* locus on chromosome 5A has pleiotropic effects on a number of spike features including spike length and rachis toughness, and was cloned to be an *APETALA2-like* gene (Faris et al., 2003). Other traits adapted during the domestication and the more recent breeding processes include seed size (Golan et al., 2015, Eckardt, 2010), flowering time (Buckler et al., 2001, Cockram et al., 2007) and plant height architecture (Borojevic and Borojevic, 2005, Zhang et al., 2014), among others.

Acquisition of these adaptation traits enabled the spread of wheat from the Fertile Crescent region to the rest of the world through different routes. Wheat reached Europe through Asia Minor (Turkey) to Greece. From Greece, it spread northward to northern Europe and westward to Italy, France and Spain. Wheat is believed to have reached the UK in about 5000 BP (Shewry, 2009). Wheat also spread to China through the Russian Gansu (Dodson et al., 2013) and to Africa through Egypt.

1.4 Wheat the global food crop: Global production and consumption

Today wheat is grown in more than 128 countries around the world, spanning all the five non-polar continents of the globe. Wheat is the third most-produced cereal in the world, behind rice and maize and has the widest range of cultivation - from latitude 67 °N (Russia and the Scandinavia) to latitude 45 °S (South America and Oceania) (Shewry, 2009). About 649 -716 million tonnes/annum of wheat was produced globally from 2010 -2013 (Figure 1.2). The majority of this was produced in Asia, Europe and North America (FAO, 2014). For 2015, the FAO Cereal Supply and Demand Brief (10/09/2015) forecast a 723 Million tonnes production data for wheat (FAO, 2015). This represents a 1 % increase from 2013 production. However, the demand for wheat is expected to rise by 15 % in the next decade (Edgerton, 2009). This rise in demand is mainly driven by the increasing consumption of wheat in developing countries as a result of increasing urbanisation, as well as, increase in the use of wheat for feed and fuel.

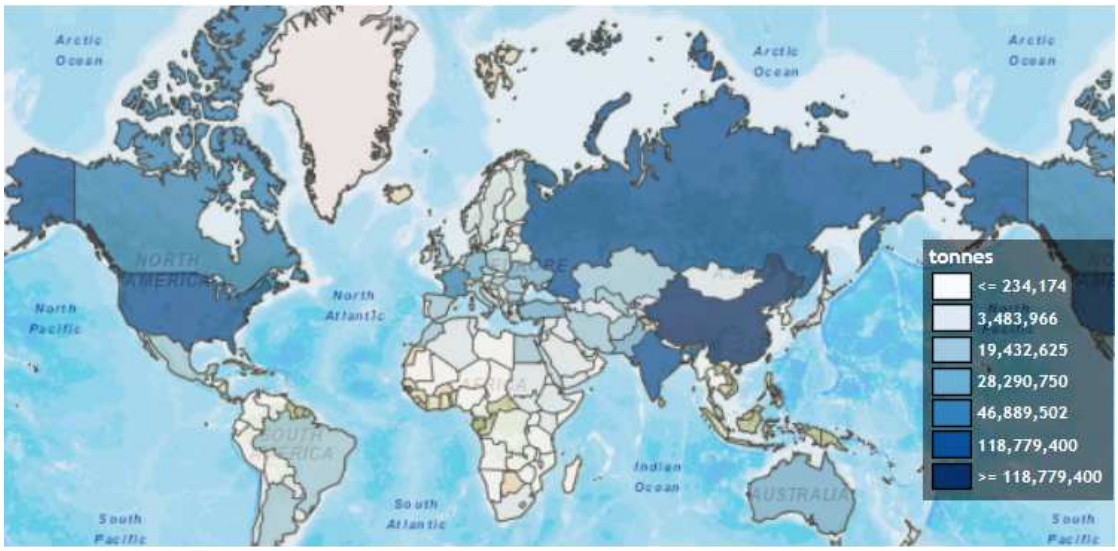


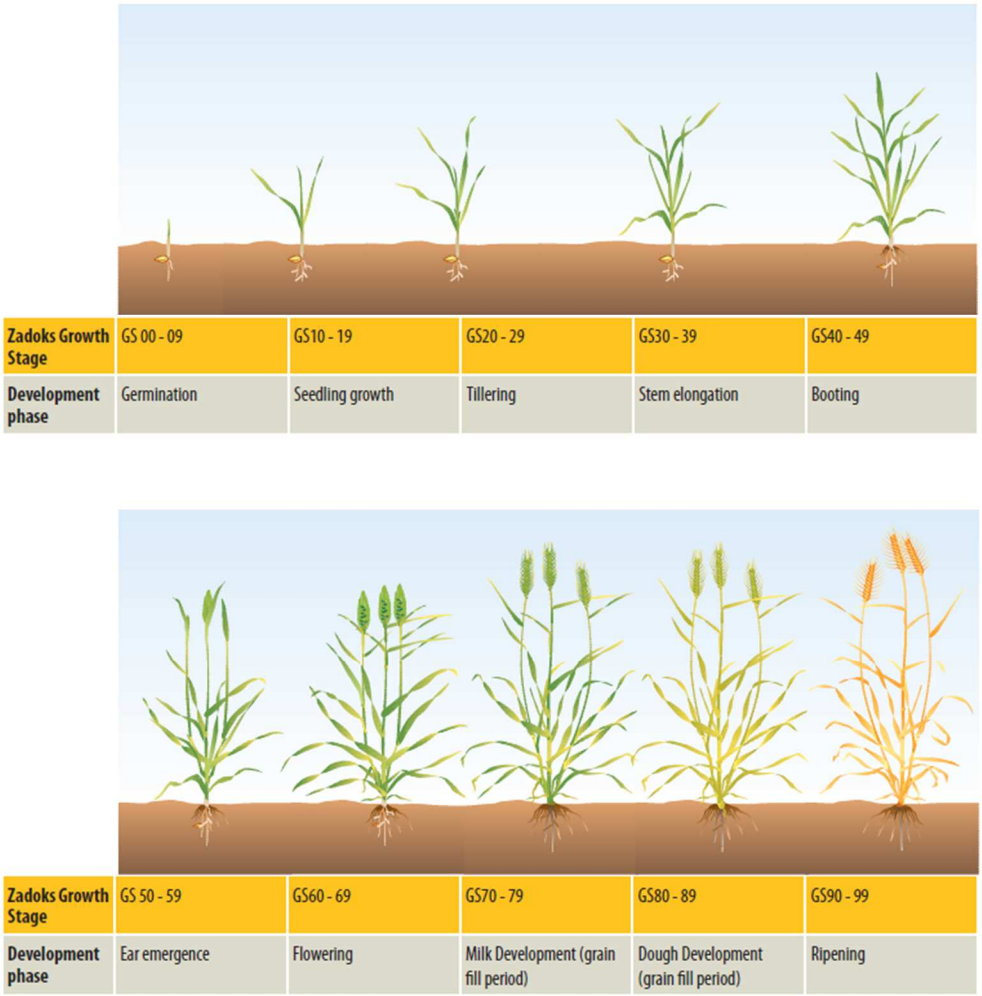
Figure 1.2: Average wheat production quantities from 2010 to 2013 by country. Data obtained from the Food and Agriculture Organization of the United (2014)

Although often underestimated, wheat plays an importantly nutritional role in human diets in that it provides about 20 % of the protein consumed globally. It also provides 20 % of global calories intake. These are taken in the form of leavened and unleavened bread, pasta, noodles, cakes, pastries and the likes. Wheat also has cultural and religious significance in many different societies around the world (Shewry, 2009).

1.5 Growth and development of the wheat plant

Based on their growth habit, modern bread wheat varieties are classified as spring or winter wheat. Winter wheat varieties have requirements for a time of prolonged cold treatment called vernalisation (Li et al., 2013, Dennis and Peacock, 2009) before they can flower. As their name suggests, winter wheats are grown over the winter and are harvested in the following summer. Spring wheats, on the other hand, do not require vernalisation for flowering and can be cultivated within a four-month growth cycle starting from early spring to summer or can alternatively be grown over the winter.

Irrespective of their growth habit and morphology, wheat plants follow the same pattern of development from seed germination to grain development. Many references are available for the description of the various stages of the wheat plant's development, however, the Zadok reference is the most widely used (Zadoks, 1974). This consists of 10 decimal scales/stages with each scale having many different subscales that represent the various processes during each developmental stage (Figure 1.3). In summary: Stage 0 encompasses the different seed germination processes. Stages 10 to 30 describe the vegetative phases of growth which include leaf emergence, tillering and shoot elongation. The reproductive phases of wheat development are represented in Stages 40 through to 60, and these encompass booting, ear emergence and flowering. The last three decimal stages (70 - 90) describe grain development processes, at which end, the plant senescence and dies.



Zadoks Growth Stage	GS 00 - 09	GS10 - 19	GS20 - 29	GS30 - 39	GS40 - 49
Development phase	Germination	Seedling growth	Tillering	Stem elongation	Booting

Zadoks Growth Stage	GS 50 - 59	GS60 - 69	GS70 - 79	GS80 - 89	GS90 - 99
Development phase	Ear emergence	Flowering	Milk Development (grain fill period)	Dough Development (grain fill period)	Ripening

Figure 1.3: The wheat plant's development: The various stages during the wheat plant development according to the Zadok classification system. Diagram obtained from Grain and Research Development Corporation growth stage guide (GRDC, 2015)

1.6 Grain development in bread wheat

A grain refers to the dry indehiscent seed produced by many cereal crops following sexual reproduction. The start of grain development marks the beginning of the end of the wheat growth cycle. Grain development in wheat can be broadly divided into three stages. The first stage involves cell division, differentiation and expansion. This stage is followed by the grain filling stage and finally the grain maturation and desiccation stage (Shewry et al., 2012).

The first stage of grain development is initiated by the double fertilisation event that takes place after pollination. Following pollination, two haploid male nuclei from a pollen grains migrate through the micropyle - a small opening of the embryo sac. One

of these nuclei fuses with the haploid egg nuclei of the female gamete to form the diploid zygote, and this subsequently divides mitotically to form the embryo. The other haploid male nucleus fuses with the two central haploid nuclei of the embryo sac to form the triploid primary endosperm (Friedman, 1998, Hamamura et al., 2012, Evers and Millar, 2002).

Upon formation, the nucleus of the primary endosperm divides multiple times without any accompanying cytokinesis (cell division). This division leads to the formation of a coenocyte, a multinucleate cell with a continuous undivided cytoplasm which has a large vacuole (Olsen, 2004). Subsequent cellularisation of the coenocyte occurs by the formation of the cell wall, and this is completed in wheat grains about four days after fertilisation (Olsen, 2001). Cellular differentiation begins by the formation of the aleurone layer from the peripheral cells while the inner layers cells form the starchy endosperm (Becraft and Yi, 2010).

During the grain filling period, photosynthates (mainly sucrose), as well as protein, from the leaf and stem are transported to the grain in a tightly regulated manner (Borrill et al., 2015b, Wan et al., 2008). The accumulation of storage components (starch and storage protein) in the central endosperm cells result in a substantial expansion of the grain (Shewry et al., 2009, Shewry et al., 2012). The end of the grain filling period is reached when the grain attains it maximum dry weight with ~40 % moisture content, a stage known as physiological maturity. The fully developed grain now enters a maturation stage in which it continues to desiccate for 1-2 weeks to < 20% moisture content (harvest ripeness).

The fully developed wheat grain is therefore made up of the diploid embryo and the triploid endosperm which consists of the starchy endosperm and the aleurone layer (Figure1.4). These are surrounded by the bran consisting of the pericarp and seed coat. These outer layers are formed from maternal tissues (ovary) and they function to protect the developing embryo as well as serve as channel of transmission of environmental cues (Radchuk and Borisjuk, 2014, Chen et al., 2014b).

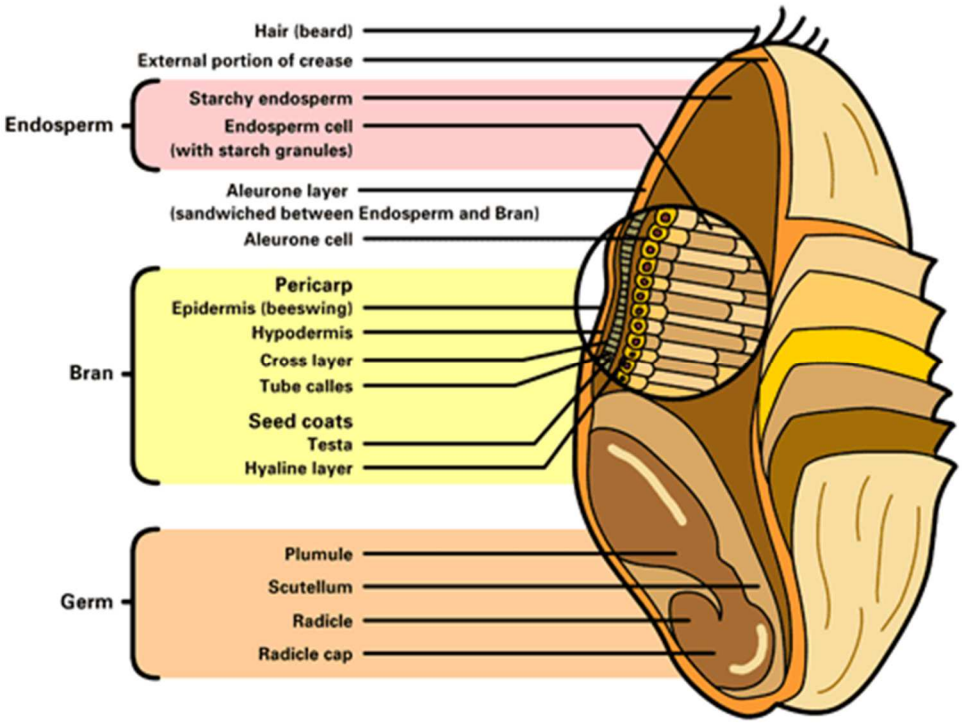


Figure 1.4: Morphological structure of the wheat grain. Photo obtained from www.millersgrainhouse.com

1.7 Seed dormancy, germination and after-ripening

Many definitions have been proposed for dormancy, with the most common definition being the inability of seeds to germinate under condition favourable for germination (Bewley, 1997). As germination is an irreversible process, plants have to time their seed germination to coincide with conditions that are favourable for the development and ultimate survival of the emerging new plant. In view of this adaptive role of dormancy, I have adopted a definition which describes dormancy as any mechanism that prevents the growth of a fully developed and viable embryo during unfavourable seasons even when exposed to a temporarily favourable condition (Graeber et al., 2012). Besides emphasising the adaptive role of dormancy, this definition highlights dormancy as an active process that is achieved through different mechanisms. In addition, this definition also emphasise the antagonistic interaction between the growth potential of a viable embryo and the constraints imposed on the embryo's growth by different dormancy mechanisms in a dormant seed.

Depending on the type and site of growth constraint on the embryo, dormancy can be classified into physiological dormancy (PD), morphological dormancy (MD), morpho-phyisoloigcal dormancy (MPD), physical dormancy (PY) and combinatorial dormancy (PY + PD; Baskin and Baskin, 2004, Nikolaeva, 2004). PD is the most common form of dormancy and can be broken by stratification (cold treatment), scarification (breakage of the seed coat) or hormone treatments. MD occurs in seeds with underdeveloped embryo while MPD combines the features of both PD and MP. PY is caused by a physical hindrance to water movement into the seed by the seed coat (Finch-Savage and Leubner-Metzger, 2006). Also, dormancy expressed in seeds after dispersal from the parent plant is referred to as primary dormancy while dormancy re-introduced as a result of unfavourable environmental condition in long dispersed seeds is referred to as secondary dormancy.

1.7.1 Mechanisms of dormancy imposition and release

Many of our current understanding of the control of dormancy in plants is gained from genetic, molecular and physiological studies in the model plant Arabidopsis, as well as other dicot plants. These studies have revealed many different mechanisms by which dormancy is imposed. Some of these mechanisms discovered in model plants are described briefly below.

1.7.1.1 *Hormonal regulation of seed germination*

Hormones play a vital role in the imposition and release of dormancy, with the interaction between Abscisic Acid (ABA) and Gibberellic Acid (GA) being the most studied. ABA is recognised to promote dormancy while GA promotes germination of the embryo. The ABA-GA interaction is modulated both at the biosynthesis and the signalling level. Holdsworth et al. (2008) and Graeber et al. (2012) provided excellent reviews of this interaction (Figure 1.5).

Some of the evidences for the involvement of ABA in dormancy regulation was derived from genetic studies involving ABA insensitive mutants - *abi1 and abi2* - caused by mutations in *Protein Phosphatase 2C protein* - *ABI1* and *ABI2,* respectively (Koornneef et al., 1984, Gosti et al., 1999, Merlot et al., 2001). The mutants displayed reduced dormancy phenotype even though the level of ABA in the seeds remained the same as

wild-type levels. It has been demonstrated that these phosphatases negatively regulate SNF1-related protein kinases (SnRK2, SnRK3 and SnRK6) which mediate ABA-dependent regulation of ABA-responsive genes (Fujii and Zhu, 2009, Ma et al., 2009).

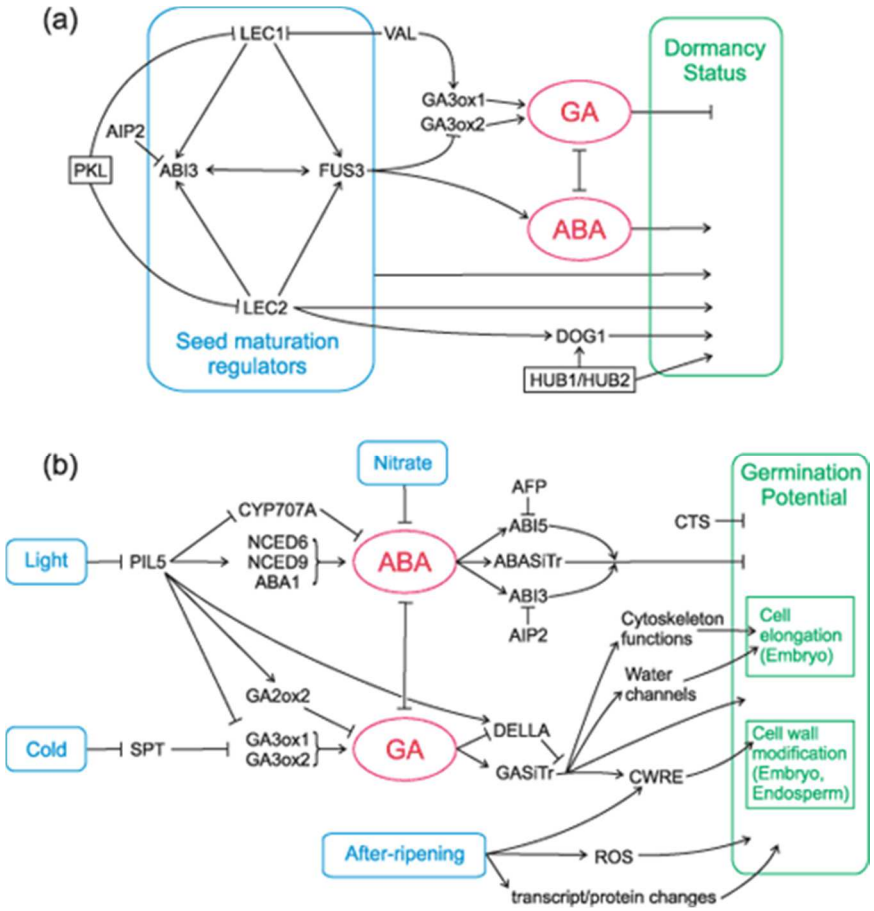


Figure 1.5: Genetic interactions in the control of dormancy in the model plant Arabidopsis: The molecular interactions between the various genes (black fonts) involved in the regulation of dormancy imposition and release during seed maturation (a) and in imbibed seed (b). The hormones ABA and GA are shown in red while the boxed genes represent chromatin remodeling factors. An arrow represents positive interaction and a bar a represents negative or repressive effect (Holdsworth et al, 2008).

In addition, the enzymes involved in ABA biosynthesis and degradation play critical roles in the regulation of dormancy. For instance, an ABA-deficient mutant, *aba1*, with defects in Zeaxanthin epoxidase - an enzyme involved in the early step of ABA biosynthesis - has a non-dormant embryo phenotype (Koornneef et al., 1982). Also over-expression of the gene encoding 9-cis-epoxycarotenoid dioxygenase (NCED), a rate-limiting enzyme in ABA biosynthesis, results in increased seed dormancy (Martínez-Andújar et al., 2011). Conversely, over-expression of the key ABA catabolism gene, ABA 8'-hydroxylase (*CYP707A2*) reduces the level ABA and consequently dormancy in seeds (Millar et al., 2006). Together, these point to a critical role of ABA signalling, biosynthesis and degradation in the regulation of dormancy (del Carmen Rodríguez-Gacio et al., 2009).

GA promotes germination by counteracting the action of ABA (Lee et al., 2002) and by promoting the activity of cell wall remodelling enzymes, as well as the induction of amylase production for the weakening of the endosperm (Linkies and Leubner-Metzger, 2012). Consistent with these functions, mutations in the GA biosynthesis genes - *Gibberelin-3-Oxidase-1 (GA3Ox1)* and *Gibberrelin-3-0xidase-2* (*GA3Ox2*) - result in reduced seed germination potential (Mitchum et al., 2006, Holdsworth et al., 2008). In addition, seeds with mutated copies of the three GA receptors - *Gibberellin Insensitive Dwarf-1 (GID1a, GID1b and GID1c)* - only germinate when the embryo is excised (Griffiths et al., 2006). Also, GA is involved in the regulation and integration of signal from environmental inputs, like light and temperature, into the germination process (Penfield et al., 2005, Yamauchi et al., 2004).

Together with GA, ethylene also counteracts the effect of ABA but mainly functions in promoting endosperm weakening through the activity of cell wall remodelling genes (Linkies and Leubner-Metzger, 2012, Linkies et al., 2009). Transcript level of *Amino Cyclopropane Carboxylate Oxidase* (*ACC Oxidase*), which encodes for an enzyme that catalyses the last step in ethylene biosynthesis, has been shown to increase in imbibed non-dormant seeds before germination (Linkies and Leubner-Metzger, 2012).

Similarly, ethylene signalling has been shown to be important for the regulation of germination. Ethylene reduces seeds responsiveness to ABA, however, mutations in

the ethylene receptor – *Ethylene Resistant-1 (ETR1),* as well as other components of the ethylene signalling including *Constitutive Triple Response-1* (*CTR1*) and *Ethyleine Insensitive-2 (EIN2),* result in the reduction of the inhibition of ABA signalling by ethylene (Beaudoin et al., 2000, Corbineau et al., 2014). Other hormones shown to be involved in dormancy regulation include Cytokinin (Wang et al., 2011) and brassinosteroid (Steber and McCourt, 2001) which promote germination and Auxin which promotes dormancy (Liu et al., 2013c).

1.7.1.2 *Transcriptional regulation of dormancy*

In addition to the role of hormones in the regulation of dormancy imposition and release, dormancy is also regulated at the transcriptional and post-transcriptional level. Evidence for this comes from the isolation of plants with mutations in *Reduced Dormancy 2* (*RDO2)* (Leon-Kloosterziel et al., 1996) which encode a Transcription Elongation Factor II (Liu et al., 2011). As its name suggests, mutations in this gene result in a reduced dormancy phenotype. In the same screen that identified *RDO2*, another mutation in *RDO4* (Peeters et al., 2002) which encode for the H2B MONOUBIQUITINATION1 (HUB1) similarly resulted in a reduced dormancy phenotype (Liu et al., 2007). HUB1 mono-ubiquitinates the histone H2B and this histone ubiquitination is thought to be associated with increased transcription of dormancy promoting gene (Bourbousse et al., 2012).

The roles of some other histone modifying genes in the regulation of dormancy have also been demonstrated. These include *HDA19* and its interactors, *SIN3-like 1(SNL1)* and *SIN3 like 2* (SNL2; Wang et al., 2013). Together these are involved in histone deacetylation which is thought to reduce the transcription of germination promoting genes. Consequently, *hda19* and *snl1snl2* mutants have reduced dormancy phenotype. Together, these provide a strong evidence for the regulation of dormancy at the transcriptional level.

1.7.1.3 *Other mechanisms involved in the regulation of seed dormancy*

Genes involved in the regulation of seed maturation processes like *LEAFY COTYLEDON* (*LEC1* and *LEC2), FUSCA3*, along with the *ABA Insentive 3* (*ABI3*), all function in initiating dormancy in developing seeds (Reviewed in Holdsworth et al., 2008). Consequently, mutations in any of these components result in the precocious germination of seeds. The mechanisms by which other genes which are critical to the regulation of dormancy function are not yet fully understood. For instance, the *DELAY OF GERMINATION1* (*DOG1*) gene was the first cloned QTL underlying natural variation in seed dormancy of Arabidopsis (Finch-Savage and Leubner-Metzger, 2006, Nakabayashi et al., 2012). *DOG1* is a positive regulator of dormancy and encodes a protein of unknown molecular function. However, subsequently characterisation of this gene is beginning to reveal the role of DOG1 in mediating seed dormancy response to ambient temperature (Graeber et al., 2014, Chiang et al., 2011).

1.7.2 Regulation of dormancy in cereals

Unlike dormancy research in model dicot species, understanding of the molecular regulation of dormancy in cereals are just recently emerging. This progress in dormancy research in cereals has been partly driven by the need to tackle the detrimental effects of too early seed germination in cereals crops, a phenomenon known as pre-harvest sprouting (discussed in section 1.8.2). This precocious seed germination in cereal species is generally believed to be a consequence of the unbalanced seletion for crop genotypes with low seed dormancy during the early/Neolithic domestication events and under recent conventional breeding programmes. In support of this notion, semi-wild Tibetan wheat species display deep embryo dormancy in relation to elite varieties (Lan et al., 2005). However the level and type of dormancy displayed in wild cereal species can also vary (Veasey et al., 2004). While such selection enabled early and uniform germination of seeds in elite cultivars and varieites, it invariably resulted in increased susceptibility to pre-harvest sprouting. In this section, I will briefly discuss some of the current understanding of the regulation of dormancy in cereals.

1.7.2.1 *Hormonal regulation of seed dormancy in cereals*

Just like in model diploid species, hormones play key roles in the regulation of dormany in cereals, with ABA promoting dormancy and GA promoting germination. One of the foremost genetic evidence for this in cereals came from the identification of maize mutants with viviparous (Vp; precocious germination) phenotypes. There are at least 15 *vp* mutants in maize with almost all of these impaired in genes involved in ABA biosynthesis or signalling (Giulini et al., 2011). For example the maize *Vp1*, which is an orthologues of *AB13*, is involved in ABA signalling during grain development (McCarty et al., 1989). However, most of the transcripts of the orthologues of *Vp1* in wheat have been shown to be mis-spliced (McKibbin et al., 2002), leaving a relatively low level of functional *Vp1* transcripts, most of which originate from the B genome.

Also, *Vp1* expression has in some studies been associated with the resistance to sprouting in wheat (De Laethauwer et al., 2012, HaiPing et al., 2009), however, this has not been extensively supported by genetic evidence. Similarly in rice, some mutants which displayed vivipary phenotypes (*phs1-4*) have mutations in genes involved in the biosynthesis of the carotenoid precursor of ABA synthesis (Fang et al., 2008). Consequently, these mutants have reduced ABA content.

Despite this viviparous phenotypes of mutants defective in ABA biosynthesis, majority of the natural variation in dormancy in wheat are not attributed to differences in ABA content but are rather associated with difference in ABA signaling or responsiveness. For instance, dormant wheat seeds are generally associated with high level of responsiveness to ABA inhibition of germination (Gerjets et al., 2010, Walker-Simmons, 1987). These responsiveness to ABA might be mediated by the ABA responsive protein phosphatase, *PKABA1,* which is an orthologue of *ABI1* (Anderberg and Walker-Simmons, 1992) and the ABA responsive transcription factor, *TaABF1, which is* an orthologue of *SnRK2 (Rikiishi et al., 2010)*. In rice, *Sdr4*, was the first QTL cloned for the regulation of dormancy and is involved in ABA signalling. This is because its expression is upregulated by the transcription factor - *OsVp1*. In addition, higher expression of *Sdr4* is associated with increased ABA sensitivity (Sugimoto et al., 2010).

Similarly, GA signaling also impact on seed dormancy phenotype. GA induces conformation changes upon binding to its receptor, GA INSENSITIVE DWARF1 (GID1). This enables GID1 to bind to DELLA proteins which are negative regulators of GA signaling (Willige et al., 2007, Hirano et al., 2008). This binding of GID1 to DELLA results in the degradation of the DELLA proteins and consequently the activation of GA signaling. In wheat, the Rht (Reduced height) alleles underlying the dwarfing phenotypes produce GA-insensitive DELLA proteins. Interestingly, some of the Rht alleles, like Rht3 (or RhtB1-c), with a severe dwarfing phenotype, also has a dormant seed phenotype (Flintham and Gale, 1982)

1.7.2.2 *Environmental regulation of seed dormancy in cereals*

In other to ensure that seeds germinate at the right enviromental conditions conducive for the growth and survival of the emerging plants, sucessful adaptive species incorporates environmental clues to modulate their seed dormancy response. One of the most important environmental factor controlling dormancy in cereal seeds is temperature, both during grain development (developmental temperature) and during during germination (incubation temperature). Lower temperature during grain developmental induces higher dormancy in mature grains in temperate cereals like wheat and barley (Nakamura et al., 2011a, Gualano and Benech-Arnold, 2009) Conversely in wheat, lower incubation temperature during grain germination breaks dormancy while higher incubation temperature inhibit germination (Nyachiro et al., 2002). However in rice, the effect of incubation temperature depends on the presence of hull and the stage of seed development (Ueno and Miyoshi, 2005)

The molecular pathways by which temperature clues are integrated into dormancy regulation are just being uncovered in cereals. The wheat *Mother of Flowering Time* gene (*TaMFT*) was recentlly cloned and found to underly a QTL for natural variation in wheat seed dormancy. *TaMFT* was shown to regulate dormancy when grains are developed under low ambient temperature (Nakamura et al., 2011a, Lei et al., 2013). In barley, germination response to incubation temperature is associated with increased expression of genes involved in ABA and GA metabolism. For instance, dormancy induced at higher incubation temperature in barley is assoacited with increased expression of *HvABA8'OH* - an ABA catabolism gene (Hoang et al, 2013).

The regulation of dormancy by light have been demostrated in *Brachypodium*, barley and oat. In barley, white and blue light promote dormancy (inhibit germination) in imbibed seeds (Gubler et al., 2008). In addition, far-red light also inhibit germination in *Brachypodium*, while red light promote germination (Barrero et al., 2012). Light-induced promotion of dormancy is associated with increase in ABA biosynthesis due to increased expression of ABA biosynthesis genes (*HvNCED1*), and is regulated by the *CRYTOCHROME1* genes *CRY1* and *CRY2 (Barrero et al., 2014)*.

Oxygen is also required for the release of dormancy and the progression of germination in many cereals (Alani et al., 1985). There is induction of secondary dormancy in barley seeds inbibed under low oxygen conditions (Hypoxia;Kırmızı and Bell, 2012, Hoang et al., 2013). This is associated with reductions in GA content, ABA catabolism and increased sensitivy to ABA (Benech-Arnold et al., 2006).

1.7.2.3 *Contributions of seed covering structures to dormancy control in cereal*

Except for a few variation, there is a general conservation of the structural composition of the grain among cereal species. As explained in section 1.6, this is generally comprised of the embryo, endosperm and the testa fused to the pericarp which provides covering for the grain. In addition to these covering structures, some cereals also contain hull, comprised of lemma and palea or glume, which can be removed by threshing (like wheat and rice) or cannot be removed by threshing (like barley and oat; Rodríguez et al., 2015).

These covering structures serve as physical passage for the entry of water and air to the grains and also other water soluble chemicals that affect germination. In some varieties of wheat, rice and maize, the presence of proanthocyanin is responsible for the red colour of grains and this promote dormancy of the seed (Himi et al, 2002, reviewed in Rodriguez et al, 2014). This has led to the notion that red-coloured grains are more dormant than white-colored grains. In cereals whose hull cannot be easily removed during threshing like barley and oat, the hull contribute to grain dormancy by limiting the amount of oxygen available to the mature seeds (Hoang et al 2013, Benech-Arnold, 2006). Also this oxygen deprivation can occur during imbibition as a

result of the oxidation of phenolic compounds in the hull, thereby limiting the amount of oxygen available to the embyro (reviewed in Rodriguez et al 2014). In addition, leachate from the hull could inhibit germination in some cereals (Gatford et al., 2002).

1.7.3 Seed germination

Seed germination starts with the uptake of water into the seed (imbibition) and ends with radicle extension of the embryo. In some species, uptake of water is through the seed coat, but other species have physical barriers to imbibition in their seed coat (Rolston, 1978, Baskin and Baskin, 2001). Rathjen et al. (2009) reported that the movement of water into the wheat seed occurs through the micropyle in both dormant and non-dormancy lines. (Bewley, 1997). Imbibition occurs in three phases. In the first phase, there is a rapid initial uptake of water; this is followed by a plateau phase when the rate of imbibition is static. The third phase of imbibition occurs once germination is completed, as the embryo begin to elongate.

Following the first phase of imbibition, rehydration of cell structures occurs, and this results in the resumption of metabolic activity in cells. These include respiration, protein synthesis and active transcription of genes. The energy for the initial phase of germination is driven by anaerobic respiration due to the lack of oxygen molecules within the grain. Then as oxygen uptake increases, normal aerobic respiration resumes. The final stage of germination is marked by the elongation of the embryo.

1.7.4 After-ripening

Seed after-ripening occurs during a period of dry storage of seeds. For the seeds of many different species, after-ripening is required before dormancy can be broken. Although what happens in dry seeds during after-ripening is yet to be well defined, after-ripening results in changes to the physiological status of seeds. One of such physiological changes associated with after-ripening is a reduction in ABA content and ABA sensitivity of dry after-ripened seeds (Ali-Rachedi et al., 2004). However, in wheat

seeds, ABA content remains the same in dry after-ripened non-dormant and dormant seeds (Liu et al., 2013a) while the loss of ABA sensitivity is associated with after-ripening (Schramm et al., 2012). In addition, Liu et al. (2013a) showed that after-ripening reduces ABA sensitivity at the transcriptional level by the repression of the genes involved in the ABA signalling cascade – *PP2C* and *SnRK*s.

The mechanisms by which after-ripening mediate these physiological changes are not well understood, but it has been proposed that active transcription in dry seeds may play a part in the after-ripening process. The evidence for this comes from changes (increases and decreases) in transcripts levels during after-ripening of dry seeds (Holdsworth et al., 2008). In support of this active transcription in dry seeds, lowly hydrated areas have been found in maternal tissues of dry tobacco seeds and these were associated with low level of transcription of the *β-1-3 Glucanase* genes (Holdsworth et al., 2008, Leubner-Metzger, 2005). However, it has been recently shown in sunflower seeds that, transcriptional difference does not account for the physiological difference in after-ripened seeds and dormant seed (Meimoun et al., 2014). This may suggest differences in the regulation of after-ripening in species. Other mechanisms associated with the after-ripening effect is the action of Reactive Oxygen Species (ROS) which accumulate during after-ripening resulting in selective post-translational modification of mRNA and proteins through oxidation (Oracz et al., 2007, Bazin et al., 2011, El-Maarouf-Bouteau et al., 2013).

1.8 Hagberg Falling Number (HFN): an important grain quality trait in bread wheat

The amount of α-amylase within grains of wheat determines their suitability for milling and baking purposes. High α-amylase activity results in the hydrolysis of grains' starch content, which subsequently leads to the accumulation of excess sugar in the milled grains. The consequence of this high sugar accumulation it that dough made from such flour becomes too sticky and difficult to process. This presents a problem for many automated processing infrastructures in large milling and baking companies.

In addition, there is a significant reduction in the end-use quality of flour made from grains with high α-amylase content. While a level of α-amylase activity is required for bread-making, too high α-amylase activity in flour can result in lower volume and course textured bread (Figure 1.6). However, it has recently been shown that the isoform of α-amylase, and not just the total amylase content, determines the end-use quality of grains. This is because the accumulation of a specific type of α-amylase (α-Amy3) improves the bread-making characteristic of bread wheat grains (Ral et al., 2015).

Hagberg Falling Number (HFN) is an indirect measure of this α-amylase activity within grains of wheat. As the name suggests, the HFN values of a batch of grains is determined through a Falling Number test developed by Hagberg and Perten (Hagberg, 1960, Perten, 1964). The HFN test is a viscometric method that measures the time it takes for a steel ball to fall through a viscous flour slurry. Flour made from grains with high amylase will have less starch content and high sugar content due to the activity of the amylase. Such flour slurry will therefore be less viscous, and the steel ball will fall through quickly.

A low HFN number is therefore indicative of high α-amylase activity and vice versa. To guarantee good quality end products, milling agents in Europe (NABIM in the UK) specify HFN values => 250 for grains to be accepted for milling. The two main physiological causes of low HFN in wheat grains are pre-harvest sprouting (PHS) and pre-maturity amylase (PMA). Both are seed defects that are accompanied by the accumulation of α-amylase in seeds. These will be discussed briefly.

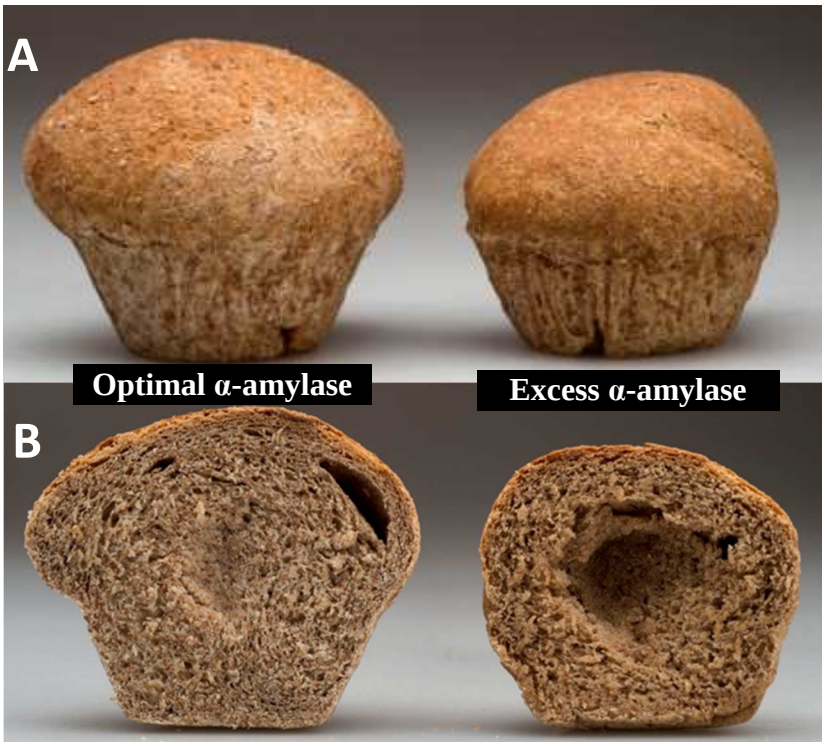


Figure 1.6: α-Amylase content affects the bread-making quality of wheat. A. Bread from flour with excessive α-amylase activity has smaller volume compared to bread made from grains with optimal α-amylase level. B. Longitudinal section of bread made from flour with too much α-amylase activity and optimal α-amylase activity; excessive α-amylase activity results in bread with course and inelastic texture.

1.8.1 Pre-harvest sprouting (PHS) in wheat

PHS refers to the too early germination of wheat grains on the mother plant before harvest. As germination induces the activity of α-amylase for endosperm weakening and starch remobilisation, the residual α-amylase in germinated grains constitutes a problem for downstream processing of grains i.e. milling and bread-making. Environmental condition during grain maturation and ripening play a crucial role in the induction of PHS. Exposure to rain just before harvest ripeness predisposes plants to PHS during subsequently raining (Mares and Mrva, 2014). Also, ambient temperature during the later stages of grain development and maturation affects the tendency of grains to sprout. This is because high temperature during this period tends to result in greater susceptibility of grains to PHS (Mares and Mrva, 2014, Barnard and Smith, 2009). These suggest a strong environmental influence on the induction of PHS.

Irrespective of this environmental dependency of sprouting induction, there exists a large variation in the susceptibility of wheat genotypes to PHS. This variation suggests

an underlying genetic control of PHS resistance that is quantitative in nature. This resistance to PHS can be expressed in different ways. Physical factors associated with PHS resistance include the morphology of the spike/ear as this can determine how much water is retained by the spikes and how much gets through to the grain (King and von Wettstein-Knowles, 2000). In support of the effect of spike structure on PHS susceptibility, Flintham et al. (2002) reported the association of the presence of awn with increased sprouting in ears. However, in most of the study of PHS resistance, genetic variation in seed dormancy accounts for majority of PHS resistance observed, this makes grain dormancy the most promising target for PHS resistance. Also, PHS resistance has been associated with seed ABA sensitivity (Walker-Simmons, 1987).

In support of the quantitative nature of the genetic resistance to PHS, many Quantitative Trait Loci (QTL) have been identified in numerous populations across the world, and these account for varying degree of resistances in these populations. These QTL are located on almost every chromosome of the wheat genome. This makes PHS resistance one of the most multi-genic traits in wheat and also highlights the complexity in the control of this trait. Table 1.2 presents a modified summary of some of these QTL as reported by Gao et al. (2013b). Some of the QTL identified in these different studies co-localise, suggesting that some main QTL are segregating in many different populations. Examples of these are the QTL on chromosome 3AS and 4AL which have been most frequently reported to have high-magnitude effect on PHS resistance. Besides these QTL studies, there have also been numerous Genome-wide Associations Studies (GWAS) with different panels of wheat varieties (Kulwal et al., 2012, Jaiswal et al., 2012, Lin et al., 2015, Rehman Arif et al., 2012). Many of the linked loci found in these association studies co-localise to previously identified QTL loci.

In addition, PHS has also been historically linked to grain colour. The dominant Red grain colour genes (*R-1)* control wheat grain colour and they are located on three independent but homoeologous chromosome 3 loci (*R-A1b, R-B1b and R-D1b*). Generally, red-grained wheat varieties are considered as PHS resistant while white-grained varieties are considered as susceptible. Results from some genetic studies (Himi et al., 2011, Flintham, 2000) suggest that the increased dormancy of red-grained line is due to the pleiotropic effect of *R-1* and not just a linkage effect. However, we

still do not yet fully understand how these genes make red-grained wheat varieties dormant.

Despite this association of grain colour to PHS resistance, there are still considerable variations in PHS resistance both within white-grained and red-grained wheat populations. In fact, some red-grained varieties could be more susceptible to PHS than some white-grained varieties (Kottearachchi et al., 2006, Mares and Mrva, 2014). Also, Flintham (2000) showed that other loci control grain dormancy in wheat besides the *R-1* loci.

1.8.2 Pre-maturity amylase (PMA) in wheat

PMA, which is also called Late Maturity Amylase (LMA), is a physiological seed defect that is characterised by the accumulation of high PI (isoelectric point) amylase in grain at the later stages of grain development, all in the absence of germination. Unlike PHS, PMA has not been studied extensively. This is probably due to the lack of observable symptoms of PMA in seeds. This has led to the incidence of PMA being unnoticed until it is detected by millers through the HFN test prior to grain receival.

There are two main classes of α-amylase enzymes in wheat. These are the high PI α-Amy1 that consists of two isoforms (α-Amy1-1 and α-Amy1-2) and the low PI α-Amy2 (Cheng et al., 2014, Lazarus et al., 1985). The *α-Amy1* genes are located on the homoeologous group 6 chromosomes (6A, 6B and 6D), while the *α-Amy2* genes are located on homoeologous loci on the long arm of chromosome 7. There is also a third class of amylase, *α-Amy3*, expressed only in immature grains (Baulcombe et al., 1987).

Table 1.2: Summary of QTL identified for PHS resistance (Gao et al., 2013b)

Trait	QTL	Linked Markers	Population	References	Comment
	4AL	*Xcdo795 - Xpsr115*			
Dormancy	4BL	*Xbcd1431.1/Xbcd1431.2*	DH	Kato et al. (2001)	Red-grained
	4DL	*Xbcd1431.1/Xbcd1431.2*			
	3AL	*Xfbb293*			
PHS	3BL	*Xgwm403, Xbcd131*	RIL	Groos et al. (2002b)	Red x White grained
	3DL	*Xgwm3*			
	5AS	*Xbcd1871*			
	1BS	*Xpsp3000*			
PHS	4BL	*Xpsp3030 - Xpsp3078*	RIL	Flintham et al. (2002)	Red-grained
	7AS	*Xpsp3050*			
	4AL	*Xpsr1327*			
Dormancy	4A	*Xgwm397 - Xgwm269*		Mares et al. (2005)	Red x White
Dormancy	3BL	*wmc527- gwm77*	DH	(Mares et al., 2009)	White-grained
PHS/	4AS	*Xwmc48 - Xgwm397*	RIL	Ogbonnaya et al. (2008)	White-grained
Dormancy	4AL	*Xgwm0637 - Xgwm937*			
	1AS	*Xwmc24-Xbarc119*			
	2AL	*Xwmc170d-Xcfd168*			
		Xgwm1045-Xgwm296			
PHS	2B	*XE36M605 - XE36M607*	RIL	Mohan et al. (2009)	Red x white grained
	2DL	*Xcfd168 - Xcfd168a*			
		Xcfd44 - Xgwm53			
	3AL	*Xwmc153 - Xgwm155*			
	3BL	*Xgwm1005 - Xgwm980*			
PHS	3AS	*BARC310*	RIL	Kottearachchi et al. (2006)	White-grained
	4AL	*BARC170*			
PHS	3A	*Xgwm155*	RIL	Kulwal et al. (2005)	Red x white
	3AS	*Xbarc310 - Xbcd907*			
PHS	4AL	*Xcdo189-Xcdo795/Xcdo*	RIL	(Mori et al., 2005)	Red
	4BL	*Xgwm495 - Xgwm375*			
PHS	2DS	*Xgwm261 - Xgwm484*	F_2 & F_6	Xiao-bo et al.	Syn x Hexaploid
Dormancy	3AS	*Xgwm5 - Xpsr394*	RIL	(Osa et al., 2003)	Red
	3A	*Xcfa2193 - Xwmc594*			
PHS	3B	*Xbarc77 - Xwmc307*	DH	(Fofana et al., 2009)	Red x White grained
	3D	*Xwmc552 - Xwmc533*			
	5D	*Xgwm469 - Xcfd10*			
	2B	*Xbarc55 - Xwmc474*			
PHS	2D	*Xwmc111 - xwpt-9997*	DH	(Munkvold et al., 2009)*	White-grained
	3D	*Xbarc1161 - xgpw4152*			
	6D	*Xcfd37 - xbarc196*			
PHS/	4AL	*Xksuf8a - Xbcd402b*	RIL	(Lohwasser et al., 2013)	Red x synthetic
Dormancy	3AL	*Xpsr903b - XATPased*			
	1A	*Xwmc611 - Xwmc333*			
PHS	2A	*Xgwm515 - Xgwm425*	RIL	(Knox et al., 2012)	Durum
	7B	*Xgwm297- Xwmc532*			
		RGC			
	3D	*Xwms1200*			
Dormancy		*Xgwm341*	BC_1F_7	(Imtiaz et al., 2008)	Synthetic
	4A	*Xgwn269c*			
		Xwms894			
	2B	*Xdup398 - Xbarc54*			
Dormancy		*Xbarc105 - Xbarc334*	RIL	(Liu et al., 2008)	White grained
	3AS	*Xbarc12 - Xbarc321*			

DH: Doubled Haploid
RIL: Recombinant Inbred Line
*Other QTL were identified in this study but only the most stable QTL are listed

During grain development, there is usually an accumulation of α-Amy2 protein between 10 - 20 days post anthesis. But the level of the enzyme declines afterwards as grains enter the maturation phase. In contrast, the α-Amy1 proteins are usually not synthesised during grain development but are involved in seed germination upon induction by GA synthesised in the embryo. However in PMA defective seeds, there is any accumulation of α-Amy1 at the late stages of grain maturation (Mares and Mrva, 2008, Barrero et al., 2013).

Just like PHS, there is a strong genetic x environmental interaction in the control of PMA. Certain environmental conditions can induce PMA. Cold and heat temperature shock have been shown to induce PMA successfully but only in PMA susceptible lines (Farrell and Kettlewell, 2008). Mrva et al. (2006) showed that there is a period of sensitivity to the induction of PMA by cold temperature, and this is between 25 - 35 days after anthesis. However unlike PHS, not many QTL have been identified for the control of PMA. Four QTL have been reported on chromosomes 2D, 3A, 6B and 7B with the 7B QTL producing the largest effect in many different germplasms.

1.9 Wheat genomics comes of age: Advances in the development of wheat genetic, genomic and germplasm resources

Up until the last few years, qualitative and quantitative genetic research in wheat was cumbersome, expensive and time-consuming. This was primarily due to the lack or shortages of genomic and genetic resources with which to study gene functions in wheat. Also due to the polyploid nature of hexaploid wheat, structural and functional redundancy between the various homoeologous gene copies made it difficult to study individual gene functions. This was further made complicated by the high repeat (transposable element) content of the bread wheat genome which makes up about 80 % of the 17 Giga base-pair (Gb) size of the hexaploid wheat genome (Choulet et al., 2010, Smith and Flavell, 1975).

In contrast to wheat, genetically tractable model plants like Arabidopsis, rice and *Brachypodium*, which have simple diploid genomes, have abundant genetic, genomic

and germplasm resources (Lamesch et al., 2012, TAGI, 2000, International Rice Genome Sequencing Project, 2005, The International Brachypodium Initiative, 2010). In addition, all of these model species have well-annotated reference genome sequences which have proved invaluable in advancing research efforts in this species. This is not the case for wheat which has a less than complete reference genome sequence.

However, within the last five years there has been a genomic revolution in wheat research, and this has led to the development of many genetic, genomic and germplasm resources. These new developments were recently reviewed by Borrill et al. (2015a). Advances in wheat genomics began with a whole genome sequencing (WGS) of the Chinese Spring variety (Brenchley et al., 2012a). This yielded 85 Gb of sequence representing a 5 x coverage of the wheat genome. As a result of the highly repetitive nature of the wheat genome and the high sequence similarity between homoeologous sequences, the assembly obtained from this WGS was highly fragmented and it was not possible to separate homoeologous sequences. Nevertheless, this represented the first step towards obtaining a reference genome sequence in wheat.

To overcome the complexity associated with WGS of wheat, a different approach that employed chromosome flow-cytometric sorting was used to separate and sequence individual chromosome arms of the Chinese Spring genome and these were sequenced at a 30 x coverage (International Wheat Genome Sequencing Consortium (IWGSC), 2014). This new genomic resource called Chinese Spring Survey (CSS) sequence was generated by the International Wheat Genome Sequencing Consortium (IWGSC) and is advantageous in that it separates homoeologous sequence. In the last few months however, the CSS assembly has been complemented by another WGS assembly of a synthetic hexaploid wheat (W7984) which has longer assembly scaffold than the CSS (Chapman et al., 2015).

While these two additional sequences resources represent significant advancements in wheat genomic, the scaffolds from these assemblies are still non-contiguous. This means there is currently no contiguous reference pseudomolecule for wheat. The only

exception to this is the assembly made from flow-sorted chromosome 3B (Choulet et al., 2014), which was the first chromosome to be purified by flow cytometry and to have its physical map constructed (Paux et al., 2008). Nevertheless many of the contig/scaffold sequences generated by the IWGSC CSS and the W7984 WGS assemblies have been genetically ordered using a population sequencing methods (POPSEQ). This involved the in-silico mapping of contigs from the CSS and W7984 assemblies in a recombinant inbred lines population (Chapman et al., 2015, Saintenac et al., 2013a).

The development of these genomic assemblies provided a reference template that has facilitated the identification of different gene models from the wheat genome (International Wheat Genome Sequencing Consortium (IWGSC), 2014, Krasileva et al., 2013). Using these gene models, it was possible to obtain a reduced representation of the wheat genome through exome capture technology. Along with RNA sequencing (RNAseq), the exome capture technology has been used to obtain gene coding sequence of many modern hexaploid wheat varieties, as well as, those of numerous landraces. This enabled the identification of unprecedented amount of Single Nucleotide Polymorphisms (SNPs) that are valuable as marker resources (Allen et al., 2011).

Consequently, numerous high-density genotyping arrays have been developed. These include the Infinium array with 9,000 SNP probes; the iSelect platform with 90,000 SNPs, of which about 47, 000 are genetically mapped SNP; and the Axiom platform with 820,000 SNPs, of which about 547,000 are genetically mapped SNPs (Cavanagh et al., 2013, Wang et al., 2014). With the development of these resources, availability of genetic markers will no longer remain a bottle-neck for genetic studies in wheat.

Besides the genomic and genetic resources described above, some other resources of great relevance to wheat research are currently in the development pipeline. This includes exome sequencing of a widely used mutagenised population for Targeting Induced Local Lesion in Genome (TILLING; Uauy et al., 2009). This resource, called in-Silico TILLING, will allow for the rapid identification of mutations in any gene of interest in a single day. Also, a wheat-specific gene expression database that integrates all the

currently available wheat RNAseq dataset is current being developed (Borrill et al, unpublished).

At the start of this Ph.D. project, only some of these resources were available. This includes the Chinese Spring WGS and the Infinium SNP platform. I also had pre-publication access to the IWGSC CSS. All the other resources were developed during the course of this Ph.D. project (Figure 1.7).

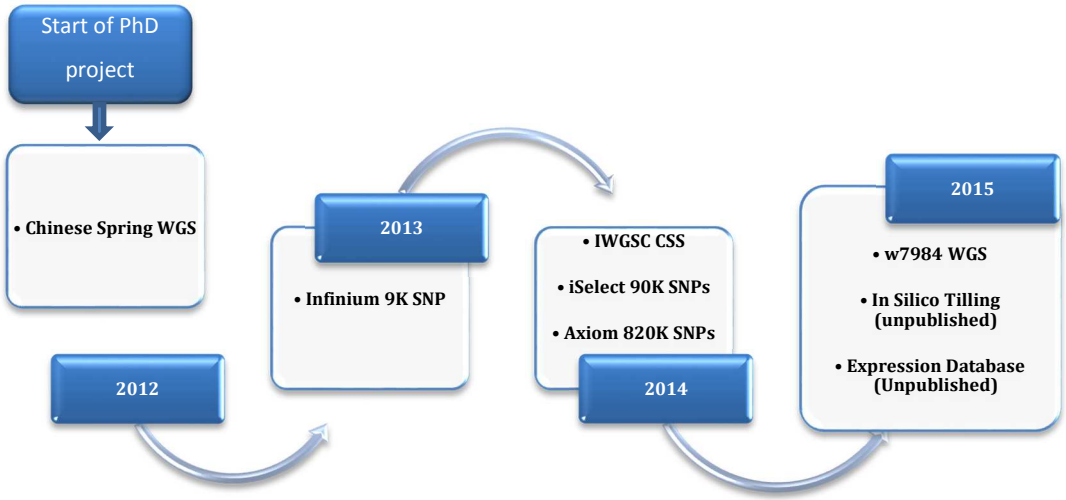


Figure 1.7: Timeline of development of wheat genomics, genetic and germplasm resources. The years highlighted refers to the year of publication.

1.10 Aims and objectives of this Project.

This study aims to elucidate the genetic, environmental and physiological aspects of major QTL controlling PHS and PMA in elite UK wheat varieties. The specific objectives of this research project are to:

1. Understand the physiology underpinning the effect of PHS and PMA resistance QTL in chosen elite UK wheat varieties. This will entails:
    a. understanding the effects of the QTL on the dynamics of dormancy loss.
    b. understanding the environmental (temperature) dependencies of each QTL effect.
2. Fine-mapping and positional cloning of one of the QTL characterised above and the development of tightly linked molecular markers to be used by breeders for selection for PHS/PMA resistance.

2 Chapter Two: Materials and Methods

2.1 Plant materials

2.1.1 Plant materials for field validation and the physiological and agronomic characterisation of PHS and PMA QTL

We developed and used isogenic F_3 and BC_3F_3 materials for field validation as well as for the physiological characterisation of selected PHS and PMA QTL. BC_1F_3 lines were used for the agronomic characterisation of the QTL effects. Prior to the development of these materials, DH lines were developed from crosses of resistant and susceptible parents listed in Table 2.1. Selected QTL were originally identified using these DH lines.

For each QTL, independent DH lines which carried the resistant parents alleles across the individual QTL intervals but which maximised the corresponding susceptible parents alleles in the background were crossed with their respective susceptible parents. F_1 lines produced from these crosses were either self-fertilised to produce F_2 plants or backcrossed "n" times with the susceptible parents before self-fertilisation to produce BC_nF_2 plants. At the F_2 and BC_nF_2 generations, independent F_2 and BC_nF_2 plants with the target QTL were selected using flanking markers. F_3 and BC_nF_3 plants were afterwards obtained from self-fertilisation of the selected F_2 and BC_nF_2 plants. Table 2.1 shows a list of the F_3 and BC_3F_3 lines used for the field validation experiment while Table 2.2 shows the number of BC_1F_3 lines used for the agronomic characterisation experiments.

2.1.2 Plant materials for fine-mapping, inheritance study and positional cloning

Two experimental populations were used for the initial and subsequent high-resolution fine-mapping of the 4A QTL. These include the Alchemy x Robigus population (AR pop.) and the Option x Claire population (OC pop.). For the initial fine mapping of the 4A QTL in the AR pop. (Chapter four, section 4.2.1.1), 14 BC_3F_3 NILs developed through the backcrossing scheme described in Figure 2.1 were used. However for the initial fine-mapping experiments in the OC pop., 131 F_4 Recombinant Inbreed Line (RILs) from 27 independent F_2 families developed through an independent but similar backcrossing strategy were used.

For the high-resolution fine-mapping experiments of the 4A QTL (Chapter five, section 5.2.2), a subset of the 131 OC F_4 RILs described above were used as well as BC_3F_2 AR RILs. In addition, F_5 RILs obtained from our breeding partners Limagrain were used. These Limagrain RILs were developed from 20 independent bi-parental or tri-parental crosses involving 31 UK wheat varieties. For the study of the mode of inheritance of the 4AL QTL, F_1 seeds were developed through reciprocal crosses between Claire and Option.

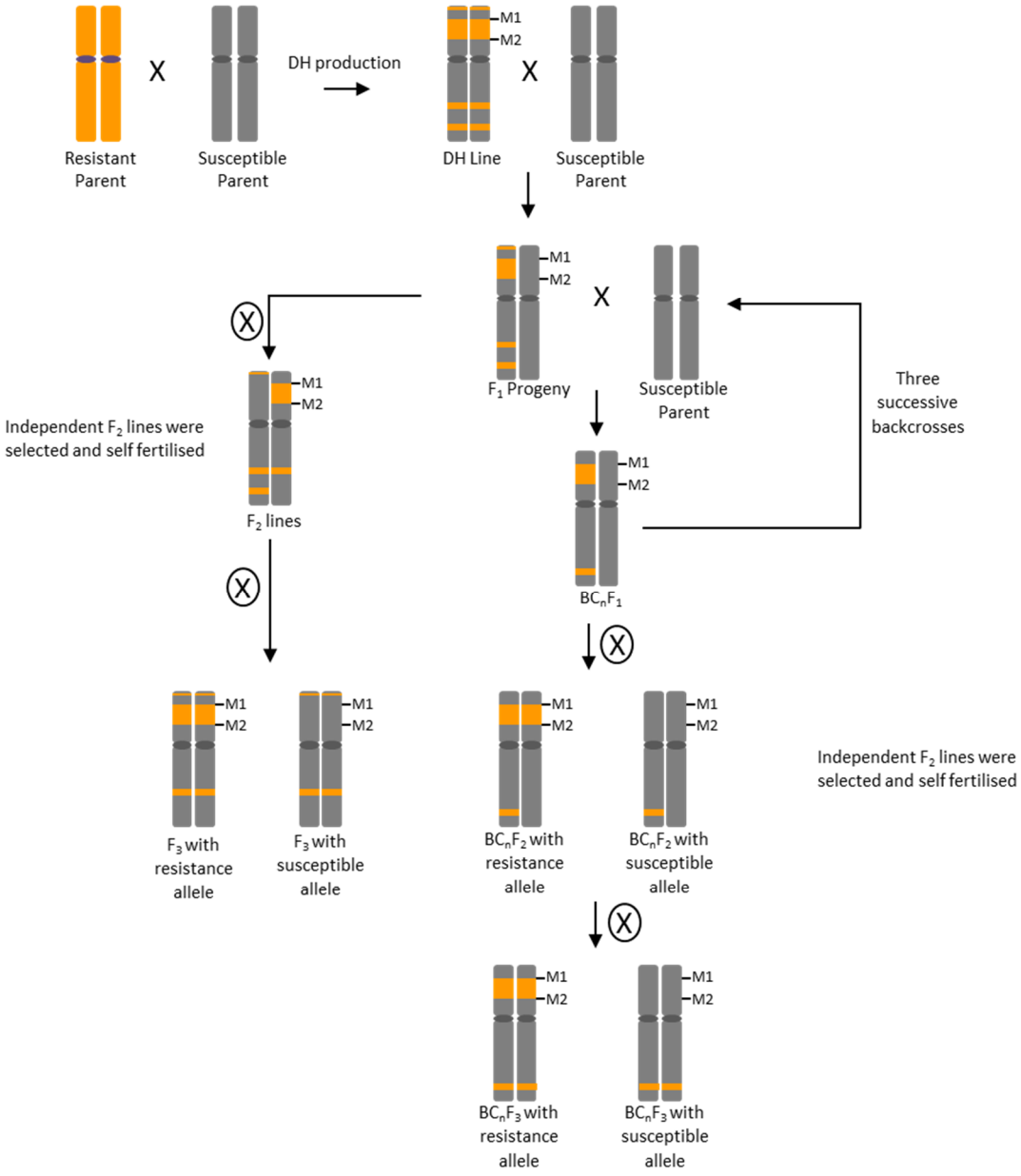


Figure 2.1: Development of doubled haploid (DH) lines and independent isogenic populations for individual QTL characterisation. DH lines were produced from the initial cross of resistant (amber coloured chromosome) and susceptible (grey coloured chromosome) parents. Selected DH lines were backcrossed recurrently with their respective susceptible parents. F_3 lines were produced by successive self-fertilisation of the progeny of the DH x susceptible parent cross, while BC_nF_3 lines were produced by successive backcrossing of the progeny to the susceptible parents. At DH, F_2 and BC_nF_2 generations, the QTL region of interest were selected with flanking markers (M1 and M2)

Table 2.1: F_3 and BC_3F_3 lines developed for field validation and physiological characterisation of the QTL effects

QTL	Alleles		F_3 Lines	BC_3F_3
1A	Susceptible	Haven	HS112-12A	HS112/3/34 HS112/3/25 HS112/15/50
	Resistant	Soleil	HS112-19B	HS112/3/79 HS112/9/66-5 HS112/9/8-5
2D	Susceptible	Avalon	AC51-7A AC514A*	AC73/10/6_1 NB69/20/30 NB69/20/30
	Resistant	Cadenza	AC51-5B	AC73/10/29 AC73/10/49 AC73/10/77
3A	Susceptible	Avalon		AC25/5/11/23_1 AC25/5/11/44 AC25/5/11/49_1
	Resistant	Cadenza		AC25/4/4/19_6 AC25/4/4/34 AC25/4/4/55_3
$3A_{SR}$	Susceptible	Rialto	RS37-1B	RS37/4/1 RS37/3/6_6 RS37/15/23_4
	Resistant	Savannah	RS37-3A	RS37/3/19 RS37/3/27_5 RS37/52/79
4A	Susceptible	Robigus		RRA33/1
	Resistant	Alchemy		RRA45/3
7B	Susceptible	Cadenza	AC150-39A AC150-5A*	AC150/5/18/42 AC150/5/18/44 AC150/5/18/50
	Resistant	Avalon	AC150-18B AC150-38B*	AC150/5/18/70 AC150/5/18/73 AC150/5/18/91

*These lines were only used for the field validation experiment and were not included in the physiological characterisation experiment

Table 2.2: Number of BC_1F_3 lines used for agronomic characterisation of the QTL

QTL	No. of susceptible BC_1F_3 lines	No. of resistant BC_1F_3 lines
2D	5	6
3A	6	7
$3A_{SR}$	5	5
7B	1	1

2.2 Growth conditions

The various germplasms used in this project including the F_3, NILs and parental materials, as well as the OC, AR and Limagrain population were grown either in the Controlled Environment Room (CER), Glass houses (GH) or in field sites across the UK. Table 2.3 gives a list of the experiment and the growth facility in which they were grown.

Materials grown in the CER or GH were first sown in 96 well trays with soil specification detailed in Borrill et al. (2015b). In brief: 15 % horticultural grit and 85 % fine peat; containing 2.7 kg m^{-3} Osmocote (3 - 4 months longevity), 0.5 kg m^{-3} wetting agent H_2Gro (Everris), 4 kg m^{-3} Maglime (Francis Flower) and 1 kg m^{-3} PG Mix fertilizer (www.yara.co.uk). They were then vernalised at 6 °C for 8 weeks. During vernalisation, leaf sample were harvested for DNA extraction from all the plants of each line, and this was used to genotyping with SSR or KASP markers located across the target QTL for each line.

After vernalisation, plants were transplanted into 1 L pots which contained 40 % medium grade peat, 40 % sterilized soil, 20 % horticultural grit; containing 1.3 kg m^{-3} PG Mix 14-16-18 (www.yara.co.uk), 1 kg m^{-3} Osmocote Exact Mini, 0.5 kg m^{-3} wetting agent H_2Gro, 3 kg m^{-3} Maglime and 300 g m^{-3} Exemptor (Bayer). The potted plants were afterwards grown in the CER or GH under the conditions detailed below.

2.2.1 CER conditions

Plants were grown in the CER under long day conditions: 16 h light (250 - 400 mmol) at 20 °C and 8 h darkness at 15 °C and at 70 % humidity. For the characterisation of the effect of grain development temperature on the QTL effects (Chapter three, section 3.2.4.5), some of the plants of each NIL were transferred from between 1 - 7 days after anthesis into a different CER with 16 h light (250 – 400 mmol) and 8 h darkness at 13 ˚C.

2.2.2 Glass houses

Plants were grown in the glass house under long day conditions: 16 h light (above 300 mmol) at a constant temperature (18 °C) and 8 h darkness at 15 °C. A relative humidity of 70 % was maintained.

Table 2.3: List of the of growth facilities used for each experiment

	Experiment	Chapter and section	Growth Facility
1	Field validation of the QTL' effect	Chapter three; section 3.2.3	Field
2	Physiological characterisation of QTL in NILs	Chapter three; section 3.2.4.3-6	Field
3	Physiological characterisation of QTL in F_3 lines	Chapter three; section 3.2.4.1-2	CER
4	Agronomic characterisation of each QTL effects	Chapter three; section 3.2.5	Field
5	Initial fine-mapping of the 4A QTL in AR NIL	Chapter four; section 4.2.1.1	GH
6	The 4A QTL mode of inheritance study	Chapter four; section 4.2.7.1	GH
7	Hormonal characterisation of the 4A QTL	Chapter four; section 4.2.7.2	CER
8	Sequence and expression analysis of the *PM19* genes	Chapter four: section 4.2.7.4	CER
9	Fine mapping experiment-1 with OC population	Chapter five; Section 5.2.2.1	GH
10	Fine mapping experiment-2 with OC population	Chapter five; Section 5.2.2.2	CER
11	Fine mapping experiment-3 with AR population	Chapter five; Section 5.2.2.3	GH
12	Fine mapping experiment-5 with Limagrain recombinant population	Chapter five; Section 5.2.3.3	GH
13	Fine mapping experiment-4 with Limagrain recombinant population	Chapter five; Section 5.2.3.2	Field

2.2.3 Field Trials

Plant materials used in three of the experiments described in this study were grown in different field sites in the UK (see Table 2.3). In addition to the field site at the John Innes Centre (JIC), Norwich UK (52.622269, 1.221398), we also used field sites of our breeding partners. These included RAGT seeds LTD (RAGT), Saffron Walden, UK (52.060236, 0.148997); KWS UK (KWS), Thriplow (52.097164, 0.105466) and Limagrain (LG), Woolpit, UK (52.225910, 0.878424). All these sites were used for the field validation experiment in a randomised complete block design with at least 4 replications. Similar design was used for the agronomic characterisation experiments, except that only one NIL per allele were used for the 7B QTL each with three plot replication. Also, only data from the RAGT, KWS and LG are reported for the agronomic characterisation experiments.

2.3 Genotyping

2.3.1 DNA Extraction

DNA extraction was done in a 96-well plate format according to the protocol developed by Pallotta et al. (2003). This is described below.

Reagents

- DNA extraction buffer
  Recipe for the DNA extraction buffer is a follow:
  - 100 mL 1.0 M Tris-HCL (pH 7.5)
  - 100 mL 0.5 M Ethylenediaminetetraacetic acid (EDTA, pH 8.0)
  - 125 mL 10 % Sodium Dodecyl Sulfate (SDS)
  - 675 mL deionised distilled water
- 6 M Ammonium Acetate
- Isopropanol
- 70 % ethanol
  Except otherwise stated, all the reagents were obtained from Sigma-Aldrich, UK.

DNA extraction protocol

1. The DNA extraction buffer was pre-heated to and stored at 65 °C
2. Young leaves (2 - 3 cm long) were harvested and placed in a 1.2 mL round collection tube (ABgene Storage Plate, ThermoFisher Scientific, UK; Cat No: AB-0564) containing one 3 mm tungsten carbide beads (Qiagen, Cat No: 69997) and freeze dried.
3. An aliquot of 500 µL of the DNA extraction buffer was added to each well of the plate.
4. The samples were then ground until fragmented plant material were suspended in the buffer and was afterward incubated at 65 °C for 45 - 60 mins. This was followed by cooling for 15 min at 4 °C
5. A 250 µL aliquot of cold 6 M Ammonium Acetate was added to each sample well, and the plate was stored at 4 °C for 15 mins. The plate was then centrifuged in a plate centrifuge for 15 mins at 5000 RCF to precipitate proteins and plant tissues.
6. An aliquot of 600 µL of the supernatant from each sample were transferred into a new storage plate, and 360 µL of isopropanol was added to each sample. The plate was incubated at room temperature for 5 mins and afterward centrifuged at 5000 RCF for 15 mins to precipitate the DNA.
7. The supernatant from Step 6 was poured away, and the precipitated DNA was washed by adding 500 µL of 70 % ethanol and centrifuging the plate at 5000 RCF for 15 mins.
8. Following centrifugation, the supernatant was discarded, and the DNA pellet was allowed to dry in the 65 °C incubator for 30 mins.
9. The DNA was then resuspended in 100 µL of water

2.3.2 Polymerase Chain Reaction (PCR)

All the PCR amplification in this study were done using Phusion® High-Fidelity DNA Polymerase (NEB, UK: Cat No.M0530S) or homemade recombinant *Taq* polymerase. The recipe for both polymerase mixes is presented in Table 2.4

Table 2.4: Reaction mixes for PCR using Phusion® or Recombinant Polymerase

PCR Recipe	Amount (Stock concentration)	
	Phusion®	Recombinant Taq
Template	2 µL (~50 ng/µL)	2 µL (~50 ng/µL)
Forward Primer	1.25 µL (10 µM)	1 µL (10 µM)
Reverse Primer	1.25 µL (10 µM)	1 µL (10 µM)
Buffer	5 µL (5 x Buffer)	5 µL (5 x Buffer)
dNTP	0.5 µL (10 mM)	0.5 µL (10 mM)
DMSO	0.75 µL	-
Polymerase	0.25 µL	0.25 µL
dH_2O	14 µL	15.25 µL

The GS4 G-Storm thermal cycler was used for the PCR reactions. The cycling programme for the PCR reactions is presented in Table 2.5

Table 2.5: Programme for Phusion® or Recombinant Polymerase PCR reactions

Steps		Temperature (time)	
		Phusion®	Recombinant Taq
Initial denaturation		98 °C (2 mins)	95 °C (2 mins)
	Denaturation	98 °C (10 sec)	95 °C (10 sec)
35 – 40 cycles	**Annealing**	58 °C - 64 °C* (30 sec)	58 °C - 64 °C (30 sec)
	Elongation	72 °C (30 sec- 2 mins)	72 °C (30 sec- 2 mins)
Final elongation		72 °C (10 mins)	72 °C (10 mins)
Store		4 °C	4 °C

*The NEB Tm calculator was used to determine the primer annealing temperature

2.3.3 Agarose Gel Electrophoresis

The size and purity of PCR amplicon or extracted RNA samples was assessed through agarose gel electrophoresis analysis. The samples were run on 1 % Agarose (Sigma-Aldrich, UK; Cat No: A9539) gel made with Tris-Acetate EDTA Buffer (TAE; Sigma-Aldrich, UK; Cat No: T9650) containing 1 in 10,000 dilution of ethidium bromide (EtBr; Sigma-Aldrich, UK; Cat No: E1510). Samples were mixed with an aliquot of 6 x loading

dye before being loaded into preformed wells of the agarose gel. Gel electrophoresis was performed at a voltage of 100 -120 V for 40 mins to 1 hour. The Gel images were visualised using UVP BioDoc-it® Transilluminator.

2.3.4 Gel Extraction and Purification

Following electrophoresis, PCR amplicon used for Sanger sequencing were purified using the Wizard® SV Gel and PCR Clean-Up System (Promega, UK; Cat No: A9281) following the manufacturer protocol with slight modifications. In brief, bands of target amplicon were excised from the gel and 10 µL of Membrane Binding Solution was added per 10 mg of the gel slice. The mixture was vortexed and incubated at 55 °C - 60 °C on a heating block until the gel slice was completely melted. The gel solution was transferred into a SV Minicolumn placed in a collection tube, and the tube was incubated for 1 min at room temperature. Following incubation, the tube was centrifuged at 13, 000 g for 1 min and the flow-through was discarded. An aliquot of 700 µL of Membrane Wash Solution was added to the column and centrifuged at 13,000 x g for 1 min, and the flow-through was discarded. Another 500 µL of Membrane Wash Solution was added to the column and was centrifuged for 5 mins with the flow-through discarded following centrifugation. Residual ethanol was evaporated by centrifuging the empty column/tube for 1 min. An aliquot of 25 - 30 µL of Nuclease-free water was added directly to the matrix at the base of the column, and this was left to incubate for 1 - 2 mins The tube was centrifuged at 13, 000 x g to elute the DNA

2.3.5 Sanger Sequencing

Sanger sequencing reactions were performed using the BigDye® Terminator v3.1 (Applied Biosystems). DNA used for these reactions was first quantified on the NanoDrop Spectrophotometer (Thermo Fischer Scientific). The recipe for BigDye®

Terminator reaction was as follow:

Component	**Amount**
BigDye® Terminator	1 µL
5 x BigDye® Buffer	2 µL
PCR product	X µL (25 – 100 µL)
Primer	0.5 µL
dH_20	Add up to 10 µL
Final Volume	**10 µL**

The Sanger sequencing reactions were performed on the GS4 G-Strom thermal cycler using the following cycling programme.

Step	**Temperature (time)**
Initial Denaturation	96 °C for 1 min
30 Cycles of	
Denaturation	96 °C for 10 sec
Annealing	50 °C for 5 secs
Extension	60 °C for 4 mins
Store	4 °C

The product of the Sanger sequencing reaction was sent to EurofinsDNA for analysis using the ABI sequencer.

2.3.6 KASPar Genotyping Assay

Kompetitive Allelic Specific PCR (KASPar) assay (Smith and Maughan, 2015) were developed for each SNP marker used in this study. Allelic-specific primers were made for each SNP allele with Primer3 (Untergasser et al., 2012, Koressaar and Remm, 2007). The primers were made in either the forward or reverse orientation with the SNP located at the 3’ end of the primer. Following primer design, the primers were directly labelled by the addition of tails sequences of the FAM and HEX fluorescence dye (See Appendix) probes to the 5’ end of the primer. A common primer was also developed for each assay, and this was designed in the opposite orientation to the allele-specific

primer. When possible, the common primer was made to be genome specific by the inclusion of an homoeologous SNP at the 3' end. The primers were mixed together in the following proportion:

Component	**Amount**
FAM labelled primer	12 µL
HEX labelled primer	12 µL
Common primer	30 µL
dH_2O	46 µL
Total volume	**100 µL**

The KASP assay was done in 384 well plates using the KASP V4.0 2x Master mix (LGC, KBS-1016-002) which contains the universal FRET cassettes, ROX™ passive reference dye, Taq polymerase, dNTP and MgCl2 in an optimised buffer solution. PCR reaction mix was prepared as follow:

Component	**Amount**
KASP V4.0 2x Master mix	2.5 µL
Primer mix	0.07 µL
DNA	2.5 µL
Total volume	**5.07 µL**

The PCR were performed using an Eppendorf® Mastercycler® Pro Thermal Cyclers with 384 well block with the following PCR cycling programme.

Steps	**Temperature (Time)**
Hotstart	95 °C for 15 min
1st PCR cycles (10 cycles)	
Denaturation	95 °C for 20 sec
Touchdown (-1 °C per cycle)	65 °C for 25 sec
2nd PCR cycles (30 cycles)	
Denaturation	95 °C for 10 sec
Annealing	57 °C for 60 sec
Store 16 °C forever	

Following PCR, fluorescent reading of the plate was done with the Tecan SAFIRE Fluorescent Scanner and genotype data analysis was done using the KlusterCaller™ software (LGC genomic)

2.3.7 SSR Genotyping Assay

For genotyping with SSR markers, the reverse primer sequences for each marker was labelled with tail sequences of the FAM (Blue), VIC (Green), NED (yellow) or PET (Red) fluorescent dye probes (Applied Biosystems). These allowed for the multiplexing of samples from different genotyping assay after PCR amplification. For each assay, the Qiagen Hotstart Master Mix (Qiagen, Cat No: 203443) was used for the PCR reaction. Prior to the PCR reaction, a primer mix was prepared as follow:

Component	**Amount**
Dye sequences	18.75 µL
Forward primer	1.25 µL
Reverse Primer	18.75 µL
dH_2O	211.25 µL
Total volume	**250 µL**

The PCR reaction mix was then prepared in a 96 well PCR plate (Thermofisher Scientific, Cat No: AB-0700) as follows:

Component	**Amount**
Qiagen Hotstart Master Mix	3.125 µL
Primer mix	0.625 µL
DNA	2.5 µL
Total volume	**6.25 µL**

The SSR PCR reactions were performed on the GS4 G-Strom thermal Cycler using the following protocol.

Steps	Temperature (Time)
Hotstart	95 °C for 15 mins
35 Cycles of	
Denaturation	95 °C for 1 min
Annealing	52 °C for 1 min
Extension	72 °C for 1 min
Final Extension	72 °C for 10 mins
Store	4 °C

Following PCR amplification, samples were diluted (1/40) and 1 µL of the diluted sample was added to the 0.1 µL of GeneScan 500 LIZ dye size standard (Thermo Fisher Scientific; Cat. No: 4322682) and 8.9 µL of formamide. Alternatively, samples labelled with different dyes were multiplexed by adding 1 µL of each sample dilution to 0.1 µL of LIZ size standard and formamide was added to a final volume of 10 µL. The samples were then run on Applied Biosystems 3730 DNA Analyzer and genotype data were afterwards analysed on the GeneScan® Analysis Software (Applied Biosystems).

2.4 Gene Expression Analysis

2.4.1 RNA Extraction

The RNA extraction protocol used is a modification of the high-throughput phenol-based method reported by Box et al. (2011). This is described below.

Reagents

- RNA Extraction Buffer (RE buffer; 0.1 M Tris pH 8.0, 5 mM EDTA pH 8.0, 0.1 M NaCl, 0.5 % SDS).
- Acidified phenol pH 4.3 ± 0.2.
- Chloroform.
- Isopropanol (2-propanol).
- 3 M sodium acetate (pH 5.2).
- 70 % ethanol.
- Nuclease-free water.

Except otherwise stated, all the reagents were obtained from Sigma-Aldrich, UK. Also, note that 1 % of 2-mercaptoethanol was added to the RE buffer just before use.

Extraction Protocol

1. About 5 - 10 grains were ground with mortar and pestle using liquid nitrogen, and the ground material was then transferred into a 1.5 mL eppendorf tube.
2. An aliquot of 500 µL of RE Buffer was added to the samples, and this was mixed thoroughly.
3. (Optional)
   A. To remove common contaminant like polysaccharides which are plentiful in grains, an aliquot of 100 - 200 µL of Ambion® Plant RNA Isolation Aid (Thermofisher Scientific, UK; Cat No AM9690) was added to each sample and mixed
   B. The sample was then centrifuged for 10 mins at 13,000 x g, and the supernatant was transferred to a fresh 1.5 mL tube.
4. An aliquot of 300 µL of 1:1 acidic Phenol (pH 4.3)/ Chloroform was added to the sample, and this was mixed thoroughly for 10 mins. The mixture was then centrifuged for 15 mins at 13, 000 x g to separate the aqueous phase from the organic phase.
5. The supernatant was removed into a fresh 1.5 mL tube, and 240 µL of Isopropanol, as well as, 30 µL of 3 M Na Acetate (PH 5.2) were added. The sample was then left at -80 °C for 15 mins to precipitate the nucleic acid.
6. The samples were then centrifuged for 30 mins at 13,000 x g and the supernatant was discarded. Care was taken not to lose the RNA pellet while decanting the supernatant.
7. The RNA pellet was washed twice with 750 µL of 70 % ethanol with centrifugation between washes at 13,000 x g for 5 mins.
8. The supernatant was discarded during each washing step, and residual ethanol was removed after washing
9. The RNA pellet was allowed to dry briefly before being resuspended in 100 µL of RNAse Free water.

2.4.2 DNAse Treatment

Extracted RNA was treated with RQ1 RNase-Free DNase (Promega, UK) to remove genomic DNA contamination. The reaction mix for the DNAse treatment is presented below

Component	Amount
RNA in water	X µL (2 µg RNA)
RQ1 10 X Reaction Buffer	1 µL
RQ1 RNase-Free DNase	2 µL
Nuclease-free water	Add to a final total volume of 10 µL
Final Volume	**10 µL**

The reaction mix was incubated at 37 °C for 30 mins. The reaction was stopped by adding 1 µL of RQ1 DNAse Stop Solution and subsequent incubation at 65 °C.

2.4.3 cDNA Synthesis

cDNA synthesis was performed using the GoScript Reverse Transcriptase (Promega, Cat No: A5000) following the manufacturer's protocol. Before the reverse transcription reaction, the RNA and primer mix were first denatured by incubation at 70 °C for 5 mins. The recipe for this RNA-primer denaturation mix is presented below

Component	Amount
Experimental RNA	X µL (1 - 2 µg)
Primer	
Oligo$(dT)_{15}$ (0.5µg/reaction)	0.5 µL
Random Primer (0.5µg/reaction)	0.5 µL
Nuclease-Free Water	to a final volume of 5 µL
Final Volume	**5 µL**

Following denaturation, the samples were cooled on ice for 5 mins and briefly centrifuged to collect condensate. A reverse transcription reaction mix (listed below) was prepared

Component	Amount
Nuclease-Free Water	7 µL
GoScript™ 5 x Reaction Buffer	4 µL
$MgCl_2$	2 µL
dNTP	1 µL
GoScript™ Reverse Transcriptase	1 µL
Denatured RNA/Primer mix	5 µL
Final volume	**15 µL**

The reverse transcription reaction was performed on the GS4 G-Strom thermal Cycler using the following cycling programme.

Step	Temperature (Time)
Annealing	25 °C (5 min)
Extension	42 °C (60 mins)
Termination	70 °C (15 mins)
Store	4 °C.

The synthesised cDNA was stored in -20^{o}C until use.

2.4.4 Quantitative PCR (q-PCR)

LightCycler® 480 SYBR Green I Master (Roche; Cat No: 04707516001) was used for the q-PCR reactions. The reactions were carried out in white 384 well plate (Roche, Cat No. 04729749001). The reaction mix for the PCR is presented below.

Component	Amount
SYBR Green Master mix (2 x Conc)	5 µL
cDNA (1 in 10 dilution)	2 µL
Forward Primer (10 µM)	1 µL
Reverse Primer (10 µM)	1 µL
Nuclease-Free Water	1 µL
Final volume	**10 µL**

The LightCycler® 480 PCR machine (Roche) was used for the reaction with the following PCR protocol:

Steps	**Temperature (Time)**
Pre-incubation	95 °C (5 mins)
Amplification (45 cycles)	95 °C (10 sec)
	60 °C (15 sec)
	72 °C (20 sec)
	78 °C (0.01 sec)
Cooling	60 °C (0.01 sec)
Melting curve	95 °C (5 acquisition per second)

The efficiency of the primer used for qPCR was first assessed using the method described by Schmittgen and Livak (2008). Only primers with amplification efficiency between the 95 - 110 % and with highly specific amplicon as verified from the melting curve were used. The wheat *Translation Elongation Factor-1A* gene (*TaEF-1A*) was used as reference gene and the expression of target genes are presented relative to the expression of *TaEF-1A*.

2.5 Physical Map Sequence Assembly and Annotation

2.5.1 BAC Extraction

Twenty-four bacteria clones containing Bacterial Artificial Chromosome (BAC) that constitute the Minimum Tilling Path (MTP) of the three main BAC clusters (cluster 16421, 285 and 7335) found in the 4AL PHS QTL region were obtained from Prof. Jaroslav Dolezel's lab at the Centre for Plant Structural and Functional Genomic, Institute of Experimental Botany (IEB), Czech Republic. BACs were extracted from the bacteria clones using the Qiagen Plasmid Midi Kit (Qiagen, Cat. No. 12143) following the manufacturer's protocol described below.

Reagents

- Qiagen Buffer P1 with RNase A solution added and stored at 2 - 8 °C .
- Qiagen Buffer P2.
- Qiagen Buffer P3 (pre-chilled at 4 °C)
- Qiagen Buffer QBT
- Buffer QC
- Isopropanol
- 70 % ethanol.

Extraction Protocol

1. A 100 mL LB culture of each bacterial clone was prepared overnight and was harvested by centrifugation in a RC5C Sorvall Centrifuge at 6000 x g for 15 mins at 4 °C using 250 mL centrifugation bottle.
2. The bacterial pellets were re-suspended in 4 mL of Buffer P1.
3. An aliquot of 4 mL of Buffer P2 was added to the sample and mixed thoroughly by vigorously inverting the bottle 6 times to lyse the cells. Cell lysis was not allowed to progress for more than 3 mins.
4. An aliquot of 4 mL of pre-chilled Buffer P3 was added to the sample, mixed thoroughly by vigorously inverting the bottle 6 times and stored on ice for 5 mins.
5. The sample was then centrifuged at 20,000 x g for 30 min at 4 °C. The supernatant containing the BAC plasmids was transferred to a new centrifugation bottle, and this was centrifuged again at 20,000 x g for 15 mins at 4 °C.
6. While step 5 was being performed, a Qiagen-tip 100 column was equilibrated by the addition of 4 mL of the Buffer QBT and the column was allowed to empty by gravity.
7. The supernatant from step 5 was added to the column and allowed to enter the resin of the column by gravity.
8. The Qiagen-tip column was washed twice by adding 10 mL of Buffer QC, allowing the column to empty by gravity after each wash.
9. BAC DNA was eluted by the addition of 5 mL of Buffer QF (pre-incubated at 65 °C) and allowing the column to empty into 15 mL tubes.
10. The BAC DNA was precipitated by the addition of 3.5 mL of isopropanol to the eluted DNA and was centrifuged at 15,000 x g for 30 mins.
11. The DNA pellet was washed with 2 mL of 70 % ethanol and centrifuged at 15,000 x g for 10 mins, and the supernatant was carefully discarded.
12. The DNA pellet was air-dried and re-suspended in 100 of distilled water.

2.5.2 BAC End Sequencing (BES)

The BES protocol is similar to the Sanger sequencing protocol described previously. However, due to the high concentration (DNA molecule/µL) of the BAC solution, no prior PCR amplification was required before the sequencing reaction was performed. The BAC sequences are contained in a pIndigoBAC-5 vector which is flanked by T7 RNA polymerase and M13 reverse promoter sites. Primer sequences for these promoters were used to sequence into the BAC. The sequencing protocol is described below.

Steps	**Temperature (time)**
Initial Denaturation	94 °C for 1 min
80 Cycles of	
Denaturation	94 °C for 10 sec
Annealing	55 °C for 5 secs
Extension	60 °C for 4 mins
Store	4 °C

The product of the Sanger sequencing reaction was sent to EurofinsDNA for analysis using the ABI sequencer.

2.5.3 BAC Sequencing

High-quality BACs DNA were supplied to The Genome Analysis Centre (TGAC, Norwich, UK) for sequencing. TGAC prepared the NGS library with an average library insert sizes of 360, 460 and 440 bps for cluster 16421, 285 and 7335 respectively. BACs were multiplexed for sequencing on Illumina Miseq lanes and overlapping 250 bp paired-end reads were obtained for each BAC.

2.5.4 BAC Assembly

The paired-end reads obtained from Illumina sequencing were assembled using the CLC Bio genomic software (www.clcbio.com). Before assembly, reads were first filtered to remove contaminant sequences by mapping to the pIndigoBAC-5 vector and the *Escherichia coli* genome. De novo assembly of reads that do not map to these

contaminant reference sequences was then done. Parameters for de novo assembly include:

Word size: 64 bp
Bubble size: 250 bp
Mismatch cost: 2
Insertion cost: 3
Deletion cost 3
Length fraction: 90 %
Similarity fraction: 95 %.

2.5.5 Transposable Element Annotation and Masking

Transposable element sequences in the assembled BACs contigs and the Barley BACs scaffolds were searched for by BLAST analysis of the assembly contig sequences against the Triticeae Repeat Database (TREP) accessible at wheat.pw.usda.gov/ITMI/Repeats. Repeat sequences identified were annotated and masked to distinguish them from non–repeat sequence.

2.5.6 Gene Annotation

Gene models in repeat-masked BAC sequence were searched for by BLAST analysis of the BAC sequence against a custom database containing wheat gene models described by (Krasileva et al., 2013) and by BLASTX analysis on NBCI (blast.ncbi.nlm.nih.gov/Blast). Gene models were also obtained by ab-initio gene prediction with FGENESH (Solovyev et al., 2006). Only FGENESH gene model with protein sequence support from NCBI or Ensembl Plant database (plants.ensembl.org) were used. Gene model with greater than 90 % protein or nucleotide sequence identity and more than 75 % sequence coverage to already annotated genes on the NCBI or Ensembl databases were considered as high confidence genes. While gene models that do not meet these criteria were considered as low confidence gene. The annotation of Barley BAC scaffolds was done by BLAST analysis of the scaffold sequences against the high confidence gene model (IBGSC, 2012) database of the

barley genome hosted at IPK Barley BLAST Server (webblast.ipk-gatersleben.de/barley/viroblast).

2.6 Phenotyping

2.6.1 Germination Index test/ Hormone Treatments

To ensure that lines were compared at similar stages of grain development/maturation, spikes of the different lines used (see Table 2.1) were tagged at anthesis and harvested at similar day post anthesis. For samples grown in the 13 °C CER after anthesis, sampling time was determined by visual observation of the grain development stage and by equivalent thermal time to samples grown in 20 °C CER. Thermal time was calculated as average daily temperature x days after anthesis.

Harvested spikes were gently threshed to obtain grains mainly from the middle part of the spike. Twenty grains were placed with crease facing down in 90 mm petri dishes containing two layers of Sartorius filter paper, and 5 mL of sterile water was added to each plate. For experiments with hormone treatment, 5 mL of 100 µM and 500 µM of Amino Cyclopropane Carboxylic acid (ACC), 50 µM and 250 µM of Gibberellic Acid (GA); 50 µM and 250 µM of Fluridone and 50 µM of Abscisic Acid (ABA) were added instead of water. All the chemicals used were obtained from Sigma-Aldrich, UK. For the experiment on the effect of incubation temperature on germination, plates were placed in a box and kept at 10 °C, 17 °C or 22 °C for seven days. During each day of incubation, germinated seeds (with ruptured seed coat) were counted and removed from the plates. The number of germinated seed in each day was used to calculate the Germination Index (GI) as described by Walker-Simmons (1987) using the formula:

$$\boldsymbol{GI} = \frac{(\mathbf{7}\boldsymbol{x}\,\boldsymbol{n}\mathbf{1}) + (\mathbf{6}\,\boldsymbol{x}\,\boldsymbol{n}\mathbf{2}) + (\mathbf{5}\,\boldsymbol{x}\,\boldsymbol{n}\mathbf{3}) + (\mathbf{4}\,\boldsymbol{x}\,\boldsymbol{n}\mathbf{4}) + (\mathbf{3}\,\boldsymbol{x}\,\boldsymbol{n}\mathbf{5}) + (\mathbf{2}\,\boldsymbol{x}\,\boldsymbol{n}\mathbf{6}) + \boldsymbol{n}\mathbf{7})}{\mathbf{7}\,\boldsymbol{x}\,(\boldsymbol{N} - \boldsymbol{M})}$$

Where n1, n2, … n7 are the number of grains that germinated on the first, second, and n^{th} days until the 7th day, respectively; N is the total number of grains per plate and M is the number of mouldy grains after the 7 days of incubation.

2.6.2 Agronomic Characterisation of QTL

Agronomic traits were measured on field grown material in different locations. Days to ear emergence was adjudged by the number of days it took for half of the plants in a 1 m^2 plot to show emergence of at least half portion of the ear (spike) from the enclosure of the leaf sheath. The height was measured after physiological maturity and was an estimate of the average height of plant in each plot. Grain morphometric measurements (TGW, Area, Width and Length) were obtained using the Marvin seed analyser (GTA Sensorik GmbH).

2.6.3 Hagberg Falling Number (HFN) Test

Harvest-ripe spikes were harvested from the field and threshed to obtain seeds. About 10 - 15 g of seeds from each sample was ground in a cyclone laboratory mill (UDY Corporation). The flour produced from each sample was allowed to equilibrate to constant moisture content at 30 °C. The moisture content of the flour was determined by drying a representative sub-sample in a 65 °C incubator and obtaining the weight before and after drying. Based on the moisture content of the flour, an amount of flour equivalent to 7 g of 14 % moisture content was used. The HFN value of the flour was measured with a Falling Number 1900 machine (Perten Instrument, Sweden). To do this, the flour was mixed with 25 mL of water in a viscometer tube, and this was shoken vigorously to homogenise the suspension. The tube, with a Falling number stirrer inserted, was placed in a hot water bath of the Falling Number machine. After an automatic stirring for 60 sec, the stirrer was released from its top position. The time it took for stirring (60 sec) and for the stirrer to fall to the bottom of the tube under gravity was recorded as the HFN value. More information can be obtained from www.perten.com/Products/Falling-Number/The-Falling-Number-Method.

2.6.4 Sprouting Test

Spikes from each line, harvested from similar time post anthesis, were allowed to after-ripening at room temperature. During this period of after-ripening, the germination statuses of the parents were monitored weekly via the GI test. When GI difference was observed between the parental samples, after-ripened spikes (two – three spikes from each plant) were arranged on metallic racks. The spikes were then misted for 5 - 7 days in a sprouting chamber containing a humidifier and a revolving platform. Misted spikes were dried and gently threshed together to collect the seeds. The seeds were examined for the symptoms of sprouting (breakage of the seed coat near the embryo), and the number of sprouted seeds in each biological replication was used to calculate the percentage of sprouting. This was then used to calculate weighted percentage sprouting averages and standard error of the mean.

2.6.5 Amylase activity assay

α-Amylase activity in grains was determined using the Ceralpha α-amylase assay kit (Megazyme). Samples of 0.25 g of grains were ground in mortar and pestle with Liquid nitrogen. Amylase extract was obtained by addition of 1.5 mL of Ceralpha Extraction Buffer and incubated at room temperature for 40 mins. Following centrifugation, the activity of amylase was assayed according to the manufacturer manual. The Ceralpha procedure is based on the endo-acting activity of α-amylase to hydrolyse a non-reducing-end blocked p-nitrophenyl maltoheptaoside (BPNPG7) in the presence of excess amylogucosidase and α-glucosidase. Following hydrolysis, the p-nitrophenyl maltosaccharide fragments are made accessible to the activity of amylogucosidase and α-glucosidase and these produce free glucose and p-nitrophenol. The addition of Trizma base solution stops the reaction and produces a colour change. The absorbance of the product is measured at 400 nM, and this corresponds to the level of α-amylase in the original sample.

2.7 Statistical Analyses

Statistical analyses were performed in Genstat (version15.2.0.8821) and Minitab (Version, 17.2.1). Statistical significance between pairs of lines or treatments was calculated in Genstat using the t-test at 95 % confidence interval. For evaluation of the genetic x environmental interaction on the effect of each QTL on agronomic traits, a Two-way Analysis of Variance (ANOVA) test was conducted in Genstat using both experimental locations and individual QTL alleles as factors. For multiple comparisons between the sprouting value of recombinant lines and parents, ANOVA analyses were done on Minitab with ad-hoc Dunnet multiple comparison test (when parental controls were included in the experiments) or the Tukey's Honestly Significant Difference(HSD) test (Tukey, 1949) (when there no parental control in the experiment. Percentages data were first Logit transformed to correct for unequal variance before being used for the ANOVA analysis.

3 Chapter Three: Characterisation of PHS and PMA QTL identified in UK Wheat Varieties

Chapter Abstract

This chapter introduces 6 QTL located on chromosomes 1A, 2D, 3A (2 loci), 4A and 7B, which confers resistance to PHS and PMA in UK wheat varieties. We validated the effects of some of these QTL in independent isogenic materials. In addition, physiological characterisation of these QTL showed that they exert their effects by affecting the dynamic of dormancy loss in grains, albeit at different stages of grain development and maturation. We also show that temperature during grain development and germination affect the expression of these QTL effects. Finally, we show that introgression of some of these QTL showed effects some agronomic traits like flowering time and plant height but did not have any consistent negative effect on yield.

3.1 Introduction

3.1.1 PHS and PMA are major threats to UK wheat production

Wheat is the most important crop cultivated in the UK, accounting for 43 % of the total arable area on commercial holding in the last three years (DEFRA, 2015). During the 2010 to 2014 growing season, an average of 14.4 million tonnes of wheat was produced per annum in the UK (DEFRA, 2014). Despite the UK being a relatively small island, It is the 14th largest producer of wheat in the world (FAO, 2014). This is mainly due to its high average yield of 7.4 tonnes/ha (2010-2013).

Wheat produced in the UK are either processed into flour for making bread, biscuits, cakes, pastries and other flour based confectioneries. It could also be used as animal feed, malting/brewing raw material or for commodity export. Based on the processability and the end-use characteristic of their grains, wheat varieties in the UK are classified into four groups using the National Association of British and Irish Millers (NABIM) classification. NABIM Group 1 comprises varieties with consistent milling and

baking qualities. Group 2 varieties show less milling potential but are still suitable for bread-making. Group 3 varieties are mainly used for biscuit, cakes and pastries while Group 4 varieties are grown mainly as feed for livestock.

The bread-making varieties (NABIM Group 1 and 2) command a higher premium than the biscuit and feed wheat varieties. For instance, the 2015 spot and forward prices of bread wheat is an average of 20 % higher than feed wheat (AHDB, 2015b). However, the yield of bread-making varieties is often lower than those of biscuit and feed wheat varieties. Also, farmers expend more cost in meeting up with the high quality requirements (protein content, specific weight and HFN value) of bread-making varieties. Some of these costs can include higher fertiliser input and/or early harvesting and artificial drying of grains to avoid sprouting damages. These measures are not environmental sustainable (Pretty et al., 2005) and do not always guarantee the production of quality grains. As a result, bread-making wheat varieties can sometimes end up being sold as feed wheat for less than their premium price. To avert this risk, there has been a decline in recent times in the cultivation of some bread-making varieties (particularly the Group 2), while the cultivation of feed wheat varieties, which are high yielding and less cost intensive, has increased (Figure 3.1).

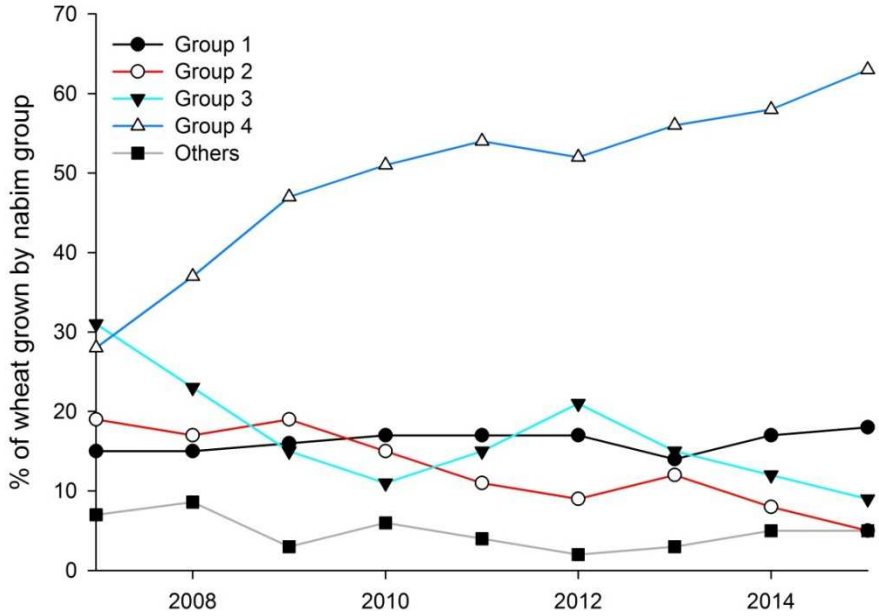


Figure 3.1: The percentages of wheat grown in the UK from 2007 to 2015 by NABIM group. The 'Others' categories refers to wheat varieties grown in the UK that are not classified by the NABIM system.(AHDB, 2015a)

The problem of low Hagberg Falling Number (HFN) caused by the incidences of Pre-harvest sprouting (PHS) and Pre-maturity amylase (PMA, described in Chapter One, 1.8) is one of the factors responsible for this decline in the cultivation of bread-making wheat varieties in the UK. The relatively wet UK weather conditions can sometimes predispose cultivated wheat to the problem of PHS and PMA. With an average summer rainfall of 85 mm (from 2000 – 2014, Met Office, 2015) during the critical stages of grain development and maturation (May-August), the UK wheat fields can be an ideal environment for PHS and PMA. However, the extent of the challenges posed by these traits varies from year to year and also from one region to the other in the UK. This is in part reflective of the yearly and regional weather variation in the UK (Figure 3.2)

For instance, in 2004, 2007 and 2012, more than 50 % of the NABIM Group 1 wheat varieties failed to meet the 250 sec HFN threshold for bread-making quality (Figure 3.3). This was particularly due to the highly conducive environmental conditions (rainfall) for the induction of PHS and PMA in these years (Figure 3.2). In contrast, in 2006 and 2013, no wheat was rejected due to this problem. On average, over the last 14 years, more than 25 % of bread-making NABIM group 1 wheat varieties did not meet the acceptable HFN threshold (AHDB, 2014). Farmers are as such forced to sell such grains as feeds. This has consequences not only on farm profitability but also on global food security. Thus, PHS and PMA constitute a major economic threat to wheat production in the UK.

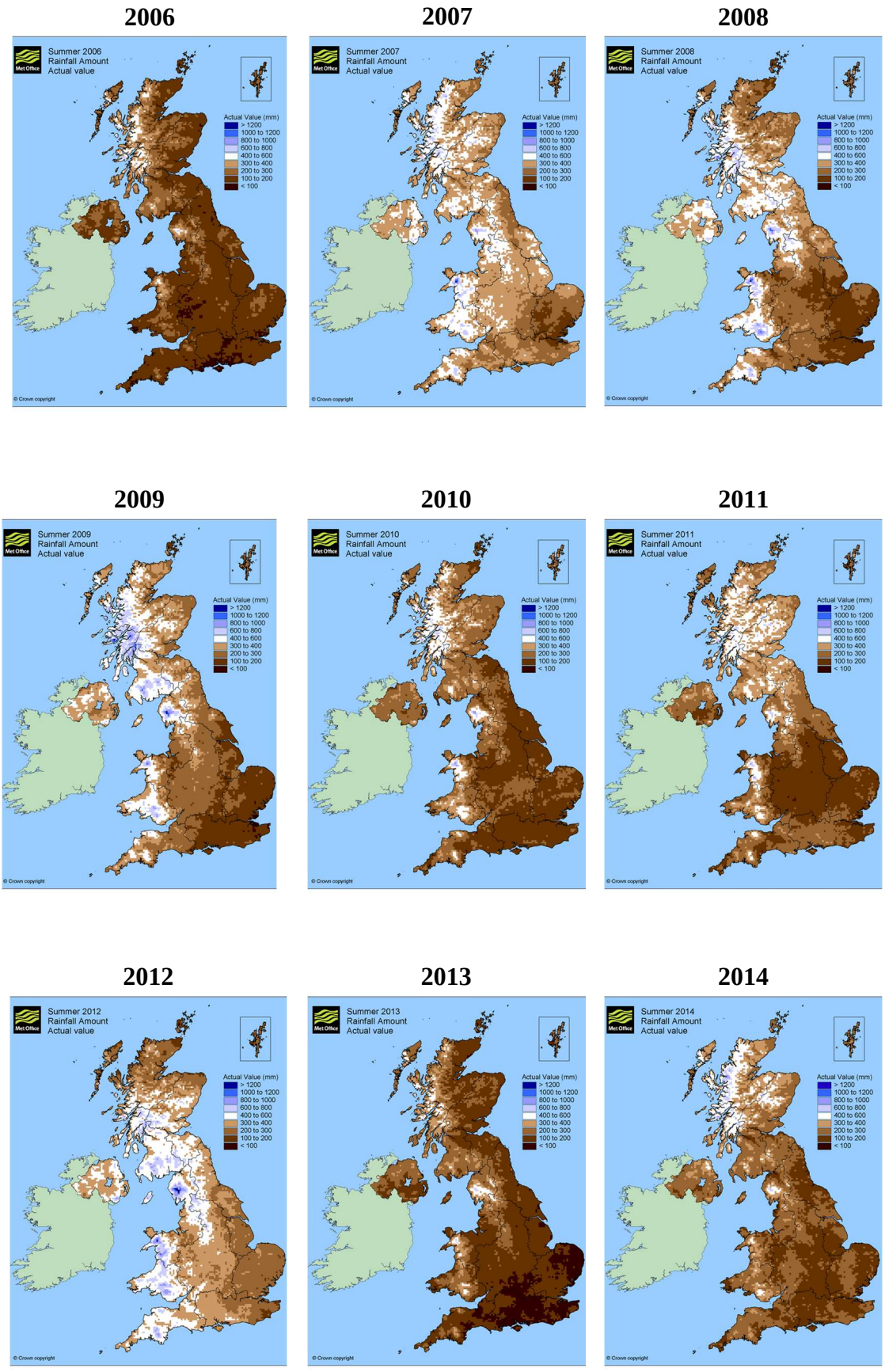


Figure 3.2: The UK climate is conducive for PHS and PMA incidences. The amount of rainfall in the UK during the summer months (May –August) from 2006 to 2014. The amount of rain is indicated by the colour scale besides each map, with increasing brown shade indicating less rainfall and increasing blue shade indicating more rainfall (Met Office, 2015).

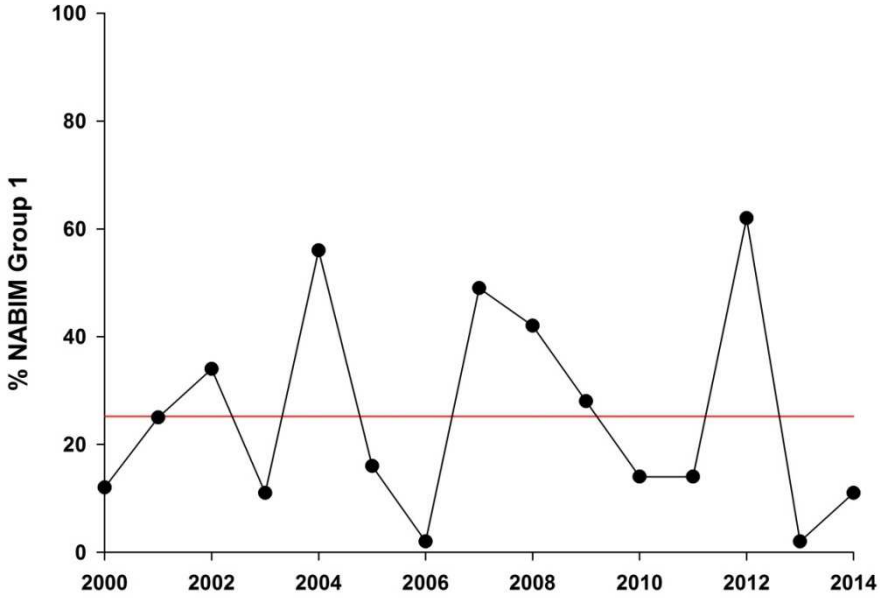


Figure 3.3: Incidences of low HFN in the UK. Percentages of NABIM group 1 varieties with HFN less than 250 sec from 2000 to 2014. The red line indicates mean percentage across the 15 year period.

3.1.2 Genetic control of PHS and PMA in the UK

Given the significance of the threats posed by PHS and PMA to UK wheat production, there have been significant research efforts aimed at stabilizing the HFN of UK bread-making wheat varieties. This have been mainly through identifying and breeding for resistance against PHS and PMA. An example of this is a Defra-LINK project titled 'An Integrated Approach to Stabilising HFN in Wheat: Screens, Genes & Understanding'. One of the outputs of this project was the identification of the genetic architecture of PHS and PMA resistance using 11 doubled haploid (DH) populations originating from 16 elite UK wheat varieties (Flintham et al., unpublished). The PHS and PMA resistances in these DH lines were assessed through a simulated sprouting test and a HFN test in field trials conducted in 2005 - 2008. The sprouting test done involved the use of overhead irrigation to induce sprouting in field grown materials. The Hagberg test, on the other hand, is a viscometric test that indirectly measures amylase activity in flours slurry and is suitable for detecting PMA, but can also detect the production of amylase that accompanies sprouting.

The field trial data, along with genetic linkage maps of the populations, enabled the identification of QTL for PHS and PMA resistance on almost all of the chromosomes of wheat (Figure 3.4). This implies an abundance of genetic variation that could be harnessed for PHS and PMA resistance breeding in the UK. However, many of these QTL were only detected once, possibly due to strong environmental influence on the control of such QTL. However, other QTL were detected in more than one trial, and still others showed consistent effects in all the trials. PHS QTL that were identified in more than one trial include QTL on wheat chromosomes 1A, 2D, 3A (two loci), 3B, 3D, 4A (2 loci), 4D, 5D (2 loci) and 7B. QTL for PMA resistances identified in at least two trials were located on chromosomes 1B, 4D, 5D, 6A and 7B (Flintham et al., unpublished).

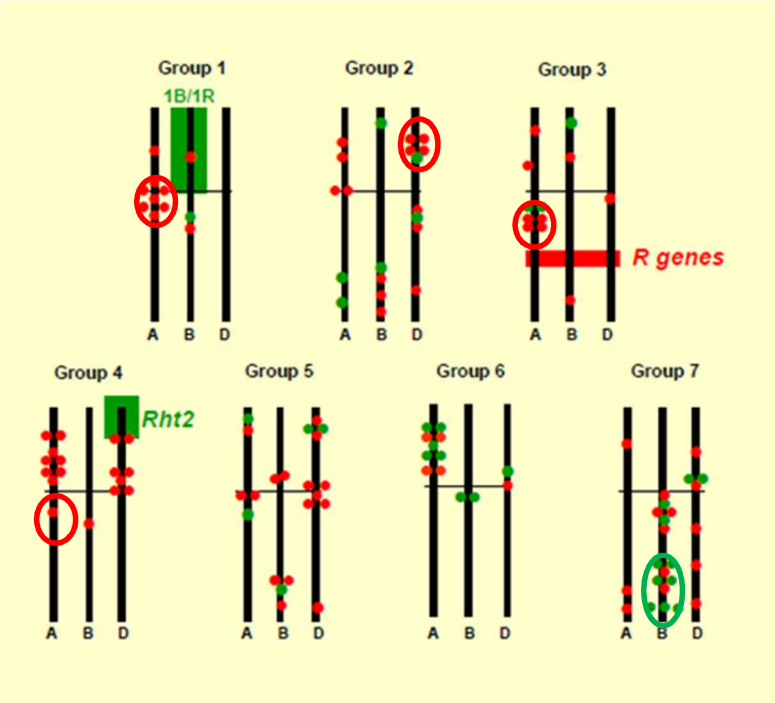


Figure 3.4: Distribution of PHS and PMA QTL identified in the HFN DEFRA LINK project. The dots represent PHS (red) or PMA (green) QTL in the different populations used for the study, therefore clustered dots may represent co-localising QTL. Green and red boxes represent location of major chromosome location known to have effects on PMA and PHS respectively. Note that the *R* genes refer to the *R-1* genes. Approximate location of the QTL selected for this present study are indicated by red (PHS QTL) or green (PMA QTL) spheres. The centromere is represented by the horizontal line across each

Of these QTL, six showed stable and consistent effects over different years and environments (Table 3.1). The 1A, 2D, 3A (two loci) QTL showed consistent effects on sprouting in all four years of field trial. The 4AL QTL effect on sprouting was observed in only two years, but it gave the highest sprouting effect observed (53 % change in sprouting between alleles) in any of the four years. Analogously, the 7B QTL consistently showed the biggest effect in the HFN test (in 2007 and 2008), but showed less pronounced effect on sprouting across the four years indicating that this QTL is primarily a PMA resistance QTL. A summary of the QTL effects and the associated genetic markers are shown in Table 3.1.

Table 3.1: Mean effects of prioritised QTL on PHS and PMA in UK wheat varieties.

QTL	Population	Beneficial Allele	Marker with highest LOD score	Change in % Sprouting				% Change in HFN		
				2005	2006	2007	2008	2006	2007	2008
1A	Haven x Soleil	Soleil	*wmc333*	26	20	26	16	19	45	31
2D	Avalon x Cadenza	Cadenza	*gwm261*	27	25	15	33	-	54	15
3A	Avalon x Cadenza	Cadenza	*wPT-1562*	29	24	19	23	-	-12	-3
3A	Savannah x Rialto	Savannah	*wmc264*	-	-	24	-	56	45	43
4A*	Alchemy x Robigus	Alchemy		-	-	23	53	-	-	-
7B	Avalon x Cadenza	Avalon	*wPT-4814*	27	4	6	24	-	107	81

Red values correspond to statistically significant effects (P<0.05). Hyphen = data not available.
*The 4A QTL was also identified in an independent Option x Claire population, with Claire providing the beneficial allele.

3.1.3 Marker development for QTL deployment in UK germplasm

For an effective deployment of the QTL into elite germplasm through marker-assisted breeding, a critical necessity is the development of genetic markers that are tightly linked to the QTL. To facilitate this, it is necessary to increase both the marker density across the QTL interval, as well as the genetic resolution through lines with recombination events within the interval. This combined approach enables a high-resolution fine-mapping of the QTL as well as identification of markers that are tightly

linked to the QTL. However, due to the polyploid nature and complexity of the bread wheat genome, the development of robust genetic markers can be difficult and time-consuming.

Single Nucleotide Polymorphisms (SNP) are an abundant form of DNA variation in plant and animal genomes (Rafalski, 2002, Akhunov et al., 2009) and have become the preferred source of genetic marker because of their ease of use. Recent advances in wheat genomics has led to breakthroughs in SNP mining and discovery in hexaploid wheat (Borrill et al., 2015a). These have led to the development of high-density SNP arrays like the Infinium, iSelect and Axiom platforms (discussed in Chapter One, section 1.9). A number of studies (Allen et al., 2011, Winfield et al., 2012) have also reported on high-throughput SNP mining in UK wheat varieties, including seven of the parents used in this study - Alchemy, Avalon, Claire, Cadenza, Rialto, Robigus and Savannah. Winfield et al. (2012) identified about 100,000 exome-based SNPs in UK wheat varieties. These include 7,886 SNPs between Avalon and Cadenza, 15,031 SNPs between Rialto and Savannah and 16,815 SNPs between Robigus and Alchemy.

3.1.4 Aims

As introduced earlier, the consistent effects of six of the QTL previously identified suggest that they could be of potential use in stabilising HFN in UK wheat varieties. However, these QTL span very large genetic intervals and as such are not readily suitable for marker-assisted breeding. Also, these QTL have not been validated in independent materials, and very little is known about how, where and when these QTL exert their effects, as well as their pleiotropic effects on other agronomic traits. To address these questions, the genetic, physiological and agronomic characterisation of these QTL was been undertaken in this chapter.

3.2 Results

3.2.1 Development of isogenic lines for the characterisation of the QTL effects

Previous studies aimed at understanding PHS and PMA resistance have been performed using varieties that have diverse genetic background which could possibly interfere with the outcome of the study. To avoid this confounding effects, independent isogenic materials, F_3 and BC_3F_3, were developed for the characterisation of the selected QTL effects. These were developed through a backcrossing scheme described in Chapter 2 (section 2.1.1).

The independent F_3 materials developed for each QTL differ mainly across the respective QTL interval but have ~82 – 87 % recurrent parents' alleles in their background. These lines however still have background differences (due to limited backcrossing) which might reduce, or in some case, interfere with the expression of the QTL effect. Nonetheless, the rapid development of the F_3 lines provided early resources for the characterisation of the QTL effect.

Like the F_3 materials, the BC_3F_3 NILs differ at the QTL interval but have a greater degree of genetic homogeneity in the background (~95 %) due to recurrent backcrossing to the susceptible parents. The NILs therefore provided an improved resource to study the individual effects of each QTL on PHS and PMA by reducing confounding background effects. Also, since the NILs differ only across the QTL region, it is possible to perform more precise genetic and transcriptional characterisation of the QTL intervals through targeted SNP mining and differential expression analysis.

3.2.2 Targeted SNP discovery for marker development using isogenic lines

The discovery of a high number of SNPs in some of the parental lines used in this study has been earlier introduced. However as at the start of this Ph.D., the precise locations of many of these SNPs across the wheat genome were unknown and this limited their use in research and breeding programme. Therefore, to facilitate targeted polymorphism discovery precisely within the target QTL region, we obtained pre-publication access to the iSelect SNP array . We used this to genotype independent resistant and susceptible BC_3F_3 NIL pairs for each QTL along with their respective parents.

Using this approach between 8 and 13 % of the 90,000 iSelect SNPs were found to be polymorphic between parental pairs (Table 3.2). These numbers represent a whole genome SNP distribution between the parental pairs, many of which are not located within the QTL region of interest. To enrich for polymorphisms within the respective QTL regions, a SNP filtering approach that takes into account the background homogeneity of the NILs except at the target QTL interval was employed. Two criteria were used to search for SNPs that map within the QTL region of interest. First, SNPs must be monomorphic within each allele group. Second, SNPs must be polymorphic between allele groups (Figure 3.5).

Table 3.2: Number of polymorphic SNPs found between parents and between allele groups

Population	QTL	Number (percentage) of SNPs between Parents	Number of polymorphic SNPs between allele groups	Number of polymorphic SNP mapped to QTL chromosome arm	Percentage of Polymorphic SNP on QTL chromosome arm (%)
Haven x Soleil	1A	6764 (8 %)	353	195	55.2
Avalon x Cadenza	2D	11235(13 %)	163	20	12.3
Avalon x Cadenza	3A	11235 (13 %)	221	77	34.8
Avalon x Cadenza	7B	11235 (13 %)	131	82	62.6
Rialto x Savannah	$3A_{SR}$	8670 (10 %)	181	159	87.8
Alchemy x Robigus	4A	8598 (10 %)	352	46	13.1

Using these criteria, between 131 and 353 polymorphic SNPs were found between the NIL pairs for each QTL. The SNPs were assigned a putative chromosome arm location based on the best BLASTN hit of the sequence surrounding the SNP against the IWGSC CSS (see chapter One, section 1.9) chromosome arm sequences (International Wheat Genome Sequencing Consortium (IWGSC), 2014). Unsurprisingly, for most of the QTL, a large proportion of the SNPs map *in-silico* to the chromosome arm harbouring the targeted QTL region (Figure 3.6 and Table 3.2). For instance, for the 1A, $3A_{SR}$ (in the Savannah x Rialto population) and 7B NIL pairs, 55 %, 88 % and 63 % of the SNP mapped to the 1AL, 3AL and 7BL chromosome arms respectively (Figure 3.6). This represents an average of 29-fold enrichment compared to the 2.4% that would have been expected if the SNPs were uniformly distributed across the entire wheat chromosome arms.

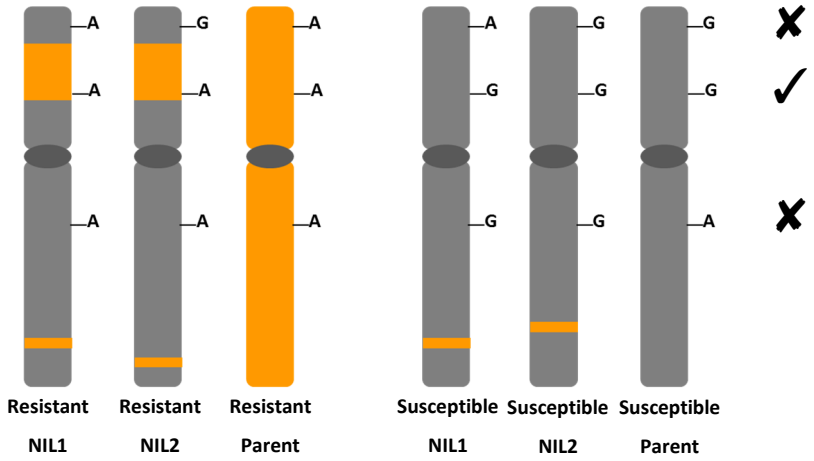


Figure 3.5: Targeted SNP discovery within target QTL regions. Following the genotyping of isogenic materials and parents, only SNPs which were monomorphic within allele group but polymorphic between allele groups were selected (✓). SNPs that did not meet these criteria were not carried forward (×)

Similar to the other QTL described above, a significant proportion (35 %) of the polymorphic SNPs found between the 3A QTL NIL mapped to the 3AL chromosome, (Figure 3.6). This percentage is less than was observed for the 1A, $3A_{SR}$ and 7B QTL regions. It is important however to note that the chromosome arms assignment was based on the best hit obtained from BLAST searches. It is possible that some SNPs might have been wrongly assigned to homoeologous chromosome arm since the SNP fragments used for the BLAST analysis were relatively short (< 200 bp) and there is high

sequences similarity among homoeologues (Krasileva et al., 2013). This seems to be the case for the 3A NIL pairs because about 27 % of the polymorphic SNPs map to the 3B chromosome which has a more complete assembly (Figure 3.6). For the 4A isogenic material, the 4AL chromosome arm harboured the highest percentage of the polymorphic SNP but the other chromosomes arms also showed a significant amount of variation between this NIL pairs. One reason for this is that, unlike the other NIL pairs, the contrasting NIL pair (susceptible and resistant) used for the 4AL QTL was developed from two independent DH lines which differ in their sizes of the resistant parent (Alchemy) introgression.

An unexpected result was that, the NIL pairs for the 2D QTL have more variations mapping to the 5BL and 5AL chromosome arms than to the 2DL region (Figure 3.6). Although the reasons for this are not yet clear, we suspect an unusual inter-chromosomal translocation between the 2DL and 5BL chromosome arms. It is possible that this translocation occurred within the QTL interval and had been retained after each backcross while selecting for the QTL.

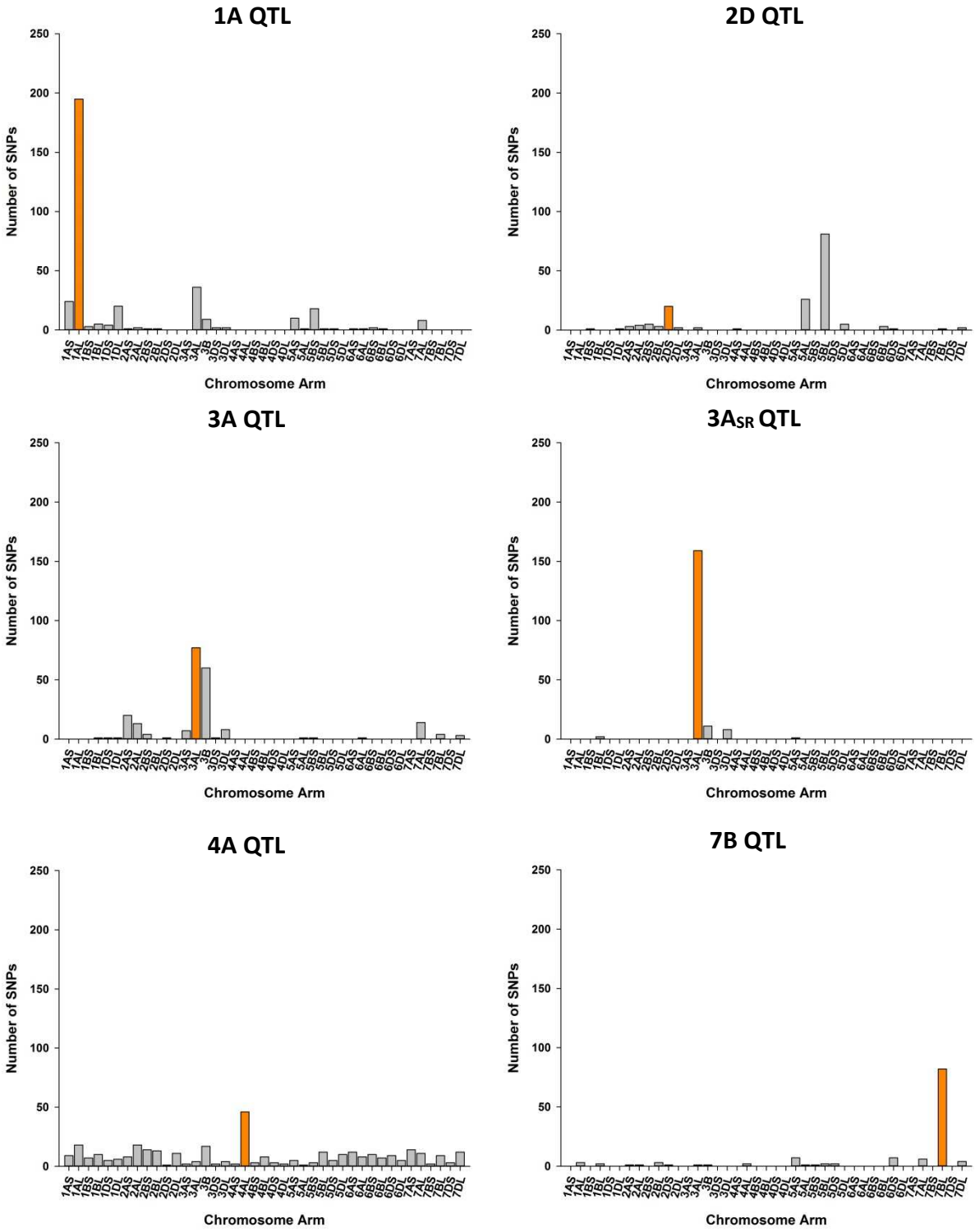


Figure 3.6: The distribution of polymorphic SNPs across the wheat genome. The locations of polymorphic SNPs between resistant and susceptible allele groups for 1A, 2D, 3A, $3A_{SR}$, 4A and 7B QTL are plotted across all the chromosome arm sequences of the bread wheat genome. The chromosome arm where each QTL is located is represented by the amber bars while the other chromosomes are represented by grey bars.

3.2.3 Field validation of PHS and PMA QTL

The PHS and PMA QTL chosen for characterisation in this study were identified in DH populations but have not been validated in independent genetic materials. In addition, DH populations do not allow for the independent study of individual QTL effects because such DH populations may have multiple QTL segregating in them. Therefore, to validate the QTL effects in independent genetic material, we used independent F_3 lines with contrasting resistant and susceptible alleles of each QTL. A list of these F_3 lines is presented in the general materials and methods chapter (Chapter Two, Table 2.1).

As at the time of this experiment, F_3 lines had not been developed for the 3A and 4A QTL. Field trials were conducted on the available F_3 lines at four different locations in the UK including: the John Innes Centre (JIC), RAGT Seeds (RAGT), Limagrain UK (LG) and KWS UK (KWS). PHS and PMA resistance in these lines were assessed through the HFN test. It is worth noting that the HFN value is inversely related to the amount of α-amylase produced in grains as a consequence of sprouting or PMA defects. So a high HFN value is indicative of high resistance to PHS and/or PMA.

Except for the $3A_{SR}$ QTL, resistance effects were observed for all the QTL in at least one experimental location. However, these effects were not consistent across all locations (Figure 3.7). The 1A QTL showed highly significant HFN effect at the JIC and KWS field sites. Lines with the resistance allele maintained close to two-fold higher HFN scores to those of susceptible F_3 lines. Although a similar effect was observed at the RAGT site, the difference between alleles was not significant. No difference was observed at the Limagrain field site, probably due to the higher HFN values for all the QTL at this location compared to the other locations.

In comparison with the 1A F_3 lines, the 2D F_3 lines showed higher HFN value (> 250 sec) in all the experimental locations. Nonetheless, a statistically significant difference was still observed for the 2D QTL at the KWS field site, and this was in the expected direction. That is, the resistant F_3 lines showed higher HFN than the susceptible F_3 lines. Resistant F_3 lines grown at Limagrain also showed some level of resistance, but this was not sufficient to produce a statistically significant effect. Similar to the 2D QTL,

the HFN values for the 7B F_3 lines were also high. Nevertheless, the HFN scores of lines with the 7B resistance were significantly higher at the JIC and KWS field site than their susceptible counterparts. In additions, resistant F_3 lines grown at RAGT showed some level of resistance but these were not significant.

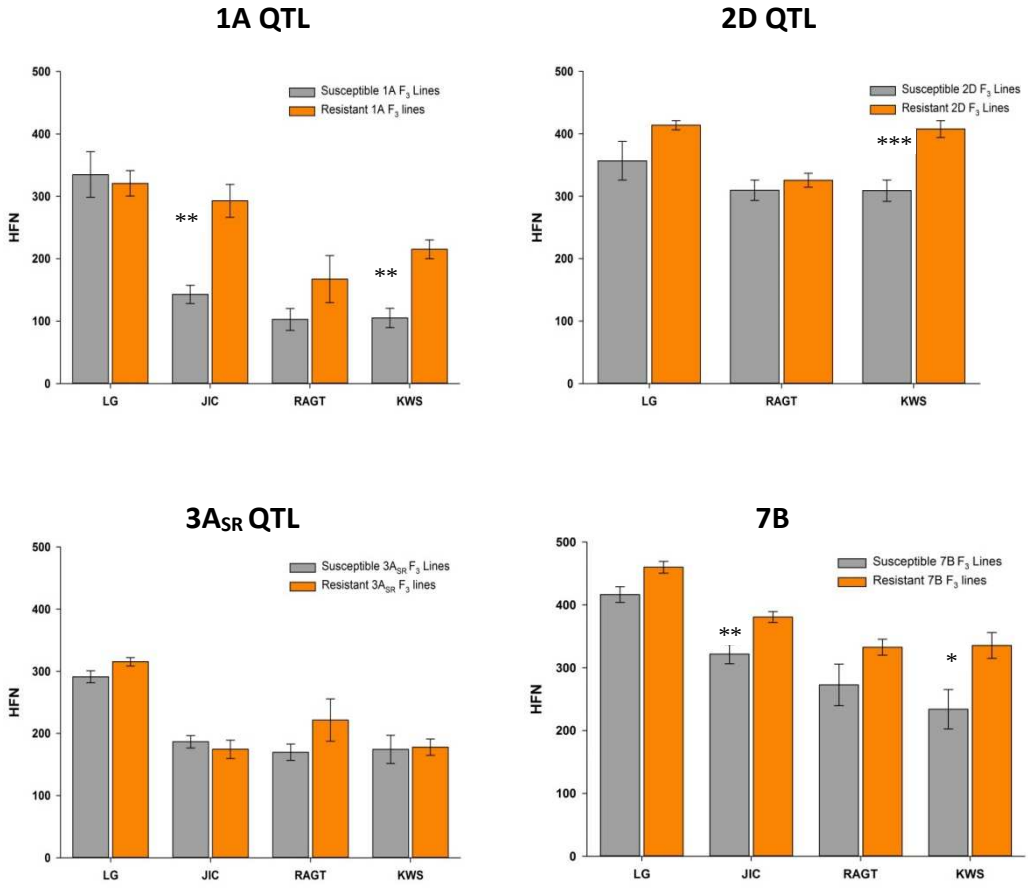


Figure 3.7: Multi-location validation of PHS and PMA QTL. The HFN score of the resistant (amber) and susceptible (grey) 1A, 2D, 3A$_{SR}$ and 7B F_3 lines grown at the John Innes Centre (JIC), RAGT Seeds (RAGT), Limagrain UK and KWS UK (KWS) field site. Error bars represent standard error of mean of at least 3 replications. Significant differences at $P < 0.05$ (*), < 0.01 (**), < 0.001 (***) are indicated.

3.2.4 Physiological characterisation of PHS and PMA QTL in F_3 and BC_3F_3 Lines

Susceptibility to PHS is negatively correlated with the depth of dormancy in grains (Gerjets et al., 2010, Gubler et al., 2005) with grains showing high dormancy being less susceptible to PHS. Given this relationship, the dynamics of dormancy loss was examined in F_3 and BC_3F_3 lines carrying alternate alleles of the 1A, 2D, 3A (only BC_3F_3), 3A$_{SR}$, 4A (only BC_3F_3) and 7B QTL. These lines are hereafter referred to as susceptible or resistant F_3 lines or NILs. This was also done at different developmental stages namely:

hysiological Maturity (PM, 40 % grain moisture content or peduncle senescence), Harvest Maturity (HM, ~20 % moisture content) and Post-harvest Maturity (PH,~14 days after HM). By examining the germination potential of seeds at these different stages, it was possible to study the nature of the QTL effects, as well as the timing of the QTL effects i.e. when does the expression of the QTL effect begin.

The rate of dormancy loss in these lines was assessed using the germination index test (GI) as described by Walker-Simmons (1987). The GI test is a weighted score which gives decreasing weights to the number of seeds that germinate in each consecutive day of a 7-day incubation period (details in Chapter 2, Section 2.5.1). The GI test is preferred to the germination percentage test because it not only estimates the proportion of seeds that germinate, but also the rate at which they germinate.

3.2.4.1 *PHS and PMA QTL show distinct effects on the rate of dormancy loss in F_3 lines*

As shown below in Figure 3.8, the rate of germination increased rapidly from physiological maturity through to post-harvest maturity in all the lines (parents, F_3 lines with resistant or susceptible alleles) and for all the QTL tested. Cadenza showed the lowest GI (0.082) at physiological maturity and the resistant 2D F_3 line, AC51-5B, had the highest GI value (0.898) at PH. However, differences were observed in the expression of the QTL effects across the various stages tested. These differences are described below for each QTL.

***1A QTL*:** The 1A QTL did not show any effect on the rate of seed germination at PM, but at HM there was a significant GI difference (P = 0.048) between the susceptible F_3 line (HS112-12A) and the resistant F_3 line (HS112-19B), with HS112-12A showing the greater GI (Figure 3.8). Although both F_3 lines showed increased germination potential at PH, the GI difference between these lines was further increased at PH (P = 0.006), still HS112-12A showing the greater GI (0.891) compared to HS112-19B (0.698). The pattern of the QTL effect in the 1A F_3 line was also mirrored in the parents at PM and PH with Haven showing a higher germination potential than Soleil.

2D QTL: In contrast to the 1A QTL, the 2D QTL showed an earlier effect, with significant GI differences between F_3 lines seen as early as PM (Figure 3.8). The GI of the

susceptible F_3 line (AC51-7A) was almost twice (P = 0.008) the GI displayed by the resistant F_3 line (AC51-5B) at PM. Although this GI difference was reduced, it was still significant at HM. At PH, there were no significant differences between the F_3 lines. While the GI scores of the parents were lower than those of their corresponding F_3 lines, the parents showed similar pattern of GI difference that was seen in the F_3 lines.

$3A_{SR}$ QTL: The $3A_{SR}$ did not show any effect at PM and HM. However, at PH, a marked GI difference (P = 0.001) was observed between the F_3 lines with the susceptible F_3 line (RS37-1B) having a GI of 0.836 while the resistant F_3 line (RS37-3A) showed a lower GI of 0.448 (Figure 3.8). This difference in GI was mainly caused by a sharp increase in the germination potential of RS37-1B from HM to PM. Although the parents - Rialto and Savannah - showed a very similar pattern of GI difference, the effect of the QTL in the parent was in the opposite direction to that observed in the F_3 lines. That is, Rialto (the susceptible parent) showed a lower GI of 0.461 compared to Savannah (the resistant parent) with a GI of 0.897.

***7B QTL*:** In contrast to all of the QTL described above which showed GI difference at one stage or another, the 7B effect on the rate of germination was not observed at any of the stages tested (Figure 3.8). The susceptible (AC150-3A) and the resistant F_3 lines showed (AC150-18B) similar GI profiles. In relations to the parents, the GI profile of both F_3 lines was similar to the susceptible 7B parent - Cadenza.

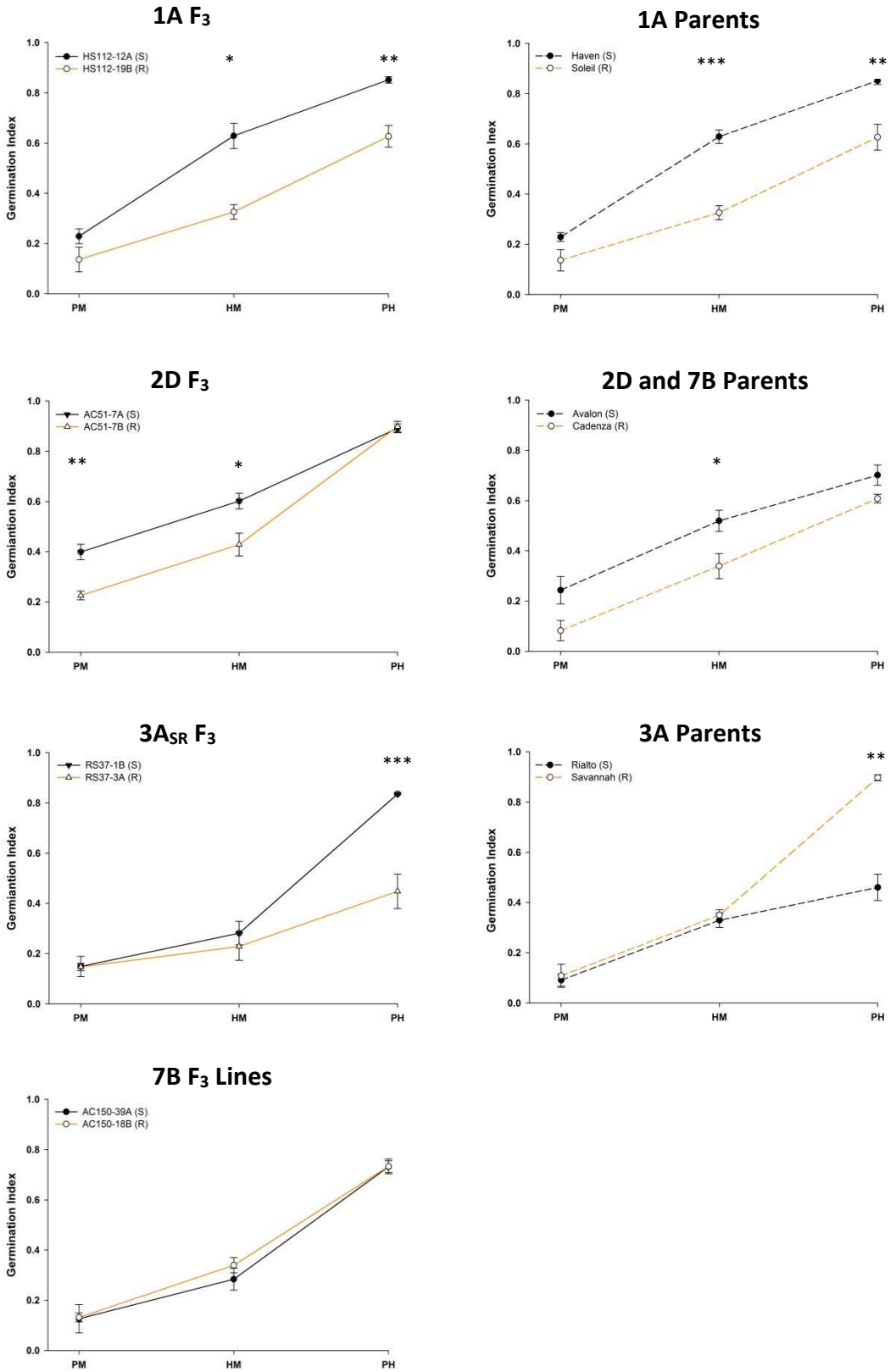


Figure 3.8: **PHS and PMA QTL show distinct effects on the rate of dormancy loss**. The germination index of seeds harvested from F_3 lines and parents for the 1A, 2D, $3A_{SR}$ and 7B QTL alleles. Seeds were tested at Physiological Maturity (PM), Harvest Maturity (HM) and Post-Harvest Maturity (PH) and germinated at 17 °C. Error bars represent SEM of 3 biological replications for each time point. Significant differences at $P < 0.05$ (*), < 0.01 (**), < 0.001 (***) are indicated.

3.2.4.2 *Imbibition temperature affects the expression of QTL effects in F_3 lines.*

Previous studies have shown that there is an inverse relationship between germination temperature and the germination index of grains of different wheat varieties (Reddy et al., 1985, Nyachiro et al., 2002), with increases in germination temperature resulting in decreases in germination index. Also, dormancy is broken in most wheat varieties when germinated at low temperature. However for most of the PHS QTL reported to date, the effect of germination temperature on the expression of these QTL effects has not been studied. Therefore, in addition to germinating the F_3 grains at 17 °C described previously, we also germinated seeds at 10 °C and 22 °C to assess the effect of germination/incubation temperature on the expression of the QTL effects.

The result of this experiment (Figure 3.9) confirmed a consistent inverse relationship between germination temperature and germination potential of seeds. At almost all stages of grain development tested, GI decreased as the germination temperature was increased from 10 °C to 22 °C. These results correlate well with the previous reports on the effect of incubation temperature on germination (Nyachiro *et al*, 2002). The only exception was at PH for the 1A and $3A_{sR}$ QTL where the GI of the susceptible parent and F_3 lines were greater at 22 °C than at 10 °C.

Furthermore, at 10 °C the susceptible and resistant F_3 lines of all the QTL showed no significant GI difference. The only exception to this was the small but significant GI difference between the 7B F_3 lines, and this was only observed at harvest maturity. The absence of the QTL effects at 10 °C is probably due to the breaking of dormancy in all the lines.

In addition, at 22 °C, the QTL effects were expressed in a very similar pattern to what was observed at 17 °C. Also, , the $3A_{SR}$ QTL parents, Savannah and Rialto, once again showed GI difference in the opposite direction to the difference observed between the F_3 lines at 22^{o}C. This further confirms the phenotype discrepancy between parental and F_3 lines for this QTL. In addition, when the GI of lines were compared at 17 °C and 22^{o}C, the effects of late expressing QTL (1A and $3A_{SR}$) were found to be greater at 22 °C than at 17 °C. This was mainly due to increases in the dormancy of the resistant lines for the QTL. Taken together, our results show an influence of temperature on the expression of our target QTL effects.

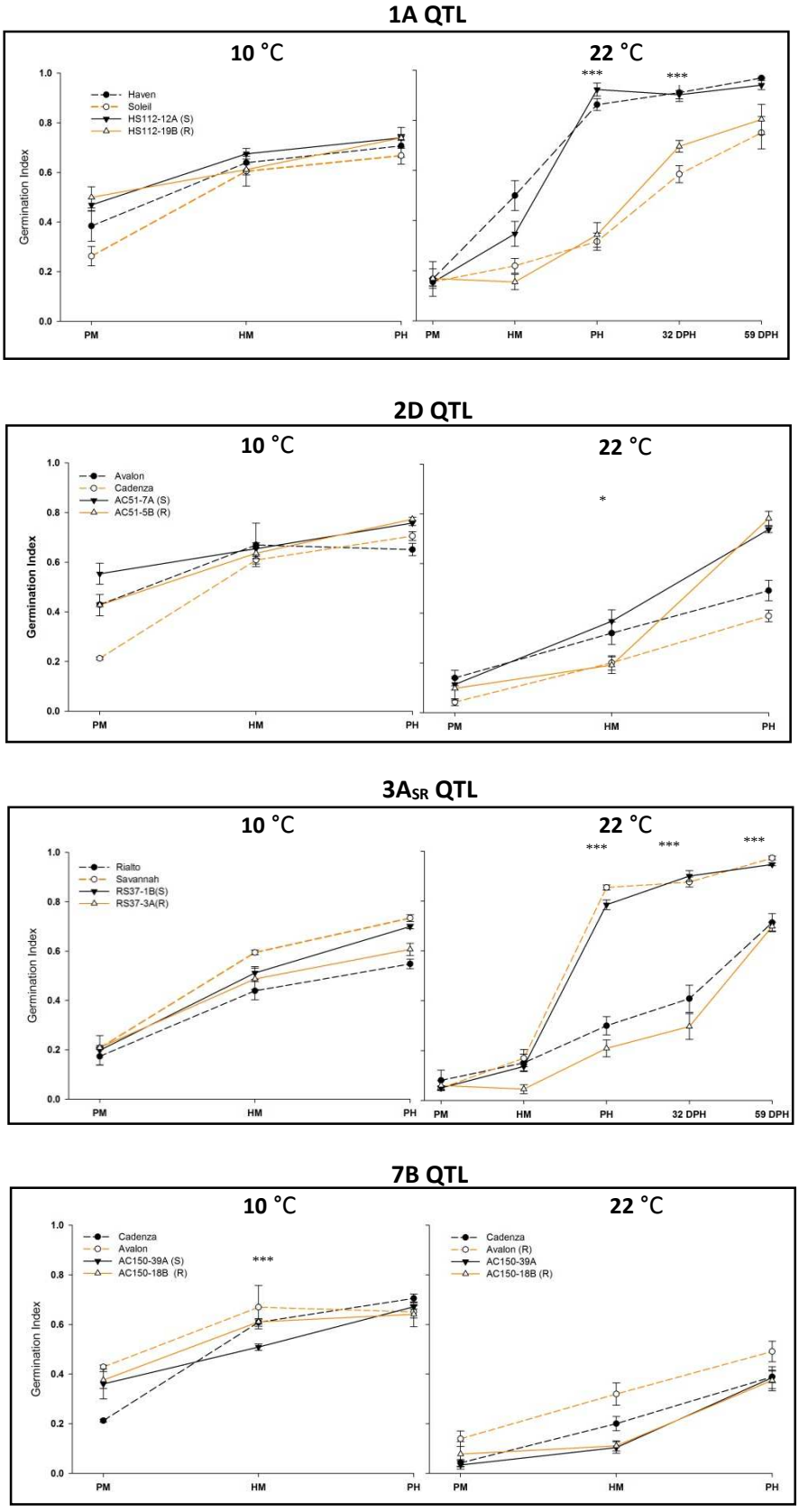


Figure 3.9: **The effect of germination temperature on the expression of QTL effect**. Seeds from parents and F_3 seeds carrying the susceptible or resistant QTL 1A, 2D, $3A_{SR}$ and 7B QTL alleles were harvested at Physiological Maturity (PM), Harvest Maturity (HM) and Post-Harvest Maturity (PH) and germinated at 10 °C, and 22 °C. For the 1A and $3A_{SR}$ lines, two more post-harvest experiments were done at 32 Days post-harvest (DPH) and 59 DPH. Error bars represent SEM of 3 biological replications for each time point. Significant differences between F_3 lines (solid lines) at $P < 0.05$ (*), < 0.01 (**), < 0.001 (***) are indicated.

To further investigate the late QTL effect expressed by the 1A and $3A_{SR}$ QTL, more post-harvest GI test were done at 32 days post-harvest (DPH) and 59 DPH. These were germinated at 22 °C since the highest GI difference was observed at this temperature. This showed that the significant GI difference observed for these QTL persisted until 32 DPH for the 1A QTL and 59 DPH for the $3A_{SR}$ QTL, even though the resistant lines showed a rapid increase in germination potential at these late time points.

3.2.4.3 *PHS and PMA QTL show mixed effects on the rate of dormancy loss in BC_3F_3 NILs*

Three independent resistant and susceptible BC_3F_3 NILs for the 1A, 2D, 3A, $3A_{SR}$ and 7B QTL, as well as their parents, were tested for the effects of the QTL on the rate of germination. For the 4A QTL, one resistant and one susceptible BC_3F_3 NIL were used. The dormancy status of these seeds was assessed through the germination index test at four different time points. These include Physiological Maturity (PM, ~40% grain moisture content); Harvest Maturity (HM, ~20% grain moisture content); Post-harvest Maturity (PH, 14 days after HM) and another PH time point at 28 DPH. The results obtained for each QTL are described below.

1A QTL: GI difference was observed between some susceptible and resistant NILs for the 1A QTL at the first post-harvest (PH) stage of grain maturation, just as was observed with the F_3 lines (Figure 3.10). However, this difference was masked by the non-homogeneity in the phenotypes of sister NILs harbouring similar alleles. That is, one of the three NILs with the susceptibility allele (HS112/15/50) behaved like the resistant NILs. Similarly, one of the three NILs with the resistant allele (HS112/9/66_5) showed similar phenotype as the susceptible NILs.

2D QTL: Unlike the F_3 lines, the resistant and susceptible BC_3F_3 2D NILs did not show any difference at the physiological or harvest maturity stages (Figure 3.7). The only difference observed was at the PH time point, where the susceptible NILs generally showed lower GI to the resistant NILs. Also, just as was observed for the 1A QTL, this GI difference between NILs was masked by the non-homogeneity in the phenotype of

sister NILs. This was because one of the three susceptible 2D NILs (AC73/10/6_1) showed similar phenotype to the resistant NILs.

3A and 3A$_{SR}$ QTL: The 3A QTL NILs did not show any observable GI difference (Figure 3.7), while the alleles groups of the 3A$_{SR}$ QTL only showed a minor but significant difference at PM. Also, all the 3A$_{SR}$ NILs (both susceptible and resistant) showed high depth of dormancy that persisted to the later post-harvest time points. This is in contrast to what was earlier observed in the F_3 lines and the reason for this is not yet clear.

4A QTL: The 4A QTL, however, showed a significant effect between the susceptible and resistant NILs tested. The effect was only observable at post-harvest maturity after a period of after-ripening (after-ripening is described in Chapter One, 1.7.4). At the peak of this effect (first PH time point), the susceptible NIL and parent showed approximately three times the GI of the resistant NIL. At 28 DPH, there was still an observable difference between the NIL, however, this difference was not significant (P = 0.061).

7B QTL: Like was observed in the F_3 lines, there was no observable GI difference between NILs of the 7B QTL except at the last post-harvest time point. However, this QTL showed a potential effect on PMA induction in grains. This will be described later in section 3.2.4.6.

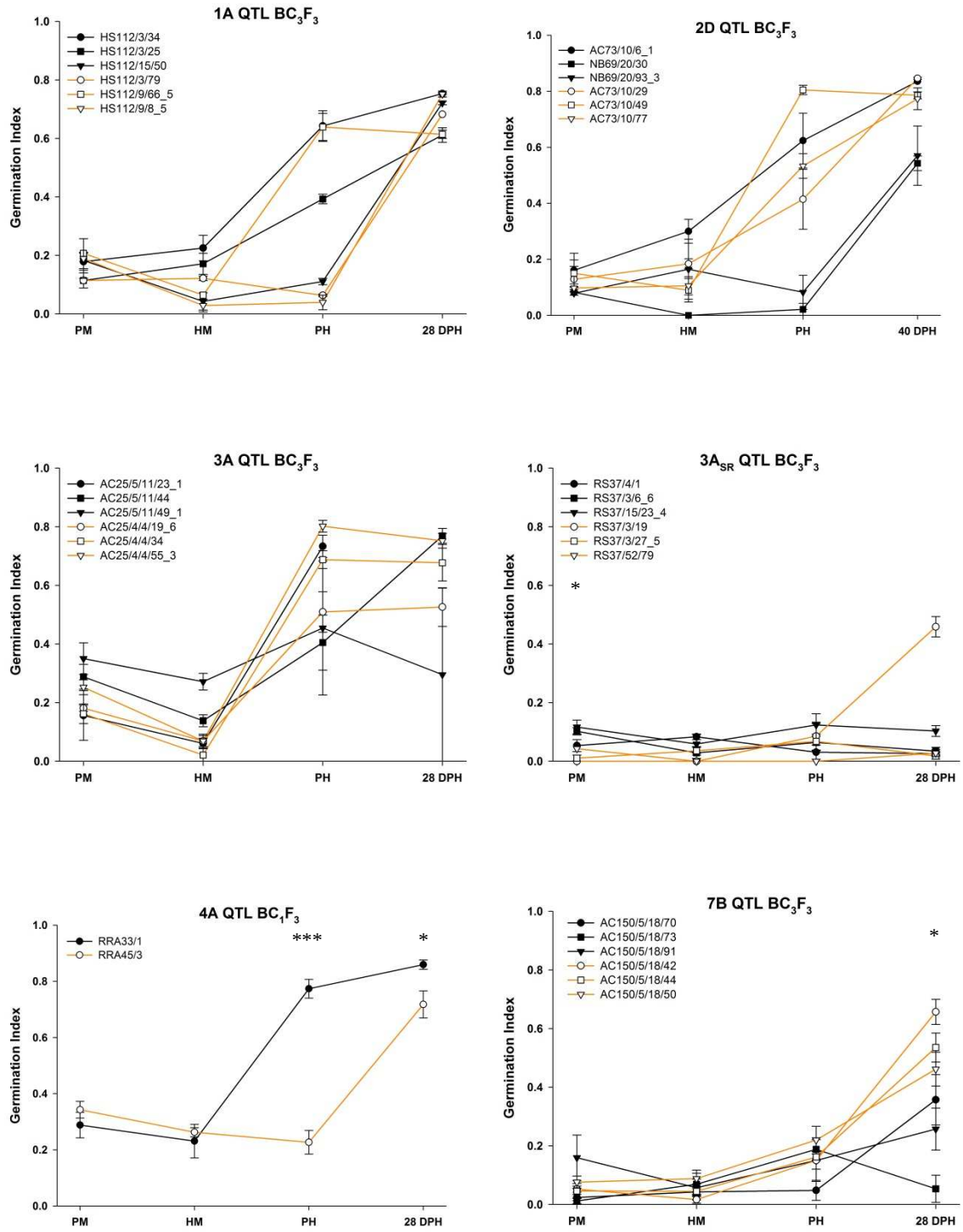


Figure 3.10: **PHS and PMA QTL show mixed effects on the rate of dormancy loss in NILs**. The germination index of seeds harvested from NILs with susceptible (black lines) and resistant (amber lines) parental alleles for the 1A, 2D, 3A, $3A_{SR}$ and 7B QTL. Seeds were tested at Physiological Maturity (PM), Harvest Maturity (HM) and two Post-Harvest time points (PH) and germinated at 16 °C. Error bars represent SEM of 3 biological replications for each time point. Significant differences based on comparison between allele mean at $P < 0.05$ (*), and < 0.001 (***) are indicated.

3.2.4.4 *Investigating the cause of the heterogeneous GI phenotype of the 1A and 2D NILs*

For two of the QTL (1A and 2D), NILs which were expected to have similar phenotypes based on the marker-assisted selection showed heterogeneous behaviour. One reason for this heterogeneity in the sister NILs phenotypes could be that the sister NILs for these QTL are not as isogenic as expected. The contrasting phenotype of the sister NILs could, therefore, be caused by residual segregating loci in the background.

Given that two out of the three NILs used in this experiment had already been genotyped with the iSelect SNP array (in section 3.2.2), it was possible to test this hypothesis. Thus, an analysis of residual polymorphisms present between sister NILs (NILs with the same QTL allele) was carried out. For the 1A QTL, line HS112/9/66_5 with the unexpected GI phenotype was compared with its sister line - HS112/9/8_5. For the 2D QTL, line AC73/10/6_1, which showed the unexpected phenotype was compared with its sister NIL – NB69/20/30.

For the 1A sister NILs, a large proportion of the polymorphism between the sister NILs also maps to the 1A chromosome (Figure 3.11 A). By comparing the genetic position of these SNPs as published by Wang et al. (2014), it was discovered that the genetic interval of the mapped SNPs between these sister NILs partly overlaps with the genetic interval of the mapped SNPs found between the allele (resistant and susceptible) groups in section 3.2.2. This implies that there are residual polymorphisms within or close to the 1A QTL interval between these sister NILs and this might be the reason for the heterogeneous phenotype. Although the 2DS chromosome arm of the 2D sister NILs appeared isogenic, more than 20% of the polymorphic SNPs between these sister NILs map to the 5BL chromosome arm (Figure 3.11 B). This is the same chromosome arm harbouring the majority of the polymorphic SNP between the resistant and susceptible 2D allele groups in section 3.2.2.

These findings (residual polymorphisms) partly provides an explanation for the heterogeneous behaviour of some of the sister NILs. However, these do not provide conclusive evidence for the heterogeneous phenotype of the NILs. More work will therefore be required to characterise further the NILs.

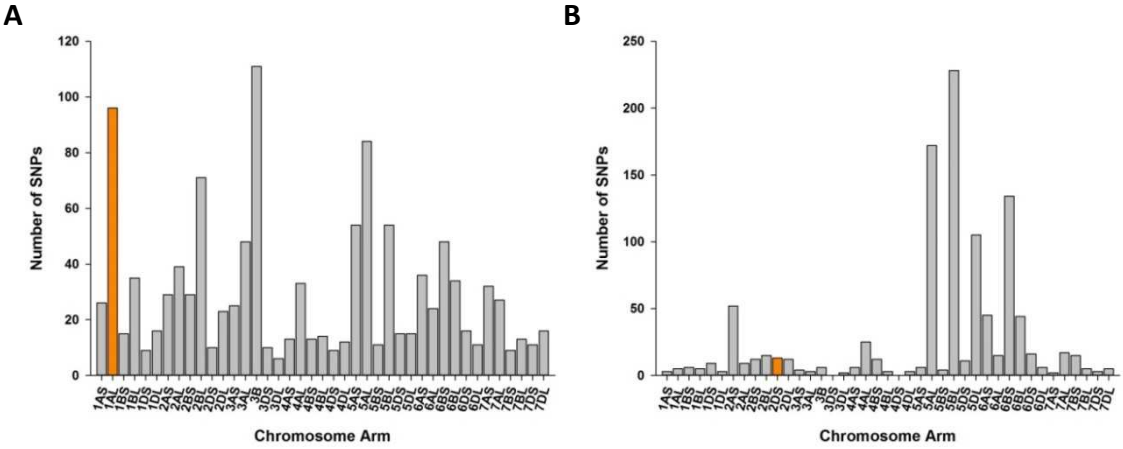


Figure 3.11: The distribution of residual polymorphisms between heterogeneous sister NILs: The location of polymorphic SNPs found between (A) HS112/9/66_5 and HS112/9/8_5 for 1A QTL and (B) AC73/10/6_1 and NB69/20/30 for 2D QTL. The chromosome arm where each QTL is located is represented by the amber bars while the other chromosomes are represented by grey bars.

3.2.4.5 *Low temperature during grain development increases the depth of dormancy in BC_3F_3 lines*

Some studies have shown that temperature during grain developmental plays a role in setting the depth of dormancy displayed by wheat seeds during maturation (Buraas and Skinnes, 1985, Reddy et al., 1985). In a similar manner, the temperature at the onset of grain development can also influence the expression of PHS resistance QTL at the later stages of grain maturation and after-ripening. In a recent study, where a PHS susceptible line and a PHS resistant line were grown at 13 °C and 25 °C, difference in germination percentage was found between the lines only when grown at 13 °C and not at 25 °C (Nakamura et al., 2011b). These results point to an effect of grain developmental temperature in the control of PHS QTL.

To investigate the effect of temperature during grain development on the level of dormancy and possibly the expression of the QTL effect, the NILs used in the previous physiological characterisation experiment (section 3.2.4.3 above) were also grown at 13 °C from anthesis onwards. The rate of germination in these lines was assessed using the germination index test. Using thermal time calculations (see Chapter Two, section

2.6.1), samples were harvested at similar stages as the samples used in the previous physiological experiment (section 3.2.4.3).

Plants of NILs grown at 13 °C from anthesis were expectedly delayed in their overall development and maturation due to the low temperature of growth compared to NILs grown at 20 °C. Also, NILs grown at 13 °C displayed a higher depth of dormancy when compared with NILs grown at 20 °C at a similar thermal duration post anthesis (Figure 3.12). This increase in dormancy, however, varied between QTL. For instance, while NILs for the 1A and 4A QTL showed a moderate increase in dormancy depth when grown at 13 °C from anthesis, dormancy in NILs for the 2D, 3A and $3A_{SR}$ was markedly increased when grown at 13 °C.

Due to this increase in dormancy, most of the QTL did not show any effect on germination rate. The only exception to this was the 4A QTL, which caused GI difference between the NILs at PH ($P = 0.021$) with the susceptible NIL having a higher germination rate than the resistant NIL. This GI difference was still observed at 35 DPH ($P = 0.016$) suggesting that this QTL effect is expressed during the after-ripening period. This 4A QTL effect was similar to that observed in plants grown at 20 °C.

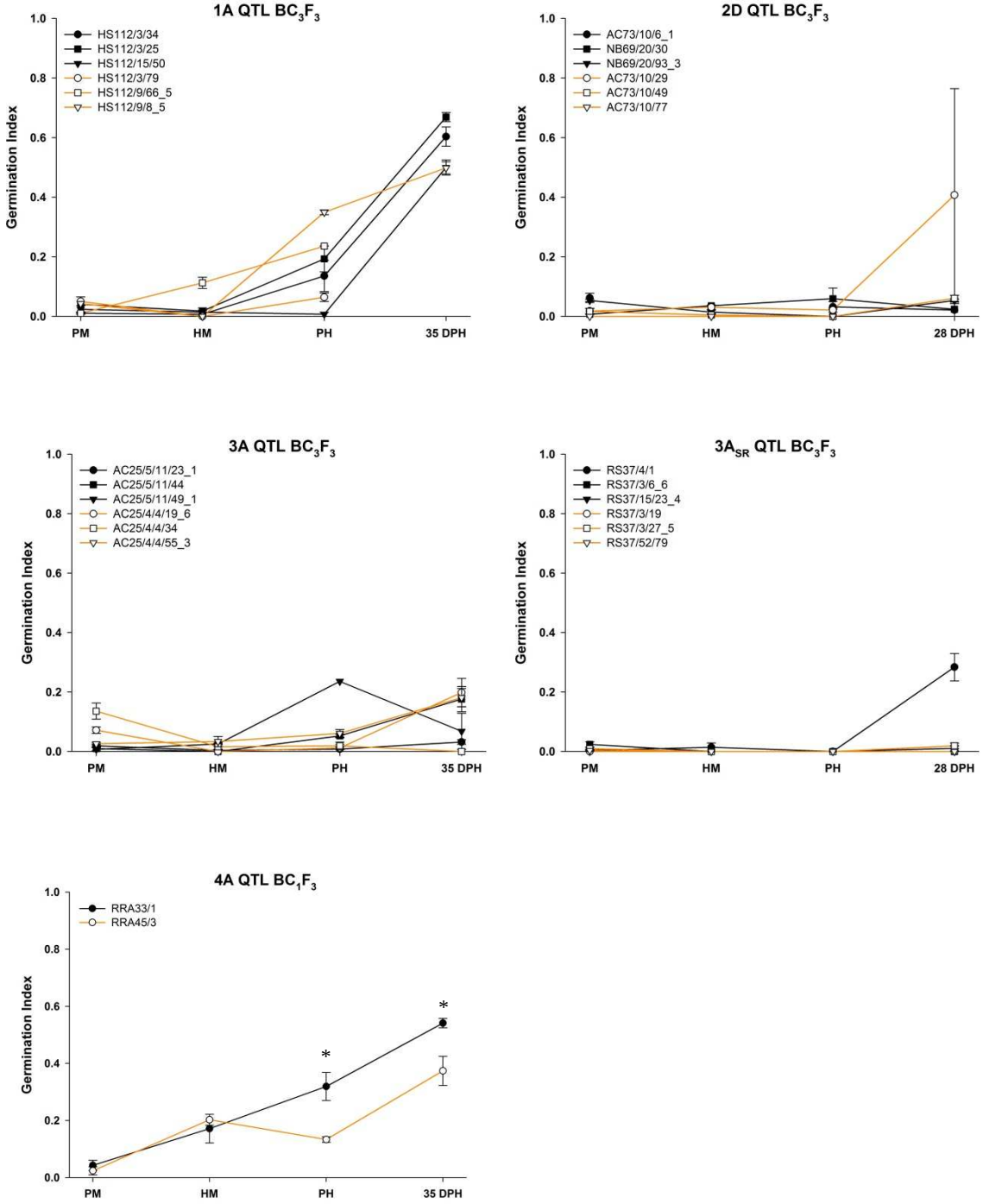


Figure 3.12: Low grain developmental temperature increases the depth of dormancy in wheat NILs. The GI of seeds harvested from NILs with susceptible (black lines) and resistant (amber lines) parent alleles for the 1A, 2D, 3A, $3A_{SR}$ and 4A QTL grown at 13°C from anthesis. Seeds were tested at Physiological Maturity (PM), Harvest Maturity (HM) and two Post-Harvest time points (PH) and germinated at 16 °C. Error bars represent SEM of 3 biological replications for each time point. Significant differences based on comparison between allele mean at $P < 0.05$ (*) is indicated.

3.2.4.6 *The 7B QTL show potential resistance effect on cold inducible PMA.*

In the F_3 and BC_3F_3 GI experiments, the 7B QTL did not show any observable effect on the germination potential of grains. This might be because the 7B QTL is mainly a PMA QTL, which is a trait that is independent of the PHS phenotype and hence the dormancy status of seeds. In view of this, the effect of the 7B QTL on PMA induction was tested in two independent F_3 lines per allele and three independent BC_3F_3 NILs per allele. This was done by measuring the α-amylase activity of grains using the Megazyme (UK) Ceralpha α-amylase test kit.

The average α-amylase activity of the susceptible F_3 lines was higher than those of the resistant lines. However, this difference was not statistically significant (P = 0.059, Figure 3.13 A). Similarly, difference in α-amylase activity was observed between the susceptible and resistant BC_3F_3 NILs of the 7B QTL. The NILs harbouring the susceptibility allele showed more amylase activity than the resistant NILs (Figure 3.13 B), however, just like in the F_3 lines, the difference observed was not statistically significant. These insignificant results were most likely due to high variability in the induction of PMA in individual seeds (Mares and Mrva, 2008). The tests, however, show a potential resistance effect of the 7B QTL on PMA induction

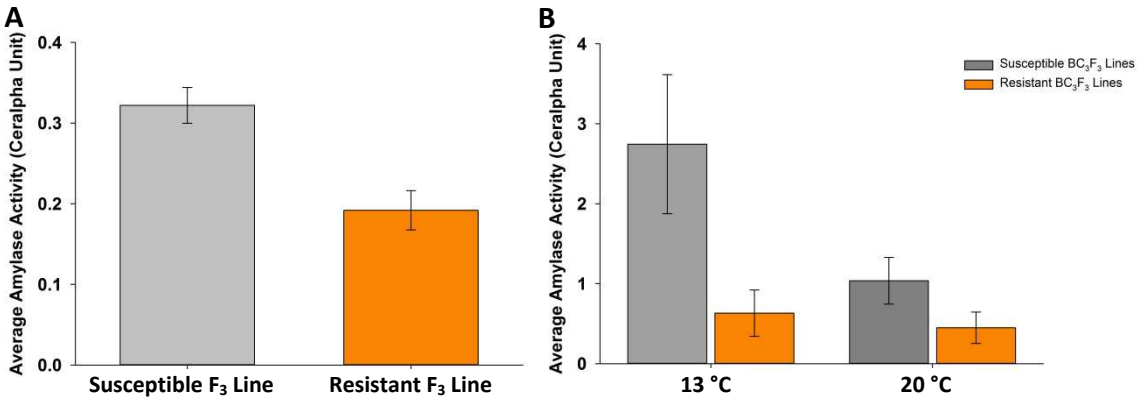


Figure 3.13: **Potential effect of the 7B QTL on PMA induction**. Average amylase activity of grains harvested from 7B resistant and susceptible F_3 lines (A) and BC_3F_3 lines (B). Error bars represent SEM of amylase activity of the 2 independent lines in (A) or 3 independent lines in (B).

In addition, the 13 °c cold temperature treatment during grain development also had a marked effect on PMA induction, especially in the susceptible 7B NIL. There was more than a two-fold increase in α-amylase activity in grains of the susceptible NILs when grown at 13 °c compared with plants grown at 20 °c. This result is consistent with the reports by Mrva et al. (2006) which showed a strong PMA induction in PMA susceptible lines when plants were transferred to the cold between 25 to 35 days post anthesis. This increase in PMA induction in the cold was not however observed in plants of the resistant NILs.

3.2.5 Effect of the PHS and PMA QTL on agronomic traits

Resistance to PHS and PMA induction is a desirable trait that ultimately improves the bread-making quality of wheat grains. However, it is important that these are not achieved at the expense of other agronomic traits that are of high breeding value. Such traits include yield, grains morphometric parameters (size, weight and thousand grain weight) and adaptive traits like plant height, and flowering time. It is, therefore, important to examine whether the PHS and PMA QTL studied have detrimental effects on these other agronomic traits either by linkage or pleiotropy.

To determine the effects of the QTL on agronomic traits, BC_1F_3 NILs carrying alternate alleles of the 2D, 3A, $3A_{SR}$ and 7B QTL along with their parents (Avalon, Cadenza, Rialto and Savannah) were sown in different locations in the UK, in field sites of collaborating breeding partners - KWS, Limagrain (LG) and RAGT. The yield, height, ear emergence and grain size parameters of these NILs were examined. Due to insufficient amount of seeds for field trials, the 1A and 4A QTL NILs were not included in this experiment. Also, for the 7B QTL, only one NIL per QTL allele was used for this experiment. Details of the number of NILs for each QTL and the experimental locations can be found in Chapter Two (section 2.2.3 and Table 2.2).

Using QTL alleles and experimental locations as factors, a two-way ANOVA analysis was carried out on the effect of the 2D, 3A and $3A_{SR}$ QTL on each of the trait. This showed a significant effect of location on grain area, grain width and yield ($P < 0.001$). The effects of the QTL however varied from traits to traits with some QTL

producing significant effect on some traits across all location (2D, 3A and $3A_{SR}$) and others showing no consistent effect on any traits (7B QTL). There was no evidence for any significant genetic (QTL) x environment interaction on any of the traits (P: 0.157 - 0.930). Table 3.3 gives a summary of the effects of the QTL on the various traits tested.

Height: Of the four QTL tested, only the 2D and the $3A_{SR}$ QTL showed effect on plant height architecture and this was significant in all the environment tested, suggesting that these effects on plant height are robust. Across all the locations, the 2D QTL resistant NILs were taller than their susceptible counterpart by an average height of 12.5 cm. The increased height of the 2D resistant NILs agrees with the height phenotype of the Cadenza parent that provided the resistance allele. In contrast, the $3A_{SR}$ resistance QTL introgression led to a reduction in height across locations. On average, the $3A_{SR}$ resistant NILs were 4.51 cm shorter than their susceptible counterparts. The increased height of the susceptible $3A_{SR}$ NILs agrees with the height phenotype of the tall and PHS susceptible parent, Rialto (Figure 3.14).

Figure 3.14: Height effect of the $3A_{SR}$ QTL. Height of parental (Savannah and Rialto) and isogenic BC_3F_3 susceptible (NIL A) and resistant (NIL B) lines for the $3A_{SR}$ QTL.

Flowering: Days to ear emergence was used for the evaluation of the flowering time trait. Field evaluation of the NILs showed that only the 3A and $3A_{SR}$ QTL showed effects on days to ear emergence, and this was consistent in all the locations tested. In the Avalon x Cadenza 3A population, the resistant NILs reached ear emergence on an average of two days later that the susceptible NILs. This was also true for the Savannah x Rialto population with the resistant allele associated with a later ear emergence phenotype. This suggest that the PHS resistance offered by the 3A and $3A_{SR}$ QTL might negatively impact on the duration of grain filling.

Grain parameters*:* The effects of the QTL on grain size and shape variation in the NILs were also examined. This was done by measuring thousand grain weight (TGW) and morphometric parameters like grain area, width and length. The 3A PHS resistance was associated with a significant reduction in TGW (average of 7.3 %) and grain area (average of 6 %) in all the locations tested. This reduction was as a result of the 3A effects on both grain length and width in the NILs harbouring the 3A resistance allele. The 2D resistance QTL was also associated with an average of 2.7 % reduction in grain length across all the location tested but this did not result in any reduction in grain area or TGW as did the 3A PHS resistance. Although the $3A_{SR}$ resistance showed delayed ear emergence phenotype similar to the 3A QTL, its effect on grain morphometric parameter was not consistent as it only showed a grain size effect in one location (LG).

Table 3.3: Summary of the agronomic effects of the PHS and PMA QTL

Traits	Locations	R effect			
		2D	3A	$3A_{SR}$	7B
Height (cm)	KWS	9.4	4.76	-4.33	-
	LG	13.3	2.31	-4.67	1
	RAGT	14.9	1.52	-4.53	-
Ear Emergence (Days)	KWS	-0.42	2.07	2.07	-
	LG	-0.59	2.36	1.67	1
	RAGT	-1.23	2.25	1.93	
TGW (g)	KWS	-2%	-7%	-3%	-
	LG	2%	-7%	-5%	-6%
	RAGT	0%	-8%	-2%	-2%
Area (mm^2)	KWS	-2%	-5%	-1%	-
	LG	0%	-6%	-3%	-4%
	RAGT	-2%	-7%	-1%	-2%
Width (mm)	KWS	0%	-2%	-1%	-
	LG	2%	-4%	-2%	-3%
	RAGT	1%	-3%	0%	-1%
Length (mm)	KWS	-3%	-3%	0%	-
	LG	-2%	-3%	-2%	-2%
	RAGT	-3%	-4%	0%	-2%
Yield (Tonne/ha)	KWS	-2%	2%	2%	15%
	LG	6%	-3%	1%	4%
	RAGT	-8%	3%	8%	11%

Red values correspond to statistically significant effects (P<0.05).

Yield: Yield evaluation showed that none of the QTL tested showed a consistent effect on yield in any of three locations tested. In most locations, there was no significant yield difference between the resistant and susceptible NILs. However, the 2D and 7B QTL showed significant yield increase in one and two of the three experimental locations, respectively. In both instances, the resistant allele provided the yield increase. In contrast, the 3A PHS resistance allele showed a detrimental effect on yield in one of the three locations (LG).

3.3 Discussions

In this chapter, the genetic, physiological and agronomic characterisations of six QTL for PHS and PMA resistance in the UK have been undertaken. Although these QTL represent a proportion of the quantitative resistances to PHS and PMA present in UK wheat germplasm, they were chosen because they were identified to be the most significant and stable QTL. These QTL include PHS QTL on chromosomes 1A, 2D, 3A (two loci) and 4A and a PMA QTL on chromosome 7B.

Selected PHS and PMA QTL co-localises with QTL identified in other studies

Some of the QTL characterised in this study have been previously identified in other studies. For instance, QTL for PHS resistance have been mapped to chromosome 1A in populations derived from white-grained hexaploid wheat (Singh et al., 2010, Munkvold et al., 2009) and durum wheat varieties (Knox et al., 2012, Singh et al., 2014). Based on comparisons with the consensus maps by Somers et al. (2004), these 1A QTL are in a similar interval with the QTL characterised in this study. In addition, a 1A QTL was also identified from recombinant inbred line population made between a red-grained and white-grained varieties but this map to the distal end of the 1AS chromosome QTL (Mohan et al., 2009). Also, the 1A QTL identified by Singh et al. (2010) was hypothesised to be associated with amylase activity. This partly agrees with the observation in field trial experiments in 2007 and 2008 (Table 3.1) where there was a significant effect of the 1A QTL on HFN measure. However, the 1A QTL produced more consistent effects on PHS rather than on PMA in all four years of field trial.

A 2D PHS QTL flanked by markers *Xgwm261* and *Xgwm4845* was identified in a population from synthetic wheat and a hexaploid wheat variety (Ren et al., 2008). This QTL maps to the same genetic interval as the 2D QTL in this study which is also flanked by marker *Xgwm261.* Interestingly, *Xgwm261* is a closely linked marker to the semi-dwarfing gene - *Rht8* (Gasperini et al., 2012). Similarly, Munkvold et al. (2009) identified a 2D QTL for PHS resistance in the same white grained hexaploid

wheat population in which the 1A QTL was identified and this map within a similar region to the 2D QTL in this study.

Two 3A QTL were characterised in this study. This includes the 3A QTL from the Avalon x Cadenza population and the $3A_{SR}$ QTL from the Savannah x Rialto population. Based on the close proximity of the markers with the highest LOD peak in both QTL (*wPT-1562* for 3A and *gwm264* for $3A_{SR}$), it is tempting to speculate that both QTL might be covering the same locus.

Very recently, Kumar et al. (2015) identified a QTL for PHS resistance in a white-grained wheat population and this appears to be located in a similar genetic interval to the 3A QTL in this study. This implies that the 3A QTL characterised in this study, although identified from a red-grained wheat population, is not associated with grain colour and might also be segregating in the white-grained wheat population. This is unlike some QTL identified on chromosome 3A that are associated with the red grain colour. Other QTL have been identified for resistance to PHS on wheat chromosome 3A (Mori et al., 2005, Groos et al., 2002a, Kulwal et al., 2005, Nakamura et al., 2011a, Liu et al., 2013b), however, many of these are too distal on the short arm or longs arm of the 3A chromosome and as such do not co-localise with the 3A QTL in this study.

Although the 4A and the 7B QTL were not identified consistently across the four years of field trial, they however produced the largest PHS effect (for the 4AL QTL) or HFN effect (for the 7B QTL) in at least two years of field trials. This suggests that both QTL are of importance but might be more difficult to identify. This highlights the importance of using the appropriate phenotyping strategy for these QTL. In support of the importance of the 4AL QTL for PHS resistance, it has been identified in many different germplasms across the world (Albrecht et al., 2015, Cabral et al., 2014, Mares et al., 2005, Flintham et al., 2002).

Selected PHS and PMA QTL were validated in independent isogenic lines

Despite the identification of many PHS and PMA resistance QTL in different studies, only handfuls have been independently validated in isogenic population. This makes it difficult to determine the robustness of these QTL and limits their use in breeding for PHS/PMA resistances. A study by Torada et al. (2008) used BC_1F_3, BC_3F_3 and BC_5F_3 Near isogenic lines to validate the 4AL QTL. Interestingly, the QTL effect was observed more clearly after each successive backcross to the dormant parent. This demonstrates the usefulness of isogenic material in the study of QTL effects.

In this study, we validated the QTL in F_3 populations, which despite not being as isogenic as backcrossed populations, still offered the opportunity for early independent validation of the QTL. The effects of the QTL were not always observed in the locations tested as there were no significant differences between lines with the resistant and susceptible parent alleles in some of the locations tested. However, when differences were observed, the differences were in the right direction i.e. lines with the resistant allele of the QTL showed higher HFN than lines with the susceptible allele. For instance, the 1A and 7B QTL showed resistance effects in two experimental locations (JIC and KWS) and the 2D QTL showed effect in one location (KWS). The $3A_{SR}$ QTL, however, could not be validated as it did not show statistically significant effects on HFN in any of the locations tested. Given that the $3A_{SR}$ showed consistent effects on HFN in DH population during field trial done in 2006-2008, it is likely that the resistance it offers is real.

The F_3 validation experiments also highlight the importance of environmental influence on the expression of PHS and PMA QTL. Although many of the QTL showed effects in more than one experimental location, the effect of the QTL were not seen in all the locations test. This is particularly true for the Limagrain trials where none of the QTL showed any effect on PHS/PMA mainly due to the high HFN value obtained at this location. This can be attributed to the influence of environmental factors in modulating the expression of the genetic resistance offered by the QTL. Such genetic x environmental interactions are common in the expression of PHS and PMA QTL (Knox et al., 2005, Kulwal et al., 2004).

QTL showed distinct effect on dormancy loss and PMA induction

In addition to the validation of the QTL, it is important to understand the mechanistic basis underlying their effects. This will inform ways by which the QTL can be effectively deployed and combined for strategic breeding purposes. The genetic, physiological and agronomic characterisation of the QTL was undertaken in this study to address three specific questions: when do these QTL exert their effects; what are the effects of grain developmental and germination temperature on the expression of the QTL effect; and do these QTL have pleiotropic effects on other agronomic traits?

It is important to conduct these characterisation studies in isogenic genetic material to reduce the complexity associated with studying the effects produced by multiple loci on the same trait. Studies that have previously undertaken physiological characterisation of PHS resistance and dormancy in wheat used diverse parental materials (Gerjets et al., 2010, Walker-Simmons, 1987). Although useful, the effect seen in these studies could be due to multidirectional interaction (positive or negative epistasis) of many different loci and might not be necessarily true of any particular locus. An example is offered by the Avalon x Cadenza population used in this project where three QTL are segregating each with different modes of action.

F_3 lines for each of the QTL studied showed distinct dormancy profiles that are characteristic of each QTL. While the 2D QTL showed an early effect, the 1A and $3A_{SR}$ QTL showed late effects which persisted even 8 weeks after harvest maturity. The 7B QTL showed no effect on the germination kinetics of susceptible and resistant line at most germination temperature and maturation stages tested. Surprisingly, the effect observed in the $3A_{SR}$ F_3 lines is mirrored in the parents but in the opposite direction i.e. the susceptible parent (Rialto) was more dormant than the resistant parent (Savannah). One possible reason for this might be that this effect does not originate from the 3A locus being studied but rather from an independent locus in which Rialto provides the resistance allele.

Nevertheless, these distinct characteristics of the QTL effects suggest that the underlying genes and molecular pathways for the control of dormancy release in

each of the QTL differ. For instance, QTL that confer early effects on PHS resistance could be involved in molecular pathways that establishes or set dormancy early in grain development. One such QTL, *Phs.ocs-3A.1*, was recently cloned in wheat. This gene mapped to chromosome 3A and was found to be a wheat homologue of *Mother of Flowering Time* (*TaMFT*; Liu, et al. 2013; Nakamura, et al. 2011) which is upregulated after physiological maturity in dormant seed. Similarly, late acting QTL could play a role in the regulation of after-ripening kinetics in wheat, discussed in Chapter One (section 1.7.4).

This different timing of action of the QTL presents an opportunity to make better varieties with robust tolerance to PHS by stacking multiple PHS resistance genes which affect distinct pathways. In addition, some of the QTL were found to be effective under at least two germination incubation temperature (17 °C and 22 °C) suggesting that they could be deployed in diverse environmental condition.

The results from the physiological characterisation of NILs did not totally agree with results of F_3 and DH population as many of the QTL did not show statistically significant effect on PHS and PMA. This was partly due (especially for the 1A and 2D QTL) to the heterogeneity in the phenotypes of sister NILs caused possibly by residual polymorphism between them. Another important factor for this could also be that experimental samples for the F_3 and DH line characterisation were from field grown material while the NILs were grown in controlled environment room which does not allow for the integration of environmental clues into the regulation of the QTL expression. Given the strong environmental and genetic interaction in the control of PHS and PMA, it is apt to suggest that future characterisation studies should be done in a combination of different environments and using different assay methods. This will allow for a more robust characterisation of the QTL.

Despite the differences in the behaviour of the CER grown NILs and field grown DH/F_3 materials for some of the QTL, the 4A QTL showed consistent effect in both sets of materials. Although the 4A QTL was not included in the F_3 characterisation experiment, it nonetheless showed effect on sprouting resistance in the DH population and BC_3F_3 NILs. In addition, the 4A QTL produced the largest effect

observed at the post-harvest maturity stages of all the QTL and this was observable when plants were grown at 13 °C and 20 °C.

Most of the QTL tested were found not to be effective when plants were grown at 13°C from anthesis onwards. This was primarily due to the high depth of dormancy observed in both the resistant and susceptible seeds of the QTL. This finding agrees with other studies on wheat where the temperature experienced during seed development affects the level of dormancy in wheat seed (Nyachiro et al., 2002, Reddy et al., 1985).

Also, low temperature during grain development showed effect on the induction of PMA in the 7B susceptible NIL material. There was about a two-fold induction in PMA in the susceptible 7B NILs when grown at 13 °C. This agrees with PMA being cold-inducible (Mrva et al., 2006, Farrell and Kettlewell, 2008). Although the mechanism behind this is not known, some studies show that low temperature increases the sensitivity of seeds to gibberellic acid (Mrva et al., 2006), a hormone known to promote α-amylase production in seeds (Kondhare et al., 2012, Barrero et al., 2013). Farrell and Kettlewell (2008) also suggested that increase in PMA following cold treatment could be due to the alteration of the GA/ABA ratio when plants are grown under cold. Irrespective of the cold induction, the 7B QTL showed potential for PMA resistance in the isogenic material tested but his was not found to be significant due to high variability in the α-amylase activity of seeds. It is therefore suggested that more seeds samples and more replication be used for future α-amylase test.

Agronomic effects of some selected PHS and PMA QTL

Agronomic characterisation of the QTL revealed that some QTL affect plant height and flowering time (ear emergence). Reduced plant height conferred by the *Rht* genes is a desirable trait as it leads, in some cases, to increase in yield and resistance to logging. Three of the four parents of the QTL whose agronomic effects were examined (Rialto, Savannah and Avalon) have the semi-dwarf *Rht-D1b* allele

while the fourth parent (Cadenza) has the tall *Rht-D1a* allele (Addisu et al., 2009, Srinivasachary et al., 2008, Ma et al., 2015).

The 2D and 3A QTL produced significant effects on height. QTL for plant height have been mapped to both QTL intervals in a separate study by Griffiths et al. (2012). In agreement with this, Munkvold et al. (2009) reported the association of a height QTL with the 2D QTL for PHS resistance, and Ren et al. (2008) showed that the markers for *Rht8* flank the 2D PHS resistance QTL. In our study, the increased height allele comes from the resistant 2D QTL allele and the susceptible 3A QTL allele. This implies that the association observed between PHS QTL and plant height is not strictly causal but may be due to genetic linkage disequilibrium.

In addition to plant height, the days to ear emergence is another important adaptive trait which is bred to coincide with favourable environmental conditions that maximise the grain filling potential of the wheat plant. Generally, selection is made for wheat varieties with early ear emergence as these are thought to have prolonged grain filling period which could result in higher yield (Worland, 1996, Griffiths et al., 2009). However, a recent study by Borrill et al. (2015b) suggests that extended grain filling period do not always lead to increased grain weight due to the limitation of grain filling capacity. In our study, the 3A QTL produced additional effects on flowering as well as grain size and TGW.

Despite the effect of some QTL on these agronomic traits, there was no consistent negative effect of the QTL on yield, and this remains the most important consideration. This means that the QTL can be effective deployed to achieve PHS resistance in the UK without negatively affecting the productivity of UK wheat farms. However more field experiments will need to be done to confirm this.

4 Chapter Four: Characterisation and physical map construction of the 4A PHS QTL

Chapter Abstract: In the previous chapter, the genetic, physiological and agronomic characterisation of six QTL that confer resistance to PHS and/or PMA was performed. This identified a QTL on chromosome arm 4AL as having a consistent effect on PHS resistance in UK germplasm. In this chapter, we have fine-mapped the 4AL QTL to a 0.2 cM interval, constructed a physical map of this interval, and identified 17 high confidence genes in this region. Further physiological and genetic characterisation of the 4AL QTL, showed the QTL effect to originate from zygotic tissues and also highlight a possible hormonal interaction in the expression of the QTL effect. These characteristics of the 4AL QTL effect, as well as transcriptional profiling of genes in the physical map, identified a copy of the *ABA-inducible plasma membrane protein 19* (*PM19-A2*) as a main candidate gene for this QTL.

4.1 Introduction

4.1.1 The 4A PHS QTL affects after-ripening kinetic of seeds

The 4AL QTL has been widely reported to have an effect on PHS resistance in many different studies (Mares and Mrva, 2014). It has been identified in several germplasm and in many locations across the world. In many cases, it accounts for a significant proportion of the genetic variation observed in the populations studied. The 4A QTL was first reported by Flintham (2000) as a major PHS resistance locus and was named ***Phs***. Flintham and colleagues subsequently mapped this locus to the long arm of chromosome 4A (Flintham et al., 2002). Based on comparison with the wheat consensus genetic maps (Somers et al., 2004), this locus co-localises to the 4A QTL previously described in chapter three. For this reason, the 4A QTL will hereafter be referred to as *Phs*.

Despite its importance, *Phs* has only been fine-mapped in several studies but has not been cloned. Importantly, the mechanisms by which *Phs* functions are not fully

understood. However, previous physiological characterisation done in this study has shed new light on a possible mechanism by which *Phs* functions. This is because the QTL was found to be only effective at post-harvest stages, after a period of dry after-ripening. This implies that *Phs* plays a role in modulating the after-ripening pathway in wheat. Furthermore, this after-ripening effect was observed in plants grown at 13 °C and 20 °C post anthesis highlighting the robustness of *Phs* effect under different environmental condition.

Although this characterisation sheds new light on the mode of action of *Phs*, the molecular mechanism underlying this effect is still largely unknown and the identity of the causal gene underlying this effect remains elusive. This limits the effective deployment of *Phs* into modern wheat varieties. This chapter aims to achieve further physiological and genetic characterisation of *Phs* effect. In addition, the gene content of the *Phs* interval will be evaluated by several approaches including: fine-mapping, syntenic comparison and the development of the *Phs* physical map.

4.1.2 Physical map construction in polyploid wheat species

Several molecular and genetic methods and resources exist for accessing genic information about any particular chromosome interval of interest. However, many of these methods, including genetic map construction, only represent the position of loci in relation to other loci and as such do not accurately represent the actual physical DNA or genic sequence around any particular loci. Such physical DNA sequence information can only be accessed by the construction of a physical map.

A physical map is a systematic ordering of overlapping DNA fragments in which distance between loci are expressed in actual nucleotide counts, base pairs (bp; Meyers et al., 2004). Physical map construction usually starts with the enzymatic and/or physical digestion of the whole genome followed by the cloning of the produced DNA fragments into individual Bacterial Artificial Chromosome (BAC) to make a BAC library. Overlaps between BACs can then be established by searching for the presence of common DNA/marker sequence between BACs (hybridization mapping). Alternatively, DNA fingerprinting, which is a whole genome restriction

mapping approach, can be used to infer connections between BACs based on the assumption that overlapping DNA fragments produce similar pattern of restriction fingerprint upon digest.

Automated assembly of overlapping BACs from restriction fingerprints can be achieved using the Fingerprinted Contig (FPC; Nelson and Soderlund, 2009) or Linear Topography Contig (LTC; Frenkel et al., 2010) software. These software use different statistical methods to validate overlap between BAC and to produce a network of all possible connections between BAC. Reduced representation of the overlapping BACs cluster is achieved by selecting for a minimum set of overlapping BAC clones needed to cover the entire target region; this is called the Minimum Tilling Path (MTP).

For many QTL cloning efforts in wheat, the development of a physical map was instrumental for the identification of the causal gene underlying the QTL effect. Within the last decade, this approach has been used to clone QTL for (1) disease resistances - *Yr36* (Fu et al., 2009), *LR34* (Krattinger et al., 2009), *Sr35* (Saintenac et al., 2013b), *Sr33* (Periyannan et al., 2013); (2) grain protein content - *GPC* (Distelfeld et al., 2006); (3) seed dormancy - *TaMFT* (Liu et al., 2013b) and (4) vernalization – *Vrn4* (Kippes et al., 2015), among others. In many of these studies, the physical map used were developed using the varieties from which the target QTL were detected. Alternatively, already developed physical map resource for the Chinese Spring reference variety can be used as this could save time and the cost associated with constructing a new physical map. However, the suitability of the Chinese Spring genotype for the study of the target trait should be considered before using the Chinese Spring BAC library.

4.1.3 Genome sequencing of the 4A wheat chromosome reveals extensive translocations

Recent advances in next-generation sequencing coupled with the use of the mitotic chromosome flow-sorting techniques to separate individual wheat chromosomes (Šafář et al., 2004), have enabled the sequencing of the 4A wheat chromosome

(Hernandez et al., 2012). With a size of 856 Mb (1C), the 4A wheat chromosome is estimated to contain about 4,300 genes.

The 4A chromosome is also unique because it underwent the most extensive intra and inter-chromosomal rearrangements compared to other chromosomes of the bread wheat genomes (Figure 4.1; Devos et al., 1995, Miftahudin et al., 2004, Liu et al., 1992). These include a pericentric inversion between the native short arm ($4AS_{native}$) and the native long arm of the chromosome ($4AL_{native}$). This resulted in most of the $4AS_{native}$ sequences being present in the modern wheat 4AL chromosome and vice versa (Figure 4.1). Following this intra-chromosomal rearrangement, there were two cyclical translocation events from the 5AL and 7BS chromosome into the modern 4AL chromosome.

Based on these various rearrangements, the modern 4A chromosome has been divided into 5 distinct chromosomal segments A - E (illustrated in Figure 4.1; Hernandez et al., 2012). Segments A and B represent the pericentric translocation events while segments D and E represent the cyclical translocations from the 5AL and 7BS chromosome arms, respectively. segment C is the only 4AL region unaffected by the various translocations. However, recent studies are suggesting that these translocation events may even be more complex and extensive than previously thought (Ma et al., 2013, Ma et al., 2014). These rearrangements might have implication on the effort to fine-map *Phs*, as we show in section 4.2.3 of this chapter that *Phs* is located in segment D of the 4AL chromosome arm.

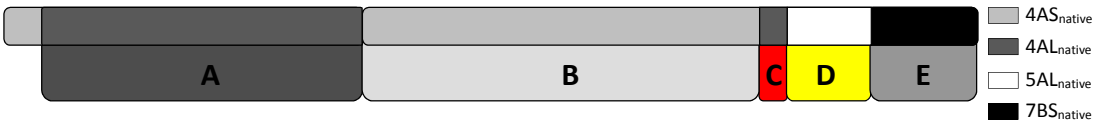


Figure 4.1: The modern 4A wheat chromosome is a product of ancient translocation events. The modern 4AL chromosome has undergone various translocations which include: sequences from the $4AL_{native}$ being transferred to modern 4AS (segment A); the $4AS_{native}$ sequencing translocated to modern proximal region of the modern 4AL (Segment B); Cyclical translocation from $5AL_{native}$ (Segment D) and $7BS_{native}$ (Segment E). Region C and the distal end of the $4AS_{native}$ region represent regions unaffected by the translocations. Adapted from Hernandez et al. (2012)

4.1.4 Aims

Given the significance of *Phs* on sprouting resistance in the UK and worldwide, it is vital that the gene(s) underlying this QTL be identified. This will facilitate the deployment of this QTL for PHS resistance but will also shed new light on the regulation of dormancy in cereals. In view of this, the aims of this chapter are to investigate the underlying gene(s) for *Phs* effect. To do this, we aim to fine-map the QTL to a narrow genetic interval. Secondly, depending on the resolution of fine-mapping achieved, we aim to examine the physical map of the *Phs* interval so as to access the full gene content of this critical QTL.

4.2 Results

4.2.1 Characterisation of F_1 seeds defines the inheritance and spatial pattern of *Phs* effect in grains

The *Phs* QTL identified in UK germplasm has hitherto only been characterised in DH and isogenic lines which have fixed resistant or susceptible *Phs* alleles across the QTL. These do not allow for the understanding of the genetic relationship (dominance /recessive) between the alleles found in the contrasting parents. Such genetic information can only be obtained from heterozygote F_1 seeds. Furthermore, by studying the phenotype of F_1 seeds made from reciprocal crosses of the parents, it is possible to determine the grain compartment where the *Phs* effect is expressed. This could either be in maternal tissues like the seed coat and pericarp which display maternal genotype, or in zygotic tissues like the diploid embryo and the triploid endosperm which have a combination of both maternal and paternal genomes in different proportions (Evers and Millar, 2002).

Therefore, to understand the genetic relationship between the *Phs* alleles and also to delimit the *Phs* effect to a specific grain components, a reciprocal cross between the susceptible Claire and resistant Option parents was done with both parents used as pollen donor (paternal) or receiver (maternal) in independent crosses (illustrated in Figure 4.2). Seeds from self-pollinated Claire and Option plants with clipped glumes were used as controls. We hypothesise that a *Phs* effect originating from maternal tissues will result in the F_1 seeds showing their respective maternal parent phenotype. While a *Phs* effect originating from a zygotic tissue will result in all the F_1 seeds from both directions of the reciprocal cross showing a similar phenotype (Figure 4.2).

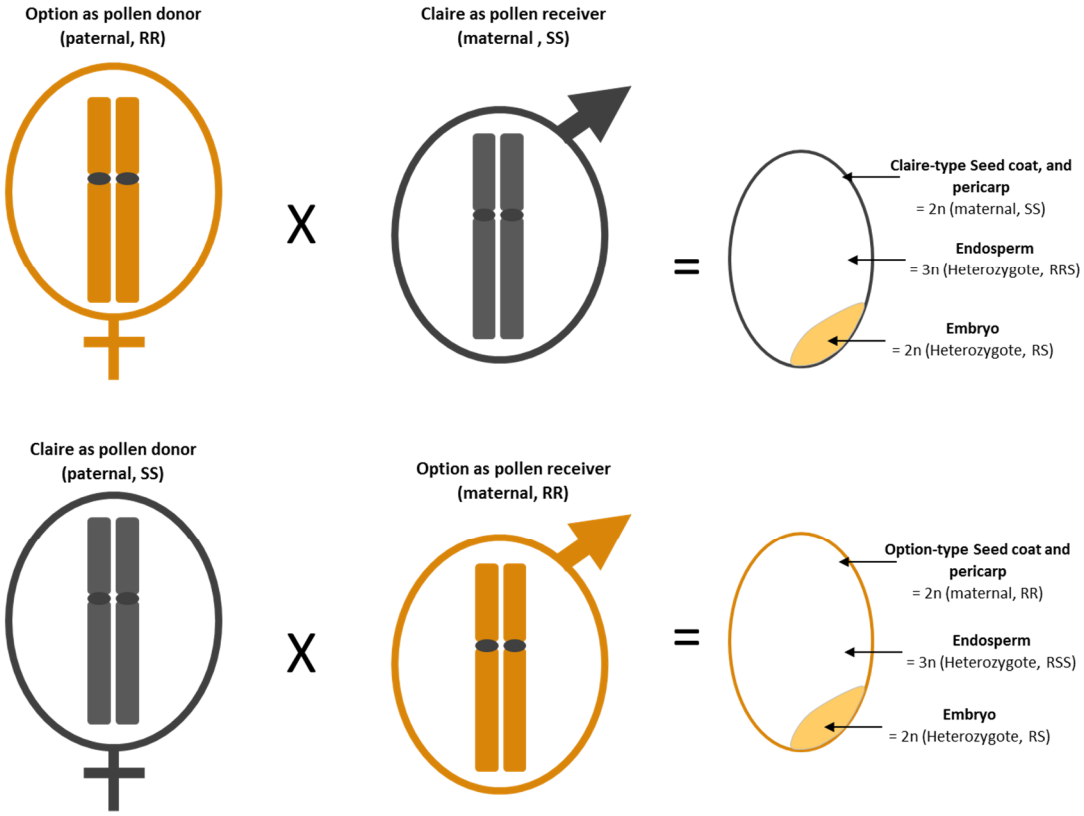


Figure 4.2: Reciprocal crossing of Claire and Option to define the location of the *Phs* effect. The reciprocal cross between Claire (grey) and Option (amber) was done with both parents supplying the maternal (oval with cross base) and paternal (oval with arrow) gametes in independent crosses. The structural and genotypic composition of the different compartment of the derived F_1 seeds in each cross are also indicated

From the reciprocal cross, we obtained 371 F_1 seeds from 10 independent Claire plants used as maternal parents and from another 11 independent Option plants used as maternal parents. However, some of these did not have sufficient seed set. Only crosses with at least 18 seeds set were used for GI phenotyping. The F_1 seeds from self-pollinated Claire and Option plants displayed the expected phenotypic difference, with Claire seeds having a significantly higher germination index compared to Option seeds (Figure 4.11). Irrespective of the parent used as the pollen receiver (maternal), the F_1 seeds from the reciprocal crosses showed similar GI phenotype. There was no statistical significant GI difference between F_1 seeds made from the maternal Claire and maternal Option crosses (P= 0.117). This suggests that the *Phs* effect is not expressed in maternal tissues but is zygotic, expressed in either the endosperm or the embryo.

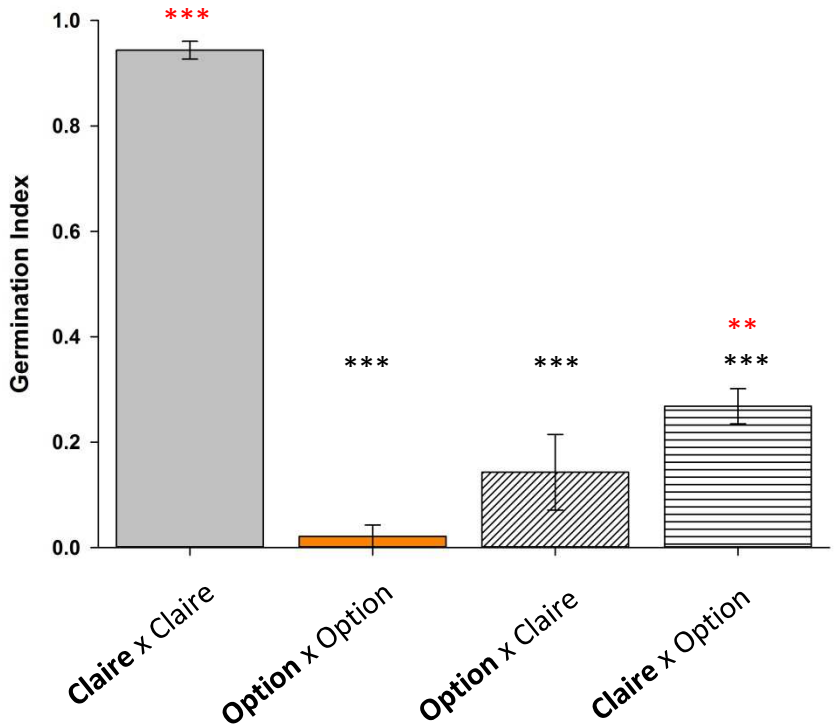


Figure 4.3: F_1 seeds from Claire and Option crosses show in-between GI phenotype: Germination index phenotype of the F_1 seeds derived from self-pollinated Claire (grey bar) and Option (amber bar) plants, as well as, cross pollinated plants (striped bar). The maternal parent in each cross is indicated with bold text. Error bars represent SEM of at least 3 biological replications. Significance value at $P < 0.001$ (***) and $P < 0.01$ (**) relative to the GI of the self-pollinated Claire (black asterisk) or Option (red asterisks) crosses are indicated.

In addition, the *Phs* phenotypes of F_1 seeds made from the reciprocal cross were found to be intermediate (used here to describe an in-between, not mid-way, phenotype) in relation to the phenotype of F_1 seeds from self-pollinated parents. F_1 seeds from Claire maternal cross showed more than a three-fold reduction in the rate of germination compared to the self-pollinated Claire plants, but they still showed significantly higher GI than seeds from self-pollinated Option plants. Similarly, F_1 seeds from the Option maternal cross had significantly lower GI than seeds from self-pollinated Claire plants and also a higher average GI compared to seeds from self-pollinated Option plant, however the later GI difference was not significant.

4.2.2 Hormonal characterisation of *Phs* highlights the possible role of its underlying gene

Seed germination and dormancy are regulated by plant hormones. Abscisic acid (ABA) promotes seed dormancy while gibberellic acid (GA), ethylene and brassinosteroids promote germination of seeds (Miransari and Smith, 2014). Since *Phs* affects the dormancy status of seeds, it is possible that the *Phs* effect could be mediated directly or indirectly by one of these hormones or by an interaction between different hormonal pathways. Therefore understanding the role played by some of these hormones in the *Phs* effect could further help in defining the function of the causal gene(s).

To investigate the possible role of hormones on the *Phs* effect, the GI phenotype of after-ripened seeds from Claire and Option incubated in different concentrations of germination stimulating hormones (or hormone inhibitors) was evaluated. Only treatments that promote seed germination were used so as to observe the effect of such treatments on the already dormant Option seeds. These included treatments with different concentrations of GA; Aminocyclopropane Carboxylic Acid (ACC) - a precursor of ethylene (Yoon and Kieber, 2013); Fluridone (Flu) - an inhibitor of ABA biosynthesis (Gamble and Mullet, 1986) and a water control treatment.

The *Phs* effect was observed between Claire and Option seeds in all the hormone treatments tested, as well as in the control (water) treatment (Figure 4.4). However, the various hormones showed different effects on the germination rate of Option and Claire. The GI of Option seeds incubated with different concentrations of ACC (100 and 500 µM) were as low as the GI of the Option water control. In contrast, both GA and fluridone-incubated Option seeds showed GI increases. Compared to the water control, 50 µM GA treatment did not significantly increase the GI of Option seeds, while there was significant GI increase with 250 µM GA treatment relative to the control ($P = 0.013$). Fluridone treatment, showed a dosage-dependent effect on the germination rate of Option seeds, with a 250 µM treatment showing three-times the GI of the 50 µM treatment ($P = 0.002$). In addition, incubation with 250 µM of fluridone significantly reduced the GI of the

non-dormant Claire seeds compared to the Claire control (P =0.001). This reduction in Claire's germination rate was not observed with the other hormone treatments.

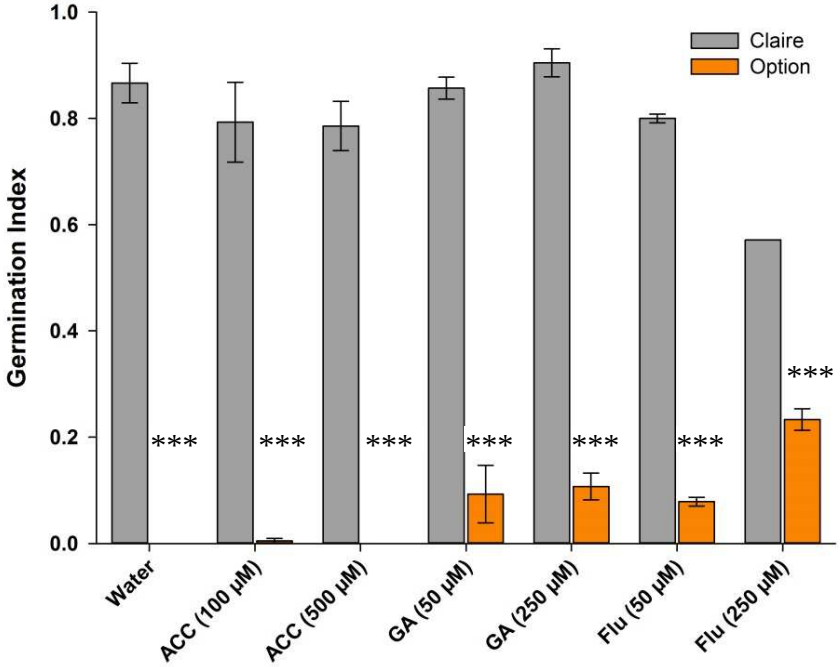


Figure 4.4: Hormonal Characterisation of *Phs* effect: Germination index phenotype of Claire (grey) and Option (amber) seeds incubated in different concentration of ACC, GA and Fluridone (Flu). Error bars represent SEM of at least 3 biological replications. *** represent P < 0.001 significance value between varieties for each treatment.

Given the effect of the individual GA and fluridone treatments on the germination of Option seeds, it is possible that combining both treatments could further improve the germination rate of Option seeds. To confirm this hypothesis, a second experiment was conducted to examine the effect of combining both treatments on the GI of Claire and Option. In addition, the effect of ABA was also tested so as to understand better the observed negative effect of fluridone on Claire's seed germination.

Due to a slightly longer after-ripening period, the seeds used in this second experiment showed a slight decrease in dormancy compared to those in the previous experiment. This was evident in the GI of seeds imbibed in water (especially in Option; Figure 4.5). Nevertheless, the *Phs* effect was still detectable between Claire and Option irrespective of the various hormones treatments. Incubating seeds in GA significantly reduced this *Phs* effect by promoting the

germination of Option seeds (Figure 4.5). Although fluridone treatment seems to increase the germination of Option, this increase in GI was not found to be significant due to the large variability of the GI of fluridone-treated Option seeds. As was hypothesised, incubation of seeds in a combination of GA and fluridone further increased the germination rate of Option seeds thereby reducing the effect *Phs* further (Figure 4.5). This suggests that GA and fluridone have an additive effect when used in combination.

Unlike in the previous experiment, fluridone incubation only slightly reduced the GI of Claire seeds compared to the control. ABA, on the other hand, did not have any effect on Claire germination rate, but it instead reduced the GI of Option thereby increasing the *Phs* effect. Taken together, these results suggest a possible involvement of GA and/or ABA in the expression of the *Phs*.

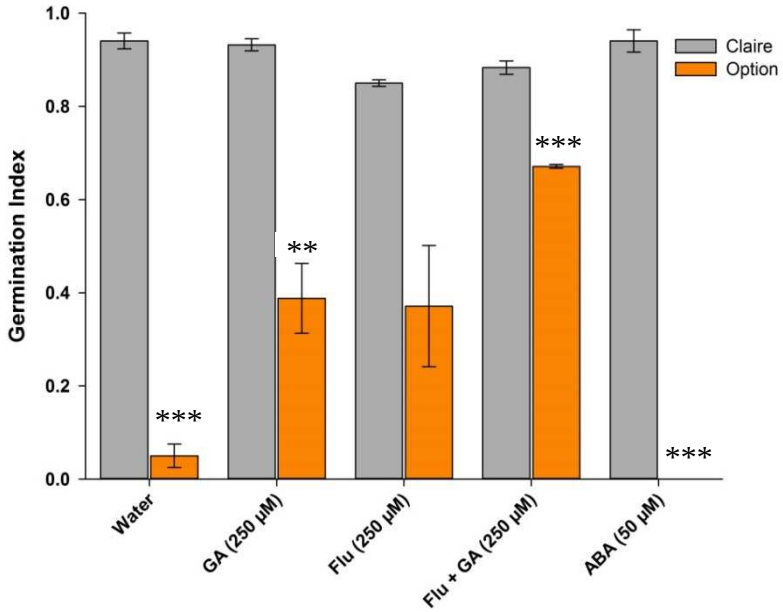


Figure 4.5: GA and Fluridone reduce the effect of *Phs*: Germination index phenotype of Claire (grey) and Option (amber) seeds incubated in different concentration of GA, Fluridone, Fluridone + GA and ABA. Error bars represent SEM of at least 3 biological replications. ** and *** represent significance value between varieties for each treatment at $P < 0.01$ and $P < 0.001$ respectively.

4.2.3 *Phs* maps to a sub-cM Interval

Physical map construction of the *Phs* interval will facilitate the identification of the causal gene(s) responsible for its effect. However, before the physical map construction of the *Phs* QTL can be undertaken, it is important to fine-map the locus to a very small genetic interval, preferably less than a cM (sub-cM). This will reduce the time and costs associated with the physical map construction of the whole QTL interval and increase the likelihood of successfully covering the interval. In the UK, the *Phs* locus was originally identified in two independent DH populations made from Alchemy x Robigus and Option x Claire crosses (referred to as AR Pop. and OC Pop. afterwards), with the beneficial (resistant) *Phs* alleles coming from Alchemy and Option. This provided an opportunity to fine-map the locus independently in both populations. Preliminary fine-mapping of *Phs* performed in the AR and OC Pop. are presented below.

4.2.3.1 *Phs* is linked to the genetic interval between SSR markers *barc170* and *wmc420* in the AR Pop.

Initial fine-mapping of the *Phs* locus was first done in the AR Pop. This was done in Near Isogenic Lines (NILs) developed through a backcrossing scheme described in Chapter Two, section 2.1.1, (Figure 2.1). NILs with overlapping recombination across the entire length of the 4AL chromosome arm were selected using five SSR markers distributed across the consensus chromosome 4AL genetic map (Somers et al., 2004). Through this approach, it was possible, to divide the resistance introgression from Alchemy into smaller chromosome segments and to examine the PHS resistance effect provided by each segment. From this schemes, 14 NILs representing 7 independent but overlapping recombination haplotypes (designated Rec Groups 1 - 7) were developed (Figure 4.6). Two independent NILs were developed for each recombination haplotype.

	BC_3F_2 Pedigree number	*wmc313*	*wmc760*	*wmc707*	*barc170*	*wmc420*
Susceptible Parent	Robigus	R	R	R	R	R
Resistant Parent	Alchemy	A	A	A	A	A
Rec Group 1	RRA 33/1/12	R	R	R	R	R
	RRA 75/5/20	R	R	R	R	R
Rec Group 2	RRA 33/1/18	R	R	R	R	A
	RRA 33/1/16	R	R	R	R	A
Rec Group 3	RRA 45/3/6	R	R	R	A	A
	RRA 45/3/15	R	R	R	A	A
Rec Group 4	RRA 73/2/8	R	R	A	A	R
	RRA 73/2/10	R	R	A	A	R
Rec Group 5	RRA 72/2/15	R	A	A	R	R
	RRA 72/2/19	R	A	A	R	R
Rec Group 6	RRA 75/5/5	A	A	R	R	R
	RRA 75/5/3	A	A	R	R	R
Rec Group 7	RRA 75/7/7	A	R	R	R	R
	RRA 75/7/8	A	R	R	R	R

Figure 4.6: Recombination haplotypes in the AR NIL population: Following a backcrossing scheme that used Robigus as the recurrent parent, five SSR markers were used to select for NILs with overlapping recombination blocks across the 4AL chromosome arm. Seven recombination haplotypes (Rec Group 1 - 7) were selected for, with two independent NILs per haplotype. The resistant (Alchemy) and susceptible (Robigus) parents alleles are represented by A (amber) and R (grey), respectively.

The understanding gained from the previous physiological characterisation experiment helped to inform the correct phenotyping strategy for the selected NILs. Since *Phs* was found to only act after a period of dry after–ripening, harvested spikes from the 14 selected NILs, as well as from the parents, were allowed a period of dry after-ripening before the induction of sprouting through over-misting in a sprouting chamber (See Chapter Two, section 2.6.4). This experiment was primarily conducted by Dr John Flintham (a consultant on this project) at the start of this Ph.D. project.

As expected, a good level of sprouting induction was observed in the susceptible parent Robigus, while Alchemy - the resistant parent - showed a significantly lower level of sprouting compared to Robigus (Figure 4.7 A). Importantly, there was little variance between the independent NILs for each recombination group. This further confirmed the genetic homogeneity of these sister NILs as opposed to the NILs used for the characterisation of some of the QTL in Chapter three (section 3.2.4.3).

Many of the NILs (from Rec Groups 2, 5, 6 and 7) showed susceptible sprouting phenotype similar to Robigus. NILs of Rec Groups 1 and 4 showed evidence of transgressive segregation because they showed a higher level of sprouting than the susceptible parent - Robigus. Interestingly, NILs belonging to Rec Group 3 showed a very high level of sprouting resistance similar to what was observed in Alchemy. This suggests that the effective *Phs* locus provided by the Alchemy introgression is present only in NILs from Rec Group 3.

I afterwards confirmed the observed sprouting phenotype of the NILs and parents reported above in an independent germination index experiment using the progeny of the NILs. As was observed in the sprouting experiment, the parents showed highly significant GI difference (GI difference of 0.8, $P<0.001$; Figure 4.7 B). Many of the NILs showed GI scores similar to Robigus implying that these NILs are susceptible to PHS like Robigus. Similar to the results of the sprouting experiment, NILs from Rec group 3 showed significantly reduced GI score (0.4) compared to the GI of Robigus ($P<0.001$). NILs of the recombination Group 6 also showed significant less GI than the susceptible control. However, the reduction in GI values was not as strong as was observed for the NILs of Rec Group 3.

Based on the phenotype data obtained from the sprouting and GI experiment, lines were classified as resistant only if they showed significantly lower sprouting to Robigus in both experiments. Using this criteria, NILs of the recombination Group 3 were classified as resistant while the other NILs, including those of Rec Group 6 which only showed lower GI but similar sprouting phenotype to Robigus, were classified as susceptible.

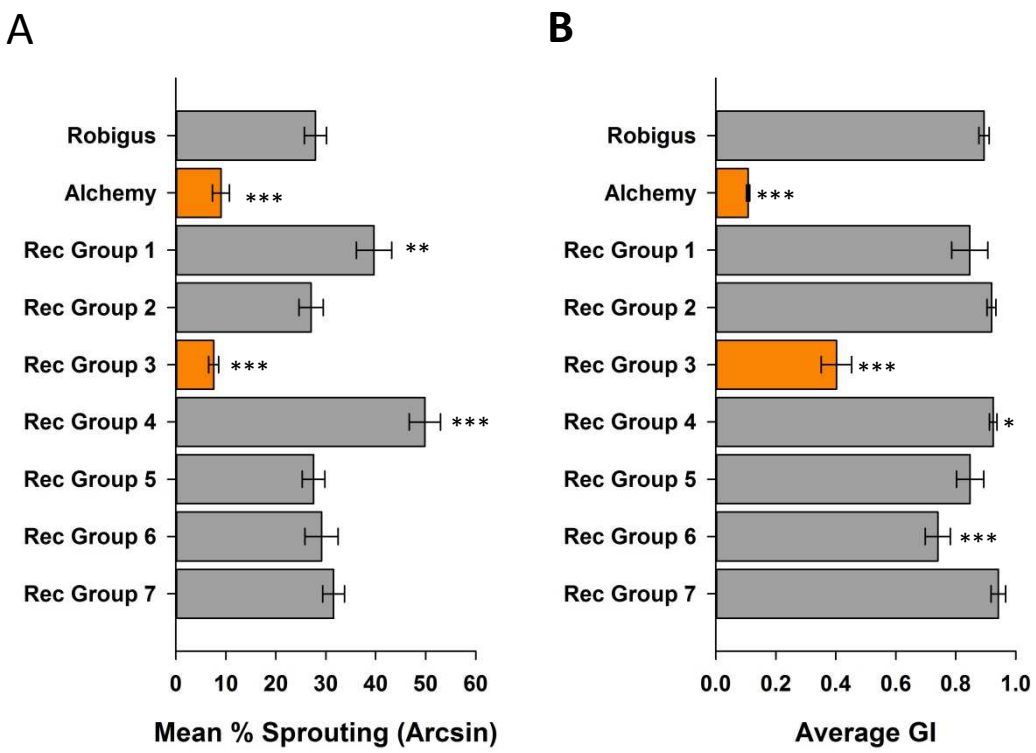


Figure 4.7: Fine-mapping of *Phs* in the AR NIL Population: Sprouting (A) and Germination index (B) phenotype of NILs containing the seven different recombination haplotypes as well as parents. Bars in both panels represent average values of two independent NILs per recombination group. NILs with susceptible or resistant phenotype are shaded grey and amber, respectively. Error bar represent SEM of at least 6 biological replications. Significance value at $P < 0.001$ is represented as *** and is based on comparison to the phenotype of the Robigus parent

The recombination in the Rec Group 3 NILs is flanked by SSR markers *barc170* and *wmc420*. However, *Phs* is not linked to either *barc170* or *wmc420*. This is because NILs from Rec Group 2 and 4 also have Alchemy allele at these markers loci, yet they show a sprouting susceptible phenotype similar to Robigus. This implies that *Phs* maps within the genetic interval delimited by *barc170* and *wmc420*. Based on the Somers et al. (2004) consensus map, the genetic distance between *barc170* and *wmc420* is approximately 21 cM. This is still too large a genetic interval for the construction of the physical map and so further fine-mapping of the *Phs* locus was done in the OC Pop.

4.2.3.2 *Phs maps to a 0.2 cM interval in the OC Pop.*

A different fine-mapping approach was adopted for delimiting the *Phs* locus in the OC Pop. Since the *Phs* locus has been identified and fine-mapped in many other studies, we took advantage of the reported linkage of *Phs* with some diagnostic SSR markers to advance the fine-mapping of this locus in our experimental population.

One of such studies reported by Torada et al. (2008) delimited *Phs* to be between markers *barc170* and *xhbe03*. This *Phs* genetic position agrees well with our earlier fine-mapped position in the AR Pop. in which *barc170* flanked the *Phs* locus. Also, the reported *barc170* and *xhbe03* markers are polymorphic in UK germplasm. Thus, by combining this genetic information, it was possible to delimit further the *Phs* locus to a narrower genetic interval than was previously achieved in the AR Pop.

To confirm the linkage of *Phs* to the genetic interval between *barc170* - *xhbe03* in UK germplasm, these two markers along with *wms894* (another polymorphic marker that maps in the same 4AL_13-0.59-0.66 physical bin as *barc170*) were used in developing F_4 Recombinant Inbred Line (RIL) created from the cross between Claire and CO69 - a DH line from the initial Option x Claire population. The development of this material was done in collaboration with our breeding partners (RAGT, KWS and Limagrain) and the University of Nottingham. In total, about 145 F_4 RILs were developed; however, only 131 of these were retained due to the loss of some lines during the RILs development process. The 131 F_4 RILs originated from 27 independent F_2 recombinant lines and had a series of recombination events between the three SSR markers which are summarised in Table 4.1. Based on the average recombination frequencies in the OC DH and RIL Pop. the genetic distance between *xhbe03* and *wms894* was calculated to be 0.2 cM; while *barc170* was found to be 2.4 cM proximal to *wms894* (Figure 4.8).

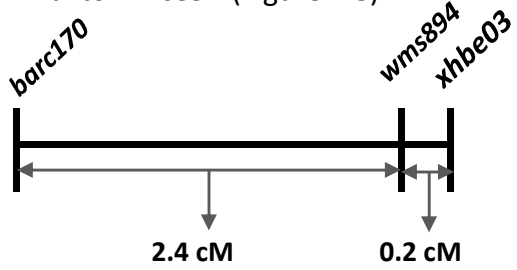


Figure 4.8: Genetic map of *Phs* SSR markers: The genetic position of SSR markers used to define the *Phs* locus. The genetic distances between these markers are indicated.

Table 4.1: Haplotype frequencies in F_4 CO RIL Population

Haplotype (*barc170-wms894-xhbe03*)	No of lines per haplotype
OCC	20
CCC	36
CCO	14
OOC	11
HOO	2
COO	18
OOO	28
HOC	1
OHC	1
Total	131

O and C denote Option and Claire alleles, respectively, while H denotes heterozygous.

Using a similar phenotyping strategies employed for the AR NILs, the sprouting characteristics of 128 homozygous RIL was examined. RILs with heterozygous genotype for any of the markers were not phenotyped. However due to the large number of lines, limited biological replications (two) of each RIL were used.

A high level of sprouting variation was observed across the entire RIL population resulting in a large coefficient of variation of 89.67 % across the experiment. However, interval mapping across the 2.6 cM interval between *barc170* and *xhbe03* showed that a significant amount of this variation can be accounted for by the interval genotypes of the RILs (Figure 4.9). Comparison of RILs with homozygous non-recombinant haplotype across the 2.6 cM interval between *barc170* and *xhbe03,* showed a highly significant difference between the susceptible (Claire) and resistant (Option) allele groups. RILs with the Option allele across this interval were highly resistant to sprouting while RILs with the Claire susceptibility allele showed significant level of sprouting. This confirms that *Phs* is linked to the *barc170* and *xhbe03* genetic interval in UK germplasm.

Furthermore, recombination at the central *wms894* marker further divides the *barc170* - *xhbe03* interval allowing for the delimiting of *Phs* to a smaller genetic interval. No significant difference was observed between RILs with Claire and Option haplotype across the *barc170* – *wms894* interval, as both allele groups showed a high level of sprouting. On the other hand, there was a highly significant difference between RILs with the Claire and Option alleles across the *wms894* - *xhbe03* interval; RILs with Option allele in this interval showed a higher level of sprouting resistance.

This implies that the effective *Phs* locus can be narrowed down to the 0.2 cM interval between *wms894* and *xhbe03*. In comparison to the 21 cM interval in which *Phs* was delimited in the AR Pop, this represents a 105 fold increase in mapping resolution of the *Phs* locus. More importantly, the *Phs* interval has been reduced to a genetic distance amenable to physical mapping.

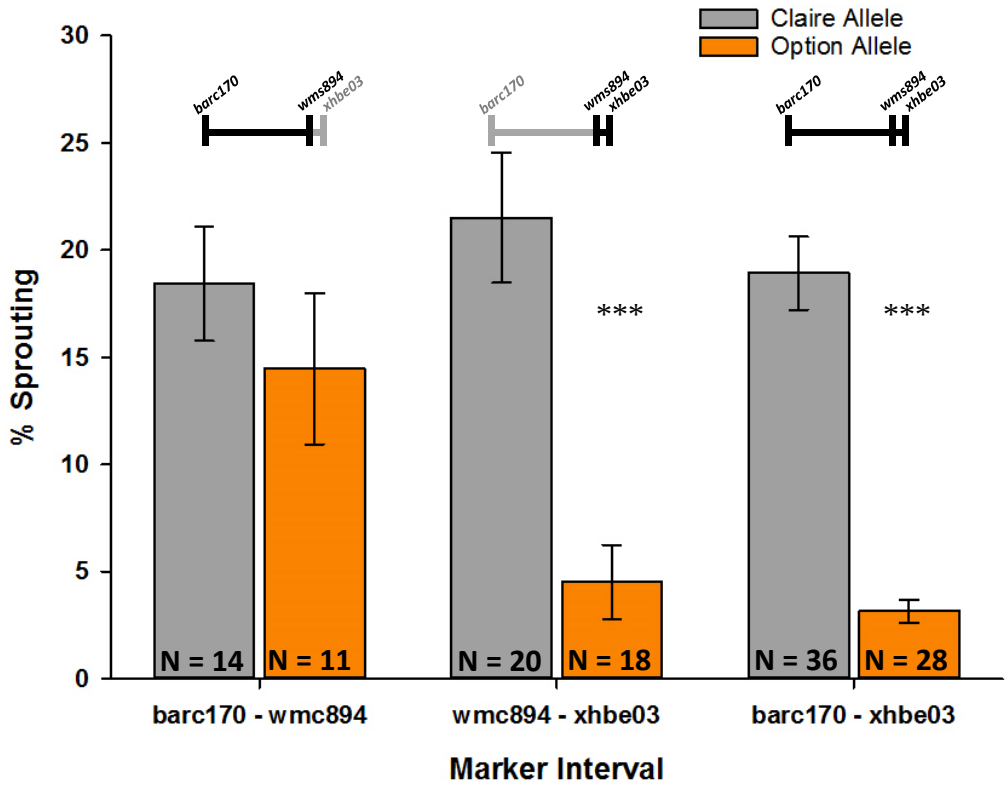


Figure 4.9: Interval mapping of the *Phs* interval in the CO population. The sprouting phenotype of 128 RILs with recombination between markers *barc170*, *wms894* and *xhbe03*. The RILs are grouped based on their haplotype across the highlighted markers in the genetic maps above each category. Claire (susceptible; grey bars) and Option (resistant; amber bar) alleles were examined in each linkage group. Error bar represent SEM of N, where N is number of RIL per haplotype. Significance value at P < 0.001 is represented as ***.

4.2.4 Synteny reveals the putative gene content of the sub-cM *Phs* interval

Synteny provides an ideal opportunity to access gene content information of specific chromosome segments in grass species with unsequenced or unassembled genomes. This is achieved by the comparison of the target interval with the corresponding interval in other grass species, like rice, sorghum and *Brachypodium*, which have sequenced and assembled genomes (The International *Brachypodium* Initiative, 2010; Paterson, et al., 2009; International Rice Genome Sequencing Project, 2005). Although it is not a member of the Triticeae, *Brachypodium* is often considered as the ideal model species for genetic and syntenic studies in the Triticeae subfamily (Brkljacic et al., 2011, Vain, 2011). This is because of its close phylogeny to members of the Triticeae subfamily and the relative completeness of its genome sequence.

In view of this, a comparison of the *Phs* interval with homologous interval in *Brachypodium* was done. To do this, probe sequences of markers *xhbe03* and *wms894* were obtained from the Torada et al. (2008) study and from RAGT Seed (*wms894* being a proprietary RAGT marker), respectively. These were used as queries in a BLAST analysis against the IWGSC CSS chromosome arms sequence. Wheat contigs with the highest significant hit to these queries were retrieved.

The sequences for *xhbe03* and *wms894* were found to be contained in wheat contigs 4AL_7174272 and 4AL_7172877 (shorthand nomenclature for IWGSC_CSS_4AL_scaff_7174272 and IWGSC_CSS_4AL_scaff_7172877), respectively. Based on the high-density wheat genetic map produced by population sequencing (POPSEQ; Chapman et al., 2015), these two contigs mapped in the same recombination bin (103.62 cM). This further supports the close proximity of the markers on the OC Pop. genetic map.

The retrieved 4AL_7174272 and 4AL_7172877 contigs sequences were subsequently used to query the *Brachypodium* genome database at Phytozome website (http://phytozome.jgi.doe.gov/pz/portal.htmL), and orthologous hits were obtained to *Bradi1g00600* and *Bradi1g00720*, respectively. This suggests that there are just 13 genes in the orthologous *Phs* interval in *Brachypodium*. Assuming

perfect collinearity between wheat and *Brachypodium*, it can be inferred that wheat orthologues of *Brachypodium* genes *Bradi1g00607* to *Bradi1g00710* should map between *xhbe03* and *wms894* markers.

Therefore using the sequence information of *Brachypodium* genes *Bradi1g00607* to *Bradi1g00710*, a BLAST search was done against the IWGSC CSS assembly to retrieve orthologous wheat genomic contigs with at least 80 % nucleotide sequence conservation. The contigs obtained from the BLAST analysis were annotated, and gene models information for these contigs was obtained by BLAST analysis against the Ensembl Plant *Triticum aestivum* server (http://plants.ensembl.org/Triticum_aestivum). The BLAST analysis returned orthologus wheat hits for *Bradi1g00600, Bradi1g00607, Bradi1g00620, Bradi1g00670, Bradi1g00677, Bradi1g00690, Bradi1g00700* and *Bradi1g0710*. The corresponding IWGSC contig and Ensembl gene model information are presented in Table 4.2 and Figure 4.6. The putative functions of the gene models were obtained from a protein similarity search against the non-redundant protein dataset of the National Centre for Biotechnology Information (NCBI) database, and these were used to assign names to the genes (Table 4.2).

No orthologous wheat gene model could be found for *Brachypodium* genes *Bradi1g00630* to *Bradi1g00660* (Figure 4.10). A similar search against the barley database did not return any orthologous sequence hits either, implying the absence of homologous *Bradi1g00630* to *Bradi1g00660* genes in the Triticeae genomes. This has wider implication on the effort in accessing the entire gene content of the *Phs* interval. Although perfect collinearity between *Brachypodium* and wheat is assumed, it is possible that the gene content in the wheat *Phs* interval is not fully represented in *Brachypodium*. This could be due to differential gene loss or duplication events in different grass species (Colasuonno et al., 2013). This represents one of the limitations of the use of synteny to infer gene content between species (Gale and Devos, 1998).

Table 4.2: Gene content of the *Phs* interval as inferred from *Brachypodium* synteny

	Brachypodium gene	IWGSC CSS Contigs	Ensembl Gene/NCBI ID	Gene Name
1	*Bradi1g00600*	4AL_7174272	Traes_4AL_F99FCB25F (90%)	*Plasma membrane protein, PM19-A2*
2	*Bradi1g00607*	4AL_7174272 4AL_7093716	Traes_4AL_DD1B27086* (90%)	*Myosin-J heavy chain protein*
4	*Bradi1g00620*	4AL_ 7074626	M28059# (86%)	*Ubiquitin Conjugating Enzyme*
5	*Bradi1g00670*	4AL_7095843	Traes_4AL_F18648C49 (90%)	*Ethylene Responsive Factor_B (ERF_B)*
6	*Bradi1g00677*	4AL_6230861 4AL_5621077	Traes_4AL_02AE47773* (92%)	*Activating Signal Co-integrator*
7	*Bradi1g00690*	4AL_7159962	Traes_4AL_2EBB63DBA (90%)	*Serine threonine Protein Phosphatase PP1-Like*
8	*Bradi1g00700*	4AL_7162562	Traes_4AL_C56125840 (92%)	*Phosphate Transporter*
9	*Bradi1g00710*	4AL_7177172	Traes_4AL_DD344FC87 (87%)	*Hypothetical protein*
10	*Bradi1g00720*	4AL_7172877	Traes_4AL_F00707FAF (91%)	*OTU cysteine Protease*

*Gene model is split across two contigs.
NCBI gene model used as no suitable gene model was found on Ensembl

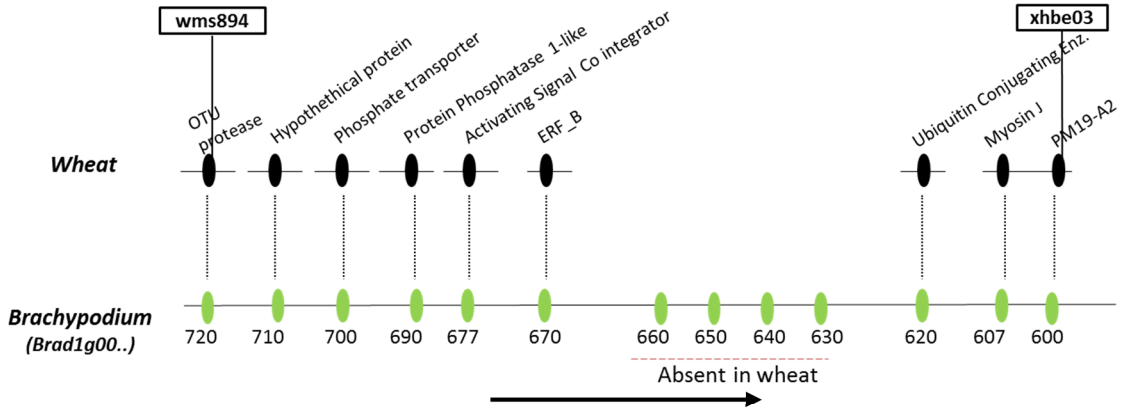


Figure 4.10: Synteny reveals the putative gene content of the *Phs* interval. Sequences of the flanking markers (*wms894* and *xhbe03)* of the *Phs* interval were used to obtain homologous *Brachypodium* genes - *Brad1g00600* and *Brad1g00720* (green oval). The homologous wheat contigs and gene models, obtained by BLAST analysis using the *Brachypodium* gene models, are represented by the black lines and ovals respectively. The chromosome orientation of the map is indicated by the horizontal arrow.

4.2.5 *Phs* is in the 4AL/5AL translocation region

Analysis of homoeologous relationships of the genes identified by synteny to be in the *Phs* interval showed that all the genes have homoeologues on chromosome 5BL and/or 5DL as opposed to chromosome 4BL and 4DL which are the truly homoeologous chromosome arms to 4AL (Table 4.3). This was also observed for the genes containing the flanking markers (*PM19-A2* and *OTU Cysteine Protease*) of *Phs,* suggesting that this pattern of 4AL/5BL/5DL homoeologous relationship spans the entire *Phs* locus. Although these genes do not represent the entire gene content of the *Phs* interval, this pattern of 4AL/5BL/5DL homoeology suggests that the *Phs* interval is within the 4A/5A translocation region represented by segment D in Figure 4.1.

Table 4.3: Homoeologous relationships of the genes in the *Phs* interval

SN	Gene Names	Ensembl Gene/NCBI ID	Homoeologues
1	*Plasma membrane protein, PM19-A2*	Traes_4AL_F99FCB25F	Traes_5BL_33ADEE93A1 Traes_5DL_426297FC7 Traes_5DL_26434784A Traes_5DL_1D17F685A Traes_5BL_47475B33B
2	*Myosin-J heavy chain protein*	Traes_4AL_DD1B27086	Traes_5DL_7D3922FD1 Traes_5BL_3367D4BA2 Traes_5BL_77A44D868
3	*Ubiquitin conjugating enzyme*	M28059	UBC4 Traes_5BL_54CC7D2D6
4	*Ethylene Responsive Factor_B (ERF_B)*	Traes_4AL_F18648C49	Traes_5BL_FBC7AF288
5	*Activating Signal Co-integrator*	Traes_4AL_02AE4777301	Traes_5DL_09AF92311 Traes_5BL_DCE4C52BA
6	*Protein Phosphatase PP1-Like*	Traes_4AL_2EBB63DBA	Traes_5BL_86F86B4A9 Traes_5DL_F7F10739D
7	*Phosphate Transporter*	Traes_4AL_C56125840	Traes_5BL_86CA8C1DD Traes_5DL_391DF2590 Traes_5DL_D473C031B Traes_5DL_E15C27E9B
8	*Hypothetical protein*	Traes_4AL_DD344FC87	Traes_5BL_717B3FD10 Traes_5DL_092BC8C4E
9	*OTU cysteine Protease*	Traes_4AL_F00707FAF	Traes_5DL_A061F2DD6

4.2.6 Physical map construction reveals the gene contents of the *Phs* interval

To access the full gene content of the *Phs* interval, we established collaboration with Prof. Jaroslav Dolezel at the Centre for Plant Structural and Functional Genomic, Institute of Experimental Botany (IEB), Czech Republic. Prof Dolezel's lab has the expertise and facilities to use flow cytometry to separate mitotic chromosome arms of grass species. Furthermore, as part of the IWGSC wheat BAC-to-BAC sequencing project, the lab is constructing the physical map of flow-sorted 4AS and 4AL chromosome arms of the Chinese Spring wheat variety. This effort produced High information Content Fingerprints from 60,140 BACs from the entire 4AL chromosome arm (which also contains the *Phs* locus), and these were assembled using the LTC software to about 924 super-contigs. In addition, Prof Dolezel's lab has generated sequences from three-dimensional (3D) BAC pools of the 4A chromosome. This makes the screening of the BAC library easy and cost effective (Cvikova et al., 2015).

In Prof Jaroslav Lab, we collaborated with Dr Miroslav Valarik and his Ph.D. student Bara Klocova, who were working on the construction of the Chinese Spring 4A physical map. From this collaboration, we obtained the physical map information of the *Phs* interval as well as the BAC clones representing this interval. These clones were afterwards processed at the John Innes Centre.

4.2.6.1 *Anchoring of the Phs interval to the 4AL physical map*

To anchor the *Phs* locus to the 4AL physical map, sequences of the wheat genes containing the flanking *Phs* markers (*PM19-A2* and *OTU cysteine protease*) were sent to our collaborating partners at IEB, Czech. These sequences were searched against the 3D BAC pool sequences of the 4AL chromosome arm, and this enabled the identification of BAC clusters (a group of BACs containing overlapping DNA fragments) containing these marker sequences. The *PM19-A2* gene containing the *xheb03* marker was found in two Chinese Spring BACs (TaaCsp4AL172K12 and TaaCsp4AL037H11), which are both part of BAC cluster 16421. The sequence of the

wms894 marker on the *OTU Cysteine Protease* gene was found in BAC cluster 412 (Figure 4.11).

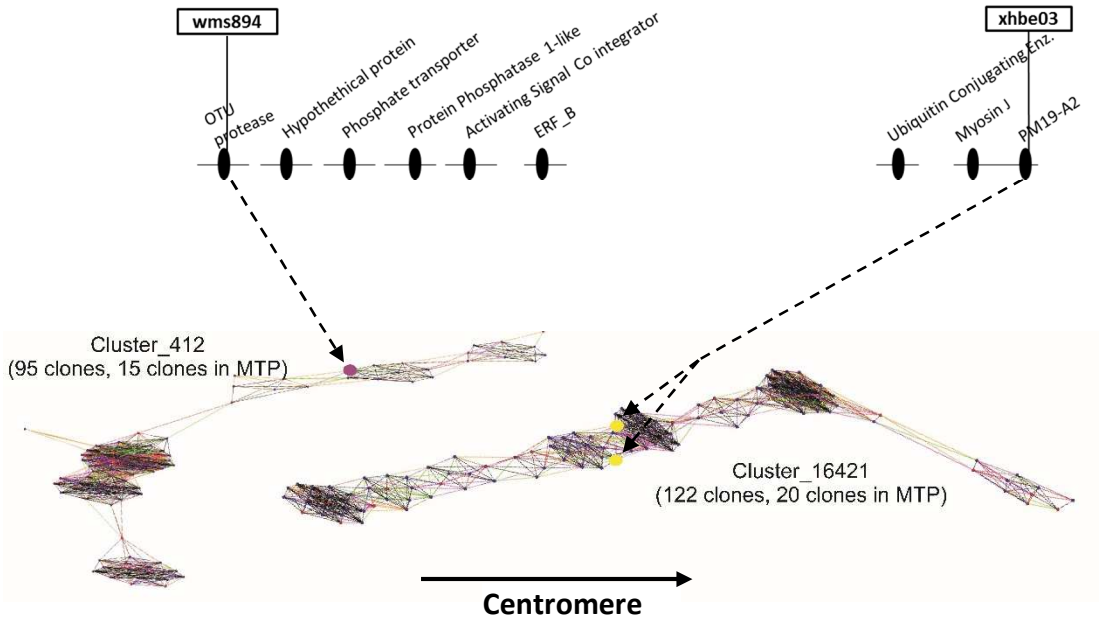


Figure 4.11: The *Phs* interval is anchored on the 4AL wheat physical map. LTC representation of the BAC clusters (Cluster 412 and 16421) which harbour *Phs* flanking genetic markers sequences. Individual BAC clones are represented by dot in the cluster while overlaps between BACs are represented by lines joining the dots. The length of the connecting lines within each cluster is inversely related to the strength of BACs overlap. BAC clones containing *xbhe03* and *wms894* markers are coloured yellow and purple respectively. The chromosome orientation of the map is indicated by the horizontal black arrow.

Cluster_16421 is a very large BAC cluster comprised of 122 BAC clones of which only 20 clones make up the MTP. The cluster was found to also contain the genes encoding for Myosin-J heavy chain protein, and Ubiquitin Conjugating Enzyme homologous to *Brad1g00607* and *Bradi1g00620* respectively. Similarly, cluster 412 is also a large cluster of BACs with 95 BAC clones of which 15 clones represents the MTP of the cluster.

Some of the other wheat genes inferred to be between *PM19-A2* and *OTU Cysteine Protease* by synteny were found in two additional BAC clusters. The genes encoding the *ERF_B, Activating signal Co-integrator* and *Protein Phosphatase* (*PP1-Like*) were all found on Cluster 285, which contains 14 BAC clones of which 4 clones make up the MTP (Figure 4.11). A Hypothetical Protein, homologous to Bradi1g00710, was

encoded by a gene found in Cluster 7335 which is made up of 5 BAC clones with a 4-BAC-clone MTP. We could not find BAC clones containing sequences of the *Phosphate Transporter* gene.

Anchorage of the *Phs* interval to these BAC clusters suggests the possibility of having more genes in the physical map than previously inferred by synteny. This is because the 9 genes identified by synteny are contained in just a small fraction of the total number of BAC clones in these clusters. In addition, we could not establish connections across the four BAC clusters spanning the *Phs* physical map. There is a minimum of three gaps in the physical map. These are located between Cluster 412 and 7335; Cluster 7335 and 285 and Cluster 285 and 16421 (Figure 4.12). One possible reason for the existence of such gap in the physical map could be the high repetitive nature of the intervening region between these clusters. This makes it difficult for the LTC programme to accurately join adjacent clusters together. Such gaps, however, could potentially have new BAC clones/clusters which could contain critical genes underlying the *Phs* effects.

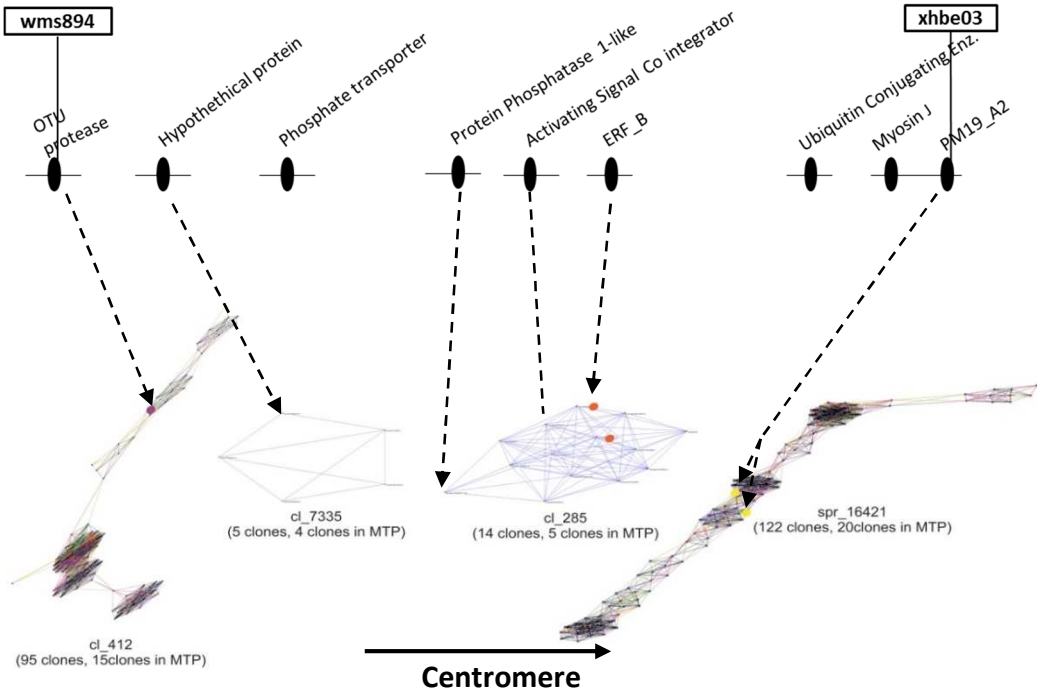


Figure 4.12: The four BAC clusters in the *Phs* physical map interval. LTC representation of the BAC clusters (Cluster 412, 7335, 285 and 16421) in the *Phs* interval. Individual BAC clones are represented by dot in the cluster while overlaps between BACs are represented by lines joining the dots. The length of the connecting lines within each cluster is inversely related to the strength of BAC overlap. BAC clones containing syntenic gene are connected to their corresponding genes by the arrows. The chromosome orientation of the map is indicated by the horizontal black arrow.

4.2.6.2 *Sequencing of BAC Clusters 16421, 285 and 7335 reveal the presence of additional genes in the Phs interval*

About 90% of the anchored syntenic genes in the *Phs* interval are located in clusters 16421, 285 and 7335, with cluster_412 containing only the *OTU Cysteine Protease*. For this reason, we prioritise the sequencing of the relevant MTP clones of these three clusters. In addition, due to the large size of Cluster 16421, only 11 of the 20 MTP clones starting from TaaCsp4AL172K12 (which contains *PM19-A2*) to the last BAC on the proximal end of Cluster 16421 (TaaCsp4AL031F05) were sequenced. This reduced the number of clones analysed from 43 to 19.

All the 19 BAC clones were obtained from our partners at IEB, Czech, and cultured for BAC extraction at the John Innes Centre. The extracted BACs were then sequenced on the Miseq Illumina sequencing platform at The Genome Analysis Centre, Norwich, UK. This produced 250 bp paired-end reads from average library insert sizes of 360, 460 and 440 bp for Cluster 16421, 285 and 7335, respectively.

4.2.6.2.1 Sequencing of the BAC Cluster 16421 reveals the presence of 11 genes

Eleven BAC clones were sequenced from cluster 16421, and these produced a total of 23,160,366 paired-end reads with the average number of reads per BAC being 2,105,488. These reads were filtered for contaminants (mainly vector and *Escherichia coli* sequences) and were assembled using CLC Genomics (described in Chapter Two, section 2.5.4). The assembly statistic for each of the BAC is represented in Table 4.4.

The assembly is made up of varying number of contigs per BAC ranging from 51 contigs in BAC TaaCsp4AL031F05 to 338 contigs in TaaCsp4AL023E24. Most of the BACs have N_{50} length greater than 2500 bp. The N_{50} is a measure of central tendency used in NGS assembly to describe the distribution of contigs length. It is defined as the minimum contig length for which, sum of all contigs above this length represents at least 50 % of the total assembly length (Ekblom and Wolf,

2014). The minimum and maximum contig lengths for the sequenced BACs ranged from 239 to 129,441 bp, respectively.

Table 4.4: Summary statistics for the assembly of Cluster 16421 BACs

BAC ID	N_{50} (bp)	Number of Contigs	Contig Length (bp)			Total base count (bp
			Min	Max	Average	
TaaCsp4AL031F05	10573	51	285	30761	2650	135164
TaaCsp4AL034K23	9588	107	281	36933	1355	145028
TaaCsp4AL121J20	2984	215	253	51766	1339	287909
TaaCsp4AL023E24	1496	338	265	59755	996	336617
TaaCsp4AL151A19	129441	156	268	129441	1495	233279
TaaCsp4AL052N17	4442	199	239	39770	996	198254
TaaCsp4AL168N08	22068	173	252	104345	1363	235732
TaaCsp4AL205J15	1356	280	259	67409	989	276983
TaaCsp4AL124E07	4721	252	252	48030	994	250496
TaaCsp4AL037H11	11856	152	253	67740	1304	198223
TaaCsp4AL172K12	10965	171	296	35476	1261	215574

Following manual annotation of the selected contigs for transposable element sequences by BLAST analysis against the Triticeae Repeat (TREP) database, the chromosome locations of the contigs were assessed by BLASTN analysis against the IWGSC CSS wheat chromosome arms sequence. Since the two different sequence assemblies (our BACs and the IWGSC CSS assemblies) are from the same genotype, we expected a very high level of sequence identity between the two assemblies. As expected, the BLASTN returned highly identical hits to 4AL IWGSC CSS contig sequences (>99.9% sequence similarity). In addition, adjacent BACs showed highly identical overlaps over long stretches of sequences. This overlap confirms the connections between the BAC as inferred by the LTC programme and it implies that the assembly across the whole BAC cluster represents a contiguous stretch of DNA sequence.

Annotation of genes on the BACs was done by a similarity search of the BACs sequences against the gene models and protein dataset of EnsemblPlant and NCBI. To complement these annotations, ab-initio gene prediction with FGENESH (Solovyev et al., 2006) was also done. Based on the availability of homology support and similarity with already annotated or characterised genes, genes models identified in the contigs were classified as either high confidence genes or low confidence genes (See Chapter Two, section 2.5.6 for details).

This annotation revealed the presence of 11 high confidence gene models in eight of the BACs of cluster 16421 (Figure 4.13). These include three genes that were previously identified by synteny to be in the *Phs* region - *PM19-A2, Myosin-J heavy chain protein* and *Ubiquitin Conjugating Enzyme*. In addition to these syntenic genes, eight new gene models were identified across the BAC cluster. These include: a gene encoding for *PM19-A1* (paralog of *PM19-A2*); *Amino-Cyclopropane Carboxylate Oxidase-1 like* (*ACC Oxidase-1*); two gene models encoding for Leucine-Rich Repeat Kinases (*LRR Kinase-1* and *LRR Kinase-2*); a gene encoding for a Hypothetical Protein; an *Agmatine coumaroyl transferase* gene, a gene encoding for Malonyl Coenzyme A:anthocyanin 3-O-glucoside-6''-O-malonyltransferase; and a *YUCCA3-like* gene.

In agreement with the gene island model for the distribution of genes in the wheat genome (Rustenholz et al., 2011), gene distribution across the BACs was not uniform. While some BACs like TaaCsp4AL121J20 and TaaCsp4AL037H11 together contained 9 of the 11 identified genes (four and five genes each respective), other BACs like TaaCsp4AL023E24, TaaCsp4AL151A19 and TaaCsp4AL151A19 did not have any high confidence gene model in them. In addition, it was possible to define the orientation of the assembly based on the overlapping regions between BACs. The order of many genes within most of the BAC was also verified based on the co-localisation of genes on the same assembly contig (implying their physical linkage).

Three other putative genes identified in this cluster were classified as low confidence genes. This includes genes with partial and low similarity matches to genes on NCBI encoding for a B3 containing protein (EMT14114; 36% protein

identity) and a Hypothetical Protein (EMT22294, 22% protein coverage with 93% identity). Another gene model obtained by ab-initio prediction showed high sequences similarity to a gene model for TRAES_3BF012800080CFD_c1 (unnamed *Triticum aestivum* protein) on chromosome 3B. However, this gene model is annotated as a low confidence gene model on NCBI. Due to the lack of substantial homology support for these genes and lack of similarity to any functional protein domain, these putative genes could be non-functional or pseudogenes and as such were not analysed further.

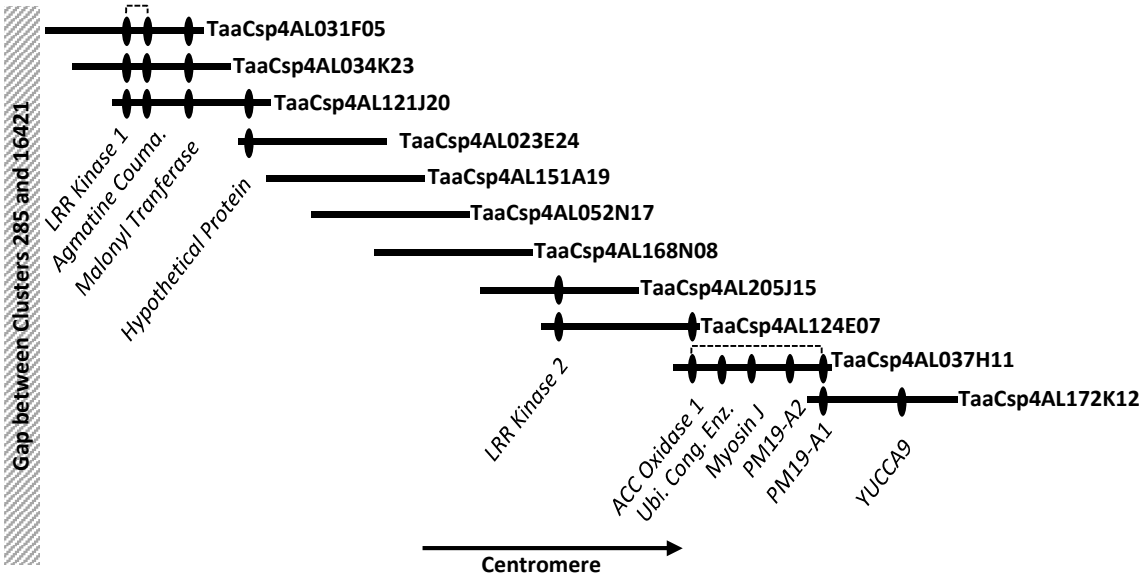


Figure 4.13: Schematic representation of Cluster 16421 BACs harbouring 11 high-confidence genes: Sequencing of the 11 BACs (solid lines) from Cluster 16421 enabled the identification of 11 gene models (ovals) across the entire cluster. BACs are arranged according to the order of overlap (not in scale) and genes found on the same assembly contigs are connected by the dotted lines. The orientation of the chromosome is indicated by the horizontal arrow, while the position of the gap between Clusters 16421 and 285 is indicated by the striped boxes.

4.2.6.2.2 Sequencing of BAC Cluster 285 reveals the presence of five additional genes

The four BACs that make up the MTP of Cluster 285 were sequenced, and these produced a total of 10,900,880 paired end-reads with an average number of reads per BAC being 2,752,220. The reads were similarly processed and assembled like Cluster 16421 BACs. The assembly statistics for each of the BACs is presented in Table 4.5. Two of the assemblies have N_{50} greater than 6 kbp while the other two reads had N_{50} less than 2 kb. Nonetheless, all the assemblies produced contigs long enough for gene annotation. The maximum contig length achieved in these assemblies was 97 kb. As with cluster 16421 BACs, all the contigs showed very high sequence similarity (> 99%) to the IWGSC 4AL chromosome arm contig sequences.

Following the annotation of the selected contigs from these BACs assemblies, five high confidence genes were identified across the whole BAC cluster (Figure 4.14). Three out of these five genes had already been identified by synteny. This includes genes encoding for *ERF_B*, *Activating Signal Co-integrator* and *Protein Phosphatase 1-like* (*PP1-Like*). The new genes identified in this cluster encode for another *Ethylene Responsive Factor* (*ERF_C*) and a *Mitogen-Activated Protein Kinase 6 – Like* (*MAPK-6*). The only low confidence gene identified in this cluster showed similarity to a *Lactation Elevated Protein* (Genbank No.: EMT21696).

Table 4.5: Summary statistics for the assembly of luster 285 BACs

BAC ID	N_{50} (bp)	Number of Contigs	Contig Length (bp)			Total base count (bp)
			Min	Max	Average	
TaaCsp4AL032F12	6093	213	307	60808	1047	222937
TaaCsp4AL012P14	1542	118	243	57216	1134	133821
TaaCsp4AL002F16	1166	166	244	16044	986	163619
TaaCsp4AL170N01	6938	130	266	97325	1500	195023

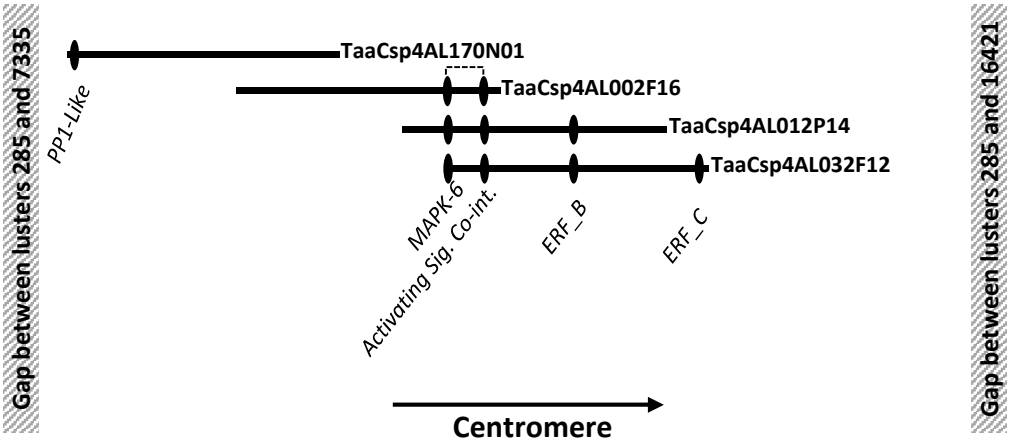


Figure 4.14: BAC sequences of Cluster 285 harbour five high-confidence genes: Sequencing of the four BACs (solid black lines) from luster 285 enabled the identification of five high confidence gene models (ovals) across the entire cluster. BACs are arranged according to the order of overlap (not in scale) and genes found on the same assembly contigs are connected by the dotted black line. The orientation of the chromosome is indicated by the horizontal arrow, while the positions of the gaps between clusters are indicated by the striped boxes.

4.2.6.2.3 Sequencing of BAC Cluster 7335 shows evidence of chromosome arm contamination

The MTP BACs of Cluster 7335 were sequenced and assembled as described for the other clusters. Most of the contigs in this cluster assembly contain transposable elements and only one gene model encoding for a Cytochrome P450 was found in the entire BAC cluster. However, repeat-masked contigs from this cluster did not show the expected high sequences identity to the IWGSC CSS 4AL chromosome arm sequences. Instead, almost all the non-repeat sequences map to the 6DS chromosome arm sequences. This is not too surprising as the 6DS chromosome has been found to be the second largest contaminant of the 4AL flow-sorted chromosome sequence (Miroslav Valarik, personal communications). . Due to the high level of contamination in this cluster, the contig sequences and the gene model found in it were not carried forward for analysis.

4.2.6.2.4 Extension of Cluster 285 BAC sequences exposes additional gene model

As previously stated in section 4.2.6.1, three gaps exist between the BAC clusters represented in this physical map (Figure 4.12) and these could prove to be very important for the fine-mapping of the *Phs* locus. One potential way of closing this gap is to use the end sequences (BES) of the BACs bordering each gap to search for highly similar wheat contigs/scaffolds from independent genome assemblies. Such contigs/scaffolds could provide additional sequences that could bridge the gaps in the physical map. However, precaution must be taken to ensure that BAC connections to the new scaffold(s) are real and not due to the presence of repetitive sequence elements.

Until now, the BAC assemblies have only been compared against the IWGSC Chinese Spring assembly. While this has helped in validating the chromosome arm location of the BAC sequences, the survey sequences did not significantly extend the BAC end sequences into the gaps. This was mainly due to the fragmented nature of these contigs (average N_{50} of 2,300 bp, IWGSC, 2014). However, an independent whole genome shotgun assembly of Synthetic hexaploid wheat - W7984 - produced by Chapman et al. (2015) has a relatively longer N_{50} of 8,300 bp.

Although this shotgun assembly is made from a different wheat cultivar, its longer contigs offer the opportunity to extend the BAC ends into the gaps between the clusters. Using this independent assembly, the end sequences of the BACs bordering the gaps at either ends of Clusters 285 (TaaCsp4AL032F12 and TaaCsp4AL170N01) were used to search for identical contigs from the W7984 assembly that could extend the end sequences of these BACs into the intervening gap. A relatively lower sequence identity of the BAC sequences to the W7984 scaffold sequence was expected due to the presence of varietal SNPs and also due to the presence of ambiguous sequences (N) in the W7984 WGS assembly.

A BLASTN search with a 4,207 bp BAC end sequence of TaaCsp4AL170N01 identified a 23,600 bp W7984 scaffold (Scaffold3222460) with high similarity to the query

(98% identity). The new scaffold extends the BAC end sequences by 3,785 bp out of which 1,032 bp are ambiguous bases (N). No gene model was found in this extended TaaCsp4AL170N01 end sequence (Figure 4.15).

Similarly, a BLASTN search with a 5,516 bp end sequence of BAC TaaCsp4AL032F12 identified a 20,800 kb W7984 scaffold (Scaffold4326385) with high sequence identity (97%). This significantly extended the BAC end by 16.2 kb of additional sequence into the intervening gap. Annotation of this additional sequence identified a new gene model for a disease resistance protein *RPM1* (Figure 4.15). In addition, efforts to use this new sequence to search the Chinese Spring BACs 3D pool sequences for new clones that could be located within the gaps did not give any positive hits.

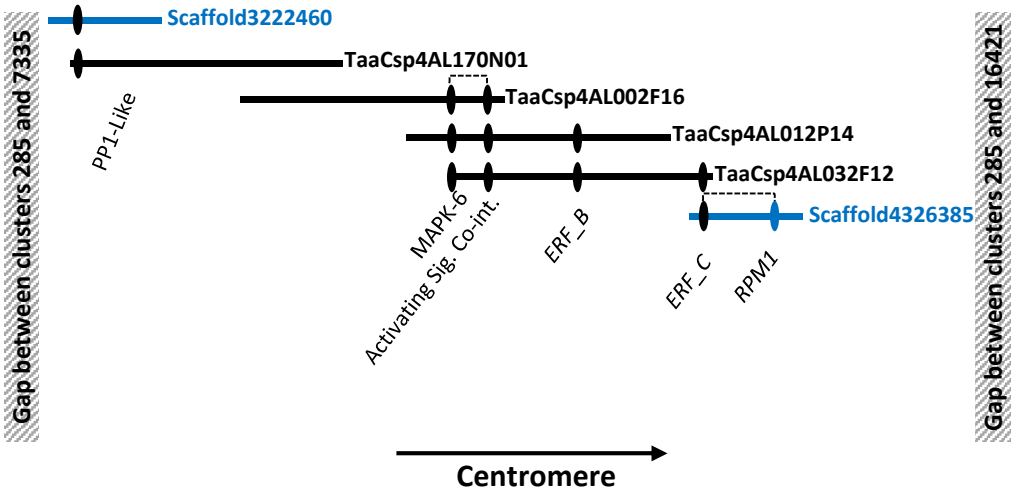


Figure 4.15: Extension of the BAC sequences of Cluster 285: The five gene models identified in Cluster 285 are represented by the black ovals. Extension of the BACs end with W7984 scaffolds sequences (solid blue line) revealed an additional gene model (blue oval). BACs are arranged according to the order of overlap (not in scale) and genes found on the same assembly contigs are connected by the dotted black lines. The orientation of the chromosome is indicated by the horizontal arrow, while the positions of the gaps around this cluster are indicated by the striped boxes.

4.2.6.3 *Summary description of the genes found on the physical map*

The physical map construction of the *Phs* interval uncovered a total of 17 high confidence genes within the *Phs* locus. These have varied gene ontologies and are involved in a wide range of biological processes. A short description of each gene's function is presented in Table 4.6 with associated references.

Several of the genes in the physical map have hormone related functions. Genes like *ACC-Oxidase-1, ERF_B* and *ERF_C,* are specifically involved in ethylene biosynthesis and signaling. No known function could be ascribed to the Hypothetical Protein, it, however, shows high homology to *ACC Oxidase-1,* and so it could also be involved in ethylene biosynthesis.

Although the *PM19* genes are not involved in hormone biosynthesis, they are reported to be inducible by ABA. They are also the only genes in the physical map already implicated in the control of seed dormancy (Koike et al., 1997b, Ranford et al., 2002). This is because they were found to be differentially expressed in dormant and non-dormant seeds of barley after imbibition. This role of *PM19* in barley is very relevant to the characteristic of the *Phs* effect.

There are also genes in the physical map which are potentially involved in plant defense responses against pest and pathogen. Genes like *LRR Kinase-1, LRR Kinase-2* and *RPM1* are members of the leucine-rich repeat receptor proteins, which primarily function in various aspect of plant defense against pest and pathogen. In addition, the *Agmatine coumaroyl transferase* gene is involved in the chemical defense response in plants through the production of hydroxycinnamic acid (Lowe et al., 2015).

Other genes found on the map are involved in the transcriptional or post-translational regulation of biological processes. These include *Activating Signal Co-integrator*, which is a transcriptional co-activator and the *MAPK-6* and *PP1-Like* genes which are ubiquitous regulators of the reversible phosphorylation of different molecular substrates. The *Malonyl transferase* gene is involved in the secondary metabolism of anthocyanin pigment stabilisation, especially in flower pigments.

Table 4.6: Summary description of identified genes in the *Phs* physical map

BAC cluster	Genes	General Description
Cluster 16421	*ACC Oxidase-1-like*	Involved in the terminal step of ethylene biosynthesis by converting aminocyclopropane Carboxylate to ethylene (Chen et al., 2014a, Yoon and Kieber, 2013)
	Agmatine coumaroyl transferase	Involved in the biosynthesis of hydroxycinnamic acid which offers chemical defence against pest. (Muroi et al., 2012)
	LRR Kinase- 1 and 2	Member of the largest subfamily of transmembrane receptor-like kinases in plants that catalyses a wide range of defence responses, and are involved in different developmental processes (Torii, 2004)
	Hypothetical Protein	Gene of unknown function but similar to *ACC-Oxidase 1-like*
	Malonyltransferase	Provides anthocyanins pigment stabilisation in plant by catalysing the substitution reaction between Anthocyanin and malonyl-CoA (Suzuki et al., 2002)
	Myosin J Protein	Catalyses the hydrolysis of ATP used for the active movement of organelles or macromolecular motors along the actin cytoskeleton. (Knight and Kendrick-Jones, 1993, Shimmen, 2007)
	PM19-A1 and A2	ABA-inducible plasma membrane protein expressed during freezing tolerance of wheat and during imbibition of dormant seed in barley (Ranford et al., 2002, Koike et al., 1997b)
	Ubiquitin Conjugating Enzyme	Involved in protein turnover and regulation by catalysing the conjugation of Ubiquitin to specific residue in target proteins (van Wijk and Timmers, 2010, Sullivan et al., 1994)
	YUCCA3	Involved in auxin biosynthesis and plant development (Cheng et al., 2006)
	Activating Signal Co-integrator	A transcriptional co-activator involved in the regulation of gene expression by interaction with multiple transcription factors. Also regulates RNA processing. (Iyer et al., 2006)
Cluster 285	*ERF_B and C*	Members of the second largest family of transcriptional factors in wheat. They are involved in the signalling of the hormone ethylene in different developmental and defence-related processes (Chen et al., 2015)
	MAPK-6	Member of a large family of kinases that function in the transduction of signal generated in response to different environmental and developmental clues (CristinaRodriguez et al., 2010). *MAPK-6* is specifically involved in jasmonate signalling (Takahashi et al., 2007).
	PP1-Like	A member of the serine-threonine protein phosphatase family which catalyses the reversible dephosphorylation of molecular substrate in different developmental process. (Lin et al., 1999)
W7984 Scaffold	*Disease Resistance protein RPM1*	A member of the nucleotide-binding Leucine-Rich Repeat Protein family similar to *RPM1* in Arabidopsis, which is an *R*-gene that recognises an effector from *Pseudomonas syring*e. (Boyes et al., 1998, Feuillet et al., 2003)

4.2.7 Physical map of the syntenic *Phs* interval in barley

In addition to the Chinese Spring physical map sequence of *Phs*, we took advantage of the physical map resources of barley (The International Barley Genome Sequencing Consortium, 2012). Barley shows a higher degree of co-linearity and gene conservation to wheat compared to *Brachypodium* (Bolot et al., 2009, Linde-Laursen et al., 1997). The physical map of barley might, therefore, be useful in identifying the genes that could be located in the gaps between the clusters. Furthermore, this will also enable a comparative analysis of the *Phs* interval between wheat, barley and *Brachypodium*.

The barley physical map developed from cultivar Morex was searched to find sequences homologous to the flanking markers of the wheat *Phs* interval. This search identified a barley consensus contig, FPcontig_2379, made from 34 overlapping BAC clones sequences. However, the homologous *Phs* interval only span 7 out of the 34 overlapping BAC clones. Dr. Nils Stein (Leibniz Institute of Plant Genetics and Crop Plant Research, IPK, Germany) and Professor Chengdao Li (Murdoch University, Australia) kindly provided us with the sequences of FPcontig_2379. Following the annotation of FPcontig_2379 sequences by BLAST analysis against the high confidence barley gene dataset (The International Barley Genome Sequencing Consortium, 2012), we identified 24 genes which are listed in Table 4.7.

Only eight out of the 24 genes identified in the barley physical map are homologous to genes identified in the wheat *Phs* physical map. These include orthologues of wheat *PM19-A1 and A2, Myosin J Protein, Ubiquitin Conjugating Enzyme, ACC Oxidase-1-like, Activating Signal Co-integrator, MAPK-6 and PP1-Like*. We also found orthologues for the Hypothetical Protein (*Bradi1g00710*) and *OTU Cysteine Protease* (*Bradi1g00720*) which contained the *wms894* flanking marker. Other genes found in the barley physical sequence include: *F-box containing protein,* two *NBS-LRR genes, Resistance Protein RGA, Zinc Knuckle, Farl- Related 5-like, UPF0480 PROTEIN AT4G32130-like, GaMyb-binding Protein, Proline-rich protein family-like* and genes encoding for two other Hypothetical Proteins.

Table 4.7: List of genes in the physical map of homologous *Phs* interval in barley

SN	Genes	Gene models and % nucleotide identity	BAC clones
1	*F-Box Containing Protein*	MLOC_57596.1 (98.7%)	HVVMRXALLmA0022M08
2	*NBS-LRR*	MLOC_28129.1 (100%)	HVVMRXALLmA0022M08
3	*NBS-LRR*	MLOC_76783.1 (100%)	HVVMRXALLmA0022M08
4	*HvPM19-1*	MLOC_66583 (100%)	HVVMRXALLmA0022M08
5	*HvPM19-2*	AK249250.1* (97%)	HVVMRXALLmA0022M08
6	*HvPM19-3*	AK249250.1 (96%)	HVVMRXALLmA0022M08
7	*HvPM19-4*	AK249250.1 (97%)	HVVMRXALLmA0022M08
8	*Myosin-J*	MLOC_72134.2 (100%)	HVVMRXALLmA0367P07
9	*Ubiquitin Conjugating enzyme*	MLOC_58932.2 (100%)	HVVMRXALLmA0367P07
10	*ACC Oxidase 1-like*	MLOC_58934.1 (100%)	HVVMRXALLmA0367P07
11	*Resistance Protein RGA*	MLOC_62062.2 (100%)	HVVMRXALLmA0367P07 HVVMRXALLeA0124O08
12	*Zinc Knuckle*	MLOC_20776.1 (100%)	HVVMRXALLmA0367P07 HVVMRXALLeA0124O08
13	*Hypothetical protein*	AF474373.1* (99%)	HVVMRXALLeA0124O08
14	*Farl- Related 5-like*	MLOC_1588.1 (96%)	HVVMRXALLeA0124O08
15	*UPF0480 PROTEIN AT4G32130-LIKE*	MLOC_47012 (97%)	HVVMRXALLrA0371B17
16	*ACC Oxidase 1-like*	MLOC_73240.1 (100%)	HVVMRXALLrA0371B17
17	*Activating Signal Contintegrator*	MLOC_5194 (90%)	HVVMRX83KhA0191G07 HVVMRXALLmA0179F14
18	*Mitogen-activated protein kinase kinase 6-like (MAPK-6)*	MLOC_51567 (100%)	HVVMRX83KhA0191G07 HVVMRXALLmA0179F14
19	*Serine Threonine Protein Phosphatase 1-like (PP1-Like)*	KD514543.1*	HVVMRX83KhA0191G07 HVVMRXALLmA0179F14
20	*Hypothetical protein*	AK371412.1*(97%) 710	HVVMRX83KhA0191G07 HVVMRXALLmA0179F14
21	*GaMYB-Binding Protein*	MLOC_8990.2 (98%)	HVVMRXALLmA0179F14 HVVMRXALLmA0051F11
22	*Hypothetical protein*	MLOC_39208 (96%)	HVVMRXALLmA0179F14 HVVMRXALLmA0051F11
23	*OTU cysteine protease*	MLOC_23495 (97%)	HVVMRXALLmA0051F11
24	*Proline-rich protein family-like*	MLOC_51079 (99%)	HVVMRXALLmA0051F11

* NCBI gene model used as no suitable gene model was found on Ensembl

4.2.8 Comparison of syntenic *Phs* segments of wheat, barley and *Brachypodium*.

Through synteny and physical mapping, we have investigated the gene contents of the homologous *Phs* segments in wheat, barley and *Brachypodium*. A comparison of these gene contents in these homologous *Phs* segment identified a region of high conservation and a region of low conservation (Figure 4.16). In the region of high conservation, there was conservation of gene content and collinearity between the three different species. In this region, orthologues could be found for the *Brachypodium* genes *Bradi1g00600* - *Bradi1g00620* and *Bradi1g00670* - *Bradi1g00720* in the other two grass species.

Despite this high level of gene conservation, differences in gene content were still observed between the species in this region. This is mainly due to the presence of additional genes in wheat and barley which are not found in *Brachypodium*. For instance, the *ACC Oxidase-1 like* and *MAPK-6* genes, which are present in wheat and barley, are missing in this region in *Brachypodium*. Similarly, *ERF_B* and *ERF_C*, which were identified in *Brachypodium* and wheat, were missing in barley. In addition, there are differences in the copy number of the *PM19* genes in the three species. While *Brachypodium* has just one copy of *PM19* (*Bradi1g00600*), wheat and barley have two and four copies of *PM19*, respectively. This suggests a duplication of the *PM19* genes in the Triticeae genomes.

In contrast, in the region of low conservation, there was no conservation of gene content between the three species. None of the genes in this region were common between any of the three species. It is however important to note that there is a gap in the physical map of wheat in this region of low conservation, and this could contain genes that could be homologous to one or both of the other species.

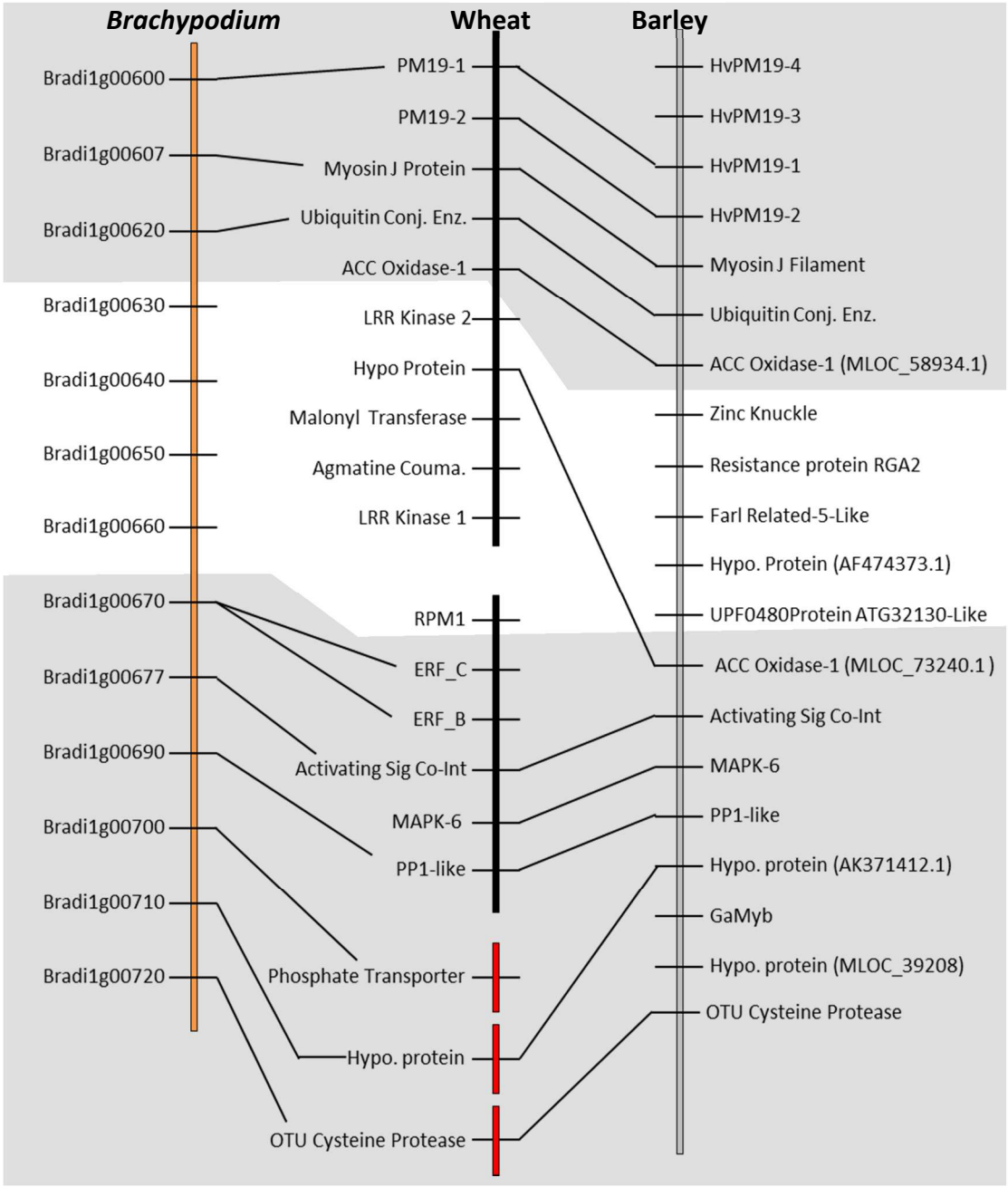


Figure 4.16: Comparison of syntenic *Phs* physical map and contigs in *Brachypodium*, wheat and barley. Genes in the homologous *Phs* interval in *Brachypodium* (amber line), wheat (black lines for physical map and red lines for IWGSC contigs) and barley (grey line) are compared against each other. The genes joined by lines across the genomes are orthologues. The region of high conservation is indicated with grey background while the region of low conservation has plain background.

4.2.9 Sequencing of flow-sorted Claire and Option 4AL chromosome arm

The bread wheat genome is very large and complex. Reduced representation methods such as flow-sorting of chromosome arms greatly reduce the complexity of the hexaploid genome of wheat for next-generation sequencing (NGS) and facilitate access to the gene space of wheat. However for successful separation by flow-sorting, there should be sufficient size differences between chromosomes. This limited chromosome flow-sorting to the Chinese Spring wheat variety which has cytogenetic (aneuploid) stocks (Sears and Sears, 1978) with sufficient size differences between chromosome arms.

To circumvent this chromosome size difference limitation, a new flow-sorting techniques called Fluorescent In situ Hybridization In Suspension (FISHIS; Giorgi et al., 2013) have been developed. In FISHIS, fluorescent labelled repetitive DNA are hybridised to individual chromosomes. With this method, it is now possible to separate the chromosomes of any wheat variety. Furthermore, advance in NGS technology has greatly reduced the cost per bp of nucleotide sequence (Poland and Rife, 2012, Ross, 2011, Swan, 2010).

Taking advantage of this novel method, we carried out the isolation and sequencing of flow-sorted 4AL chromosome arm of Claire and Option through our partners at the Institute of Experimental Botany, Czech, using the Illumina Miseq platform. The use of 4AL sequence directly from Claire and Option allows the direct comparison of the genome of the parental varieties, especially at the *Phs* locus. This also facilitated the discovery of polymorphism between these varieties and accelerated marker development for the fine-mapping of *Phs*. To the best of our knowledge, this represents a novel approach for identifying polymorphisms in candidate gene underlying the effect of QTL of interest from divergent varieties/lines.

The assembly of Claire and Options Illumina reads with the AbySS assembler (Simpson et al., 2009) gave 7,100,735 and 6,290,032 contigs respectively. Although the N_{50} values of the assemblies were low, 132 for Claire and 141 for Option, maximum contig lengths of 26,311 and 14,644 were still obtained for the Claire and Option assemblies,

respectively. In addition, an alignment of Option reads to the Claire assemblies identified more than 72,000 SNPs.

4.2.10 Analysis of Genes in the Physical Map of the *Phs* Interval

With the exception of *PM19-A1* and *PM19-A2* which have evidence for possible involvement in seed dormancy control, it is difficult to speculate if any of the other genes identified in the physical map play a role in the seed dormancy phenotype of *Phs*. A transcriptional profiling of the identified genes, was undertaken to better understand the function of these genes.

4.2.10.1 *Spatial expression pattern of genes found in Phs physical map*

The genetic and physiological characterisation of *Phs* reported in section 4.2.1 and 4.2.2 highlights some defining characteristics about the causal gene(s). One of these characteristics is an expected expression in the grain, specifically in the zygotic grain tissues, where the *Phs* effect was found to be manifested. Based on this hypothesis, we investigated the spatial expression pattern of genes identified across the *Phs* physical map.

To do this, we obtained pre-publication access to the wheat expression database developed within the Uauy lab (Borrill et al unpublished) which contains 16 different RNAseq dataset from diverse wheat germplasm across different tissues, developmental stages, and under different experimental conditions (treatments). The RNAseq reads from the different dataset were mapped to a common set of gene models obtained from the EnsemblPlant database.

Furthermore, the Transcript per Million (TPM) expression metric was used for quantifying the expression level of genes across the dataset and this allows for comparisons of expression level across different tissues and treatments (Li et al., 2010). This dataset therefore allows for a more extensive and robust transcriptional profiling of genes across a larger number of tissues than would have been possible by quantitative Real-Time PCR (qPCR). However, the expression pattern of only 11 out of

the 17 genes could be examined due to a lack of suitable (with > 95 % similarity) gene model representation in the EnsemblPlant database.

The expression analysis of untreated wheat samples (47 samples from different experiments) shows that the genes are differentially expressed in the grain, spike, leaf, shoot and root (Fig 4.17). In the grain, *PM19-A2* showed the highest level of expression compared to the other genes. *PM19-A1* was also expressed in grains but with a 6-fold ($P < 0.001$) reduction in expression level compared to *PM19-A2*. The other genes showed very low or negligible expression in grain. Conversely, transcripts for the *PM19* genes were barely detectable in the other tissues while the other genes showed a higher level of expression in other tissues than in the grain.

Myosin J protein, RPM1, Activating Signal Co integrator and *PP1-Like* all had their highest level of expression in the spike. The *YUCCA* gene also showed detectable transcript level in the spike, but was more highly expressed in the root. Of all the tissues examined, the leaf and shoot showed the lowest level of expression of these genes. As expected, both tissues also showed a similar pattern of gene expression. In these tissues, the transcript accumulation of *Activating Signal Co-Integrator*, *MAP*K and *PP1-Like* and *ACC Oxidase 1 gene* were most abundant. In the root, all the genes except the *PM19* genes were expressed. The root was also the only tissue that showed any transcript level for the *LRR Kinase 2* and *ERF_B* gene, suggesting that the expression of these genes might be specific to the root.

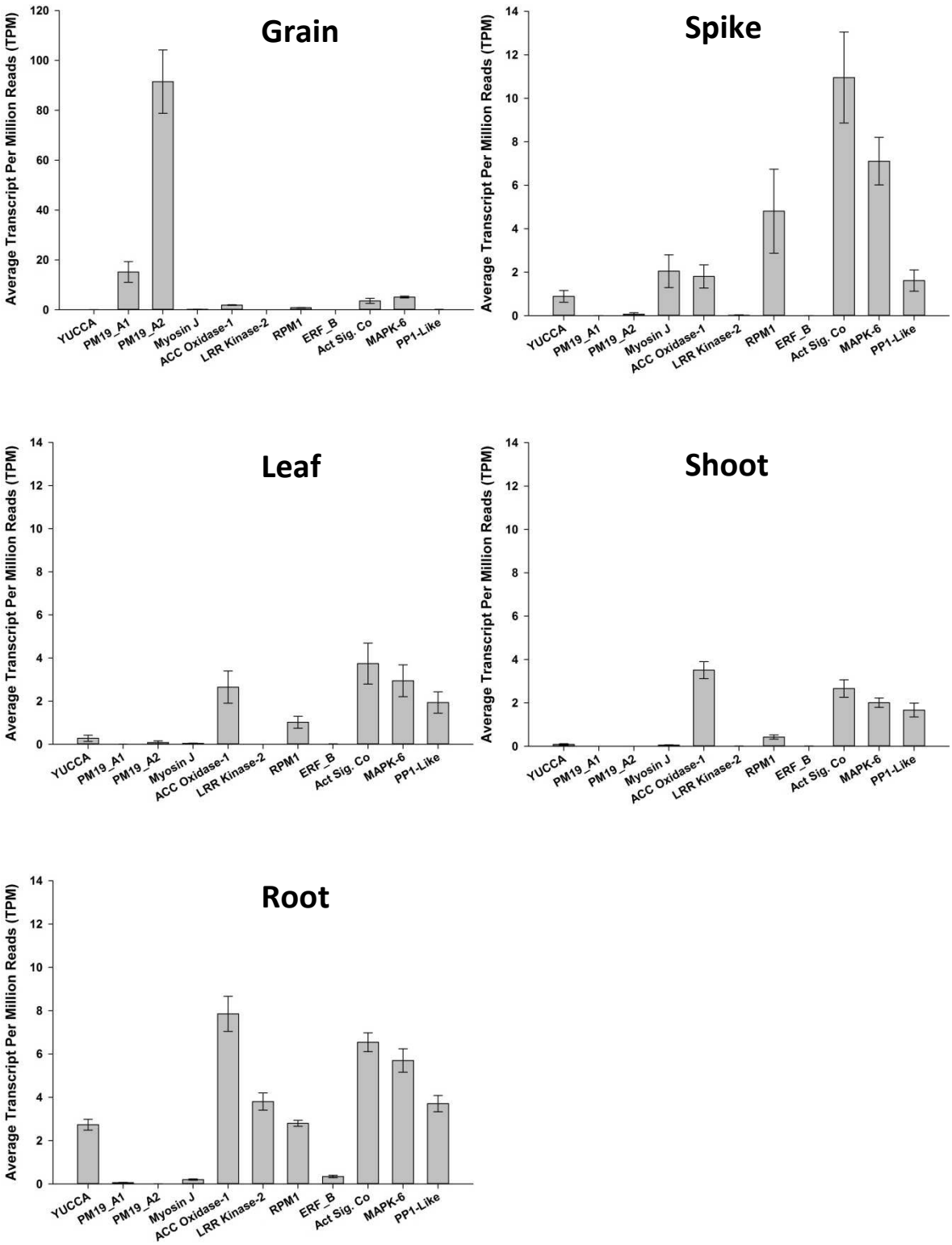


Figure 4.17: Spatial expression pattern of genes in *Phs* physical map: The figure shows normalised transcript levels (TPM) of genes identified in the physical map of *Phs* across different tissues of untreated wheat plants. The data was obtained from Borrill *et al* (unpublished) expression database. The error bars are SEM of different experimental samples from the same tissue. Note that the expression scale for the grain tissue is higher than for other tissues.

To better understand the expression pattern of these genes in the grain, where the *Phs* effect was found to be manifested, their transcript level across different time points during grain development was examined. The time points examined captured both the very early and late stages of grain development and maturation starting from two days post anthesis (DPA) to 35 DPA. This temporal transcript analysis showed that the *PM19* genes are induced late during grain maturation (Figure 4.18). The transcript level of both copies of *PM19* started accumulating from 25 DPA and reached peak levels at 30 DPA for *PM19-A2* and 35 DPA for *PM19-A1*. As was previously observed, the transcript level of *PM19-A1* was significantly lower compared to *PM19-A2* at these time points.

The *Activating Signal Co-integrator* gene was expressed early during grain development (2 and 11 DPA) and at 30 DPA. Interestingly, its expression level at 30 DPA was higher than *PM19-A1* but lower than *PM19-A2*. The *MAPK-6* gene show constitutive but very low level of expression across all the time points examined.

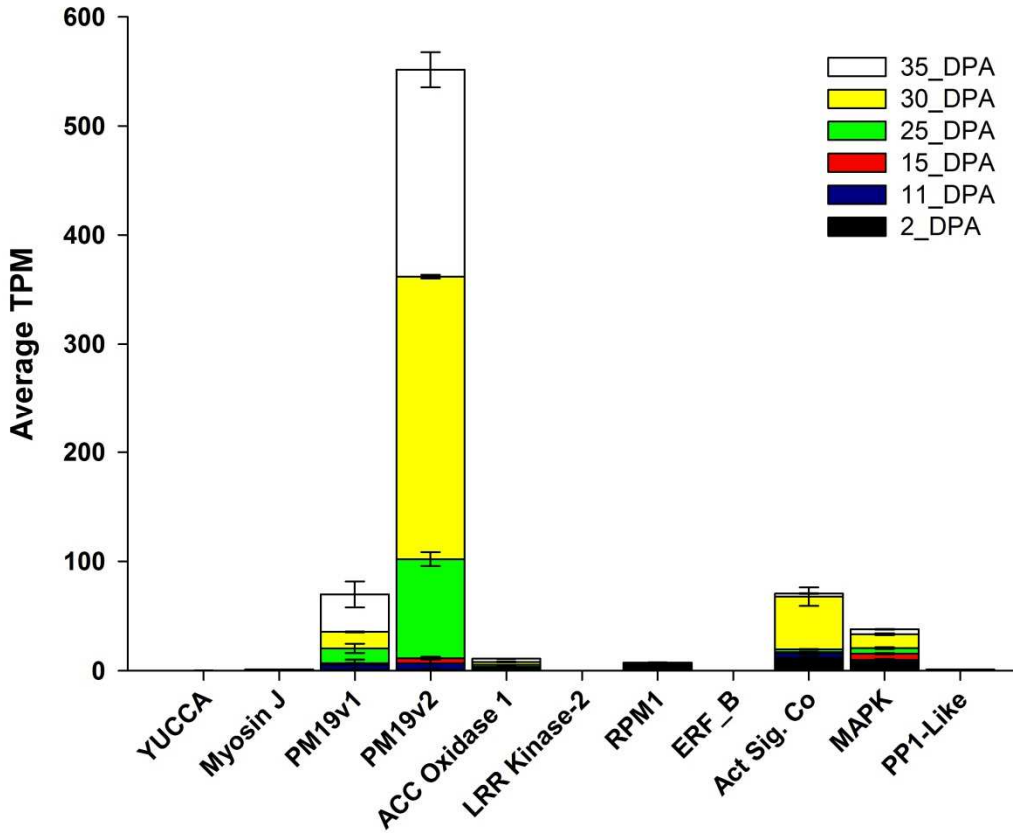


Figure 4.18: Temporal grain expression pattern of genes in *Phs* physical map. The figure shows normalised transcript levels (TPM) of genes identified in the physical map of *Phs* across different grain developmental time points in untreated wheat plants. The data is obtained from Borrill *et al.* (unpublished) expression database. The error bars are SEM of different grain samples from the same time point.

In summary, the spatial and temporal expression analyses show that the *PM19* genes are the most highly expressed genes in the grain and that their expression is specific to grains in untreated wheat samples. In addition, they are also the only genes expressed during the later stages of grain development and maturation. Although some of the other genes are also expressed in the grain, they show higher levels of expression in other tissues. Besides, transcripts of *YUCCA, Myosin J protein, LRR Kinase-2, ERF_B* and *PP1-Like* were barely detectable in the grain and it is therefore difficult to envisage how these genes could be underlying *Phs* effect in the grain.

Furthermore, the results from the previous physiological and genetic characterisation of *Phs* show the effect of the locus to be late acting, specific to the grain and possibly mediated by hormones. Of all the genes analysed, only the ABA-inducible *PM19* genes show a spatial and temporal expression pattern that agrees with these characteristics. Although the expression patterns of six other genes in the physical map are yet to be analysed, it is tempting to speculate that one (or both) of the *PM19* genes is the main candidate for *Phs*.

4.2.10.2 *Sequence and expression analysis of the PM19 genes*

4.2.10.2.1 *PM19* genes are products of ancestral duplication events

PM19-A1 and *PM19-A2* are single exon genes with sizes of 555 bp and 543 bp, respectively. In the physical map of *Phs*, the *PM19* genes are arranged in tandem (12.5 kb apart) on the same contig and have 92 % sequences identity across the coding sequence. This suggests that they might have arisen from ancient gene duplication events around the time of the wheat-barley divergence.

Similarity searches on the EnsemblPlant database using the coding sequences of these genes showed that there are two and three other homoeologous gene copies on chromosome 5BL and 5DL, respectively. These include Traes_5BL_33ADEE93A1 and Traes_5BL_47475B33B.1 on 5BL and Traes_5DL_26434784A, Traes_5DL_426297FC7 and Traes_5DL_1D17F685A on 5DL. In addition, the two 5BL genes were found to be located in tandem on the

same IWGSC CSS contig (5BL_10924933). These support the idea that the duplication event is not specific to the 4AL chromosome but that it predates the allopolyploidization event that gave rise to the modern bread wheat (Feldman and Levy, 2012).

By comparing the Chinese Spring cDNA sequences (including upstream promoter region) of homoeologous and duplicated *PM19* genes, as well as the four copies of *PM19* in barley (see section 4.2.5 and Table 4.7) and *Brachypodium PM19* (Bradi1g00600), it was possible to establish the phylogenic relationship between these genes. The phylogeny tree showed the *Brachypodium PM19* gene to be divergent from the other *PM19* genes and as such formed a clade on its own. The other *PM19* genes were separated into two main clades (Figure 4.19). One of the clades contained *PM19-A1*, Traes_5BL_47475B33B.1, Traes_5DL_426297FC7, Traes_5DL_26434784A and *HvPM19-1*. This suggests that these wheat genes have true homoeologous relationship and are also orthologous to *HvPM19-1* relationship. It also implies that Traes_5DL_26434784A might have arisen from the duplication of Traes_5DL_426297FC7.

The other clade is composed of two other subclades. One of the subclades contains *PM19-A2*, Traes_5BL_33ADEE93A1 and Traes_5DL_1D17F685A, thus suggesting true homoeology between these genes. The other subclade is barley specific and includes *HvPM19-2, HvPM19-3* and *HvPM19-4*; with *HvPM19-2* and *HvPM19-4* being closer to each other than *HvPM19-3*. This suggests two levels of duplications in this Barley *HvPm19* subclade, with the most recent being between *HvPM19-2* and *HvPM19-4.* Based on these relationships, the homoeologous wheat 5BL and 5DL genes were renamed to indicate their homoeologous relationship (Table 4.8).

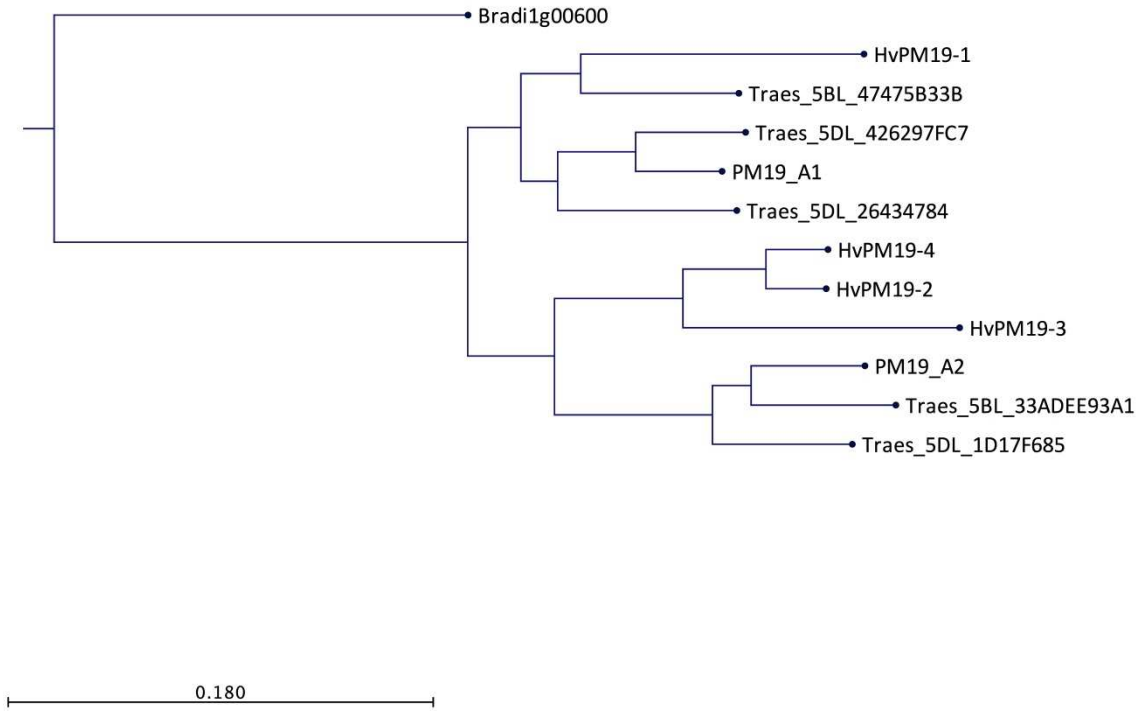


Figure 4.19: Phylogeny analysis of *PM19* genes in wheat, barley and Brachypodium. Maximum likelihood phylogeny analyses of the 4AL, 5BL, 5DL and barley and *Brachypodium PM19* genes. The analysis was done with CLC Genomics using the Juke-cantor model and a bootstrap of 100 replications

Table 4.8: Wheat *PM19* genes and their homoeologous relationships

Genes	Homoeologues	Assigned names
PM19-A1	Traes_5BL_47475B33B.1	*PM19-B1*
	Traes_5DL_426297FC7	*PM19-D1*
PM19-A2	Traes_5BL_33ADEE93A1	*PM19-B2*
	Traes_5DL_1D17F685A	*PM19-D2*
	Traes_5DL_26434784A	*PM19-D3*

4.2.10.2.2 *PM19-A1* and *PM19-A2* are polymorphic in *Phs* Parents

To further investigate the *PM19* genes, a search for functional sequences polymorphisms between resistant and susceptible *Phs* parents was done. Sequencing of the two 4AL *PM19* copies in Claire, Robigus, Option and Alchemy revealed a number of SNPs and deletions within the coding and promoter regions. Parents with similar *Phs* phenotype shared the same haplotype at most SNP position. That is, both Alchemy and Option shared the same allele at all of the SNP loci (except for one SNP locus in *PM19-A2)*, and this was different to the identical alleles between Claire and Robigus. The parental sequences of both *PM19-A1* and *PM19-A2* are presented in Appendix 4 and 5.

Sequence analysis of *PM19-A1* coding sequence, as well as its 500 bp promoter region, uncovered 9 polymorphisms between these varieties (Figure 4.20). These include seven SNPs, with six of these being in the exon. However, only one of the exon SNPs resulted in an amino acid change. The other two polymorphisms identified in *PM19-A1* were insertion/deletions (Indel). These included a 12 bp deletion towards the 3' end of the coding sequence which resulted in the deletion of 4 amino acid residue upon translation. This deletion was only observed in Claire and Robigus. A similarity search on NCBI conserved domain database (Marchler-Bauer et al., 2015) suggests that these amino acid substitution/deletions are in non-conserved residues of the PM19 protein. In addition, a 18 bp deletion in the promoter region, 206 bp away from the start codon, was also found in the resistant parents - Option and Alchemy.

A similar pattern of sequence variation was found in the *PM19-A2* sequences of the parents (Figure 4.20). A total of 13 polymorphisms were found which included 11 SNPs and two indels. Nine of the identified SNPs were located in the exons with only three of them resulting in an amino acid substitution upon translation. The exon contains a 12 bp indel which resulted in the deletion of four amino acid residues at the C-terminal of the protein. Just as was observed for *PM19-A1,* these amino acid substitution/deletions are in non-conserved residues of the protein, and it is therefore hard to predict *a priori* any effect on *PM19-A2* protein function.

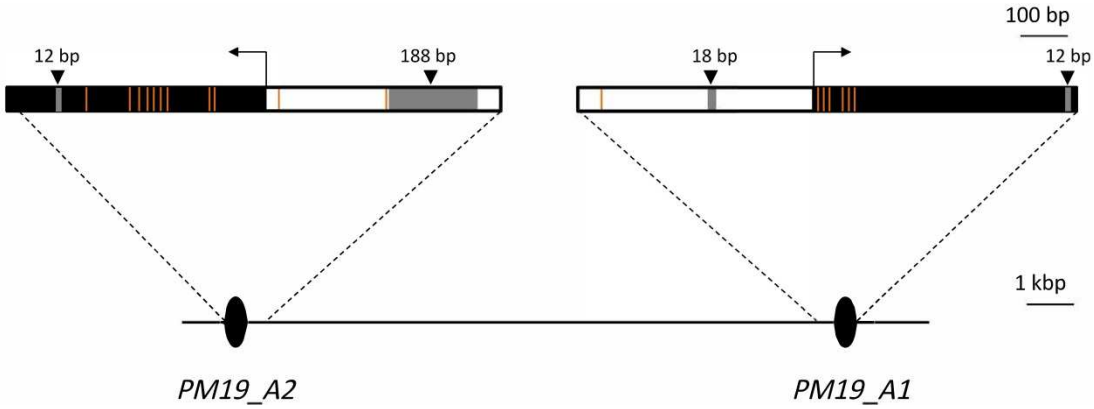


Figure 4.20: The 4AL *PM19* genes have varietal polymorphisms. Schematic representation of *PM19-A1* and *PM19-A2* on the same BAC contig. The SNPs (amber lines) and deletions (grey) found between the susceptible (Robigus and Claire) and resistant (Alchemy and Option) *Phs* parents are represented in the box above the Oval gene models. The black and unfilled regions of the box represent the CDS and promoter region of the genes respectively. The transcription start sites are also indicated with the arrow.

The most notable polymorphism in *PM19-A2* was a 188 bp indel in the promoter region, 259 bp away from the start codon. Alchemy and Option (resistant) have a deletion at this locus while Robigus and Claire (susceptible) have an insertion. Because the indel is in the promoter region which is critical for the regulation of gene expression, we checked for the presence of any known *cis*-regulatory elements in the insertion of Robigus and Claire using a database of Plant Cis-Regulatory Element, PLACE (Higo et al., 1999). This identified a core conserved ABA-responsive regulator element, ACGTG, (Oh et al., 2005) in the indel. The role of this putative regulatory sequence is unknown; therefore more experimental analysis will need to be done to examine the role of this element in the regulation of *PM19-A2*. Table 4.9 below give a summary of all the polymorphisms described for *PM19-A1* and *PM19-A2* above.

Table 4.9: Polymorphisms found in the *PM19* genes between Alchemy and Robigus

Genes	Positions*	Alchemy Allele	Robigus Allele
PM19-A1	-450	T	A
	-(224-206)	18 bp insertion	-
	12	G	T
	15	T	G
	18	C	T
	61	C	A
	72	C	A
	75	C	G
	534	-	12 bp insertion
PM19-A2	-259	-	188 bp Insertion
	-257	C	T
	-30	G	C
	116	A	C
	117	T	C
	222	T	C
	236	A	G
	237	T	C
	240	G	G
	276	T	G
	288	T	C
	386	T	C
	434	-	12 bp insertion

*The positions indicated are relative to Alchemy

4.2.10.2.3 *PM19-A2* is differentially expressed in after-ripened seeds of *Phs* parents

We hypothesise that the presence of a large deletion in the promoter region of *PM19-A2* might affect the expression of the gene such that it is differentially expressed in the varieties segregating for *Phs*. This might in turn account for the *Phs* phenotype of the parent. We first tested this hypothesis by a doing a qPCR analysis of *PM19-A2* in after-ripened seeds of Robigus and Alchemy during a period when the *Phs* effect was observable between the two varieties (as confirmed by an independent GI test). The expression analysis was done at two time points, 0 and 24 Hours after imbibition (HAI). This was because *PM19* was reported to be differential expressed in dormant and non-dormant barley seeds

at 24 HAI (Ranford et al., 2002). Care was taken to avoid sampling already germinated seeds.

The qPCR analysis showed a similar transcript level of *PM19-A2* in both varieties before imbibition, suggesting that both varieties might have accumulated the same amount *PM19-A2* at grain maturation (Figure 4.21). However there was a rapid and significant decline in *PM19-A2* expression at 24 HAI in Robigus (the *Phs* susceptible and non-dormant variety), while *PM19-A2* expression remained high in Alchemy at the same time point after imbibition (Fig 6). Furthermore, the levels of *PM19-A2* transcript in these lines at 24 HAI showed a positive relationship to the dormancy status of the seeds. In Alchemy, which had a higher level of *PM19-A2*, less than 20 % of seeds tested in the GI experiment had germinated while all of the seeds of Claire had germinated at this time point.

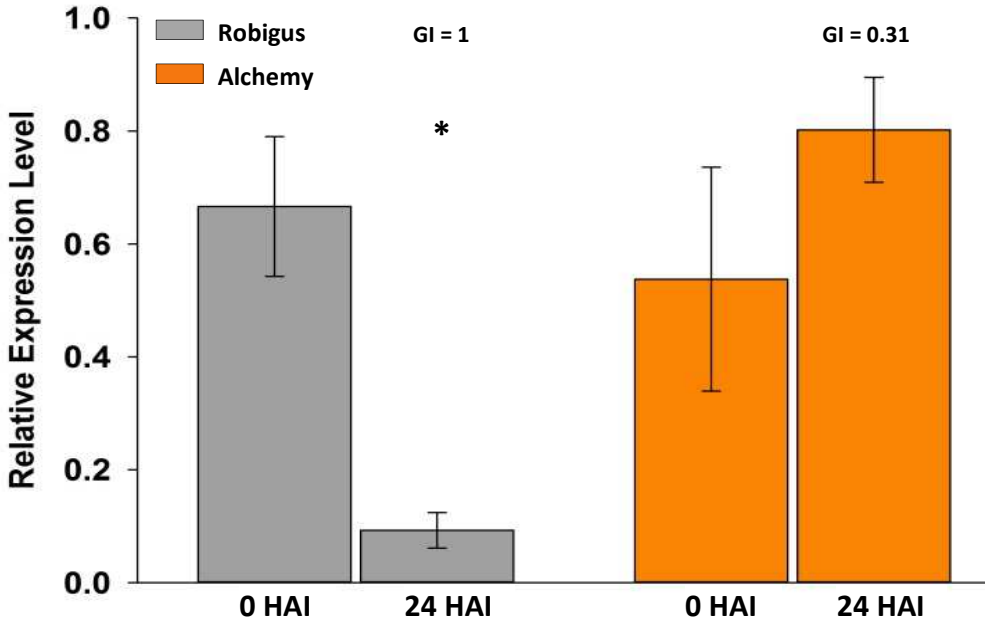


Figure 4.21: *PM19-A2* is differentially expressed post imbibition in Robigus and Alchemy. The figure shows the expression level of *PM19-A2* relative to the reference gene control (*TaEF-1A*) in Robigus and Alchemy at 0 and 24 hours after imbibition (HAI). The error bars are SEM of at least 3 biological replicates and significance at P > 0.05 is indicated (*) based on comparison between expression level at 0 and 24 HAI

To study the differential expression of *PM19-A2* further, a similar qPCR analysis was done in Claire and Option seeds. However, more post-imbibition time points were examined for Option so as to study how long *PM19-A2* expression remains high in the dormant parent after imbibition. For this experiment, only imbibed but non-germinated seeds of each variety were taken.

Similar to what was observed in the Robigus and Alchemy experiment, there was no significant difference in the expression of *PM19-A2* in both varieties before seed imbibition (Figure 4.22). However *PM19-A2* expression declined after imbibition in the susceptible parent (Claire) but not in the resistant parent (Option). Interestingly, *PM19-A2* expression in Option was found to be high in non-germinated seeds even after 8 days of imbibition. Together these results suggest that *PM19-A2* is differentially expressed post-imbibition in parents with contrasting sprouting phenotype.

While it is tempting to speculate that *PM19-A2* is the main candidate for the *Phs* effect due to this differential expression, it is impossible to untangle the cause-effect relationship between *PM19-A2* expression and the sprouting phenotype of these varieties, based solely on this experiment. The differential expression of *PM19-A2* in this experiment could be due to pre-existing differences in the physiological state (dormant/non-dormant) of the seeds prior to the experiment. To proof the causality of the expression of *PM19-A2* on the sprouting phenotype, more precise genetic association or transgenic validation will be required.

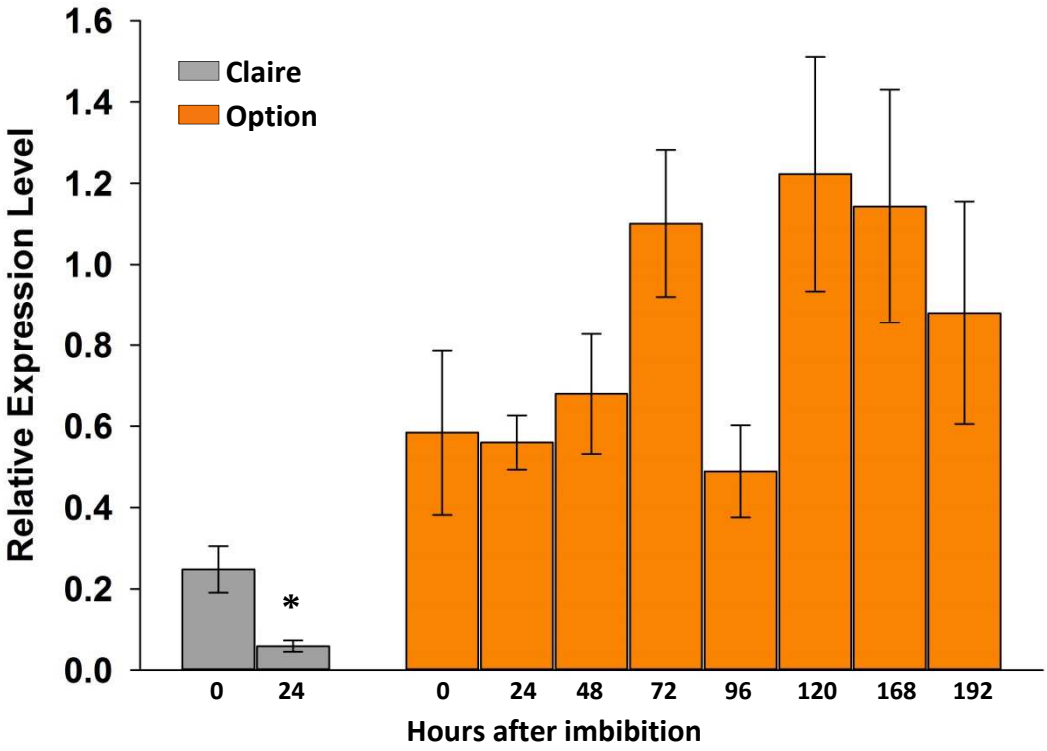


Figure 4.22: *PM19-A2* is differentially expressed post-imbibition in Claire and Option. The figure shows the expression level of *PM19-A2* relative to the reference gene control (*TaEF-1A*) at 0 and 24 hours HAI for Claire and 0, 24, 48, 72, 96, 120, 168 and 192 HAI for Option. The error bars are SEM of at least 3 biological replicates and significance at P > 0.05 is indicated (*).

4.2.10.2.4 The *PM19* genes are differentially regulated in seeds

To find out if the differential expression of *PM19-A2* observed in contrasting *Phs* parents is just specific to this copy of *PM19* or not, a similar qPCR analysis of the other gene copies of *PM19* was done before and after imbibition in the same Robigus and Alchemy samples used before. Homoeologue-specific primer pairs were designed for *PM19-A1, PM19-B1, PM19-B2* and *PM19-D2* and the specificity of amplicon obtained using these primer pairs were confirmed from the dissociation curves analysis.

The expression result shows that the different *PM19* copies are differentially expressed in seeds (Figure 4.23). In un-imbibed Robigus seeds, there was no significant difference in the transcript level of *PM19-A2, PM19-B2* and *PM19-D2*,

however the expression levels of these genes was found be significantly higher than those of *PM19-A1* and *PM19-B1* (P value between <0.001 and 0.011). Also, there was no statistically significant difference in the expression of all the *PM19* genes between Robigus and Alchemy at 0 HAI.

A similar pattern of higher expression of *PM19-A2*, *PM19-B2* and *PM19-D2* than the other two copies was observed in Alchemy as well, but this expression differences were not significant. In Robigus at 24 HAI, all the *PM19* genes examined showed significant reductions in expression levels compared to 0 HAI (P value between < 0.001 and 0.008). This was however not observed in Alchemy as there was no significant difference in the expression levels of all of the *PM19* genes.

Taken together, these results suggest that the downregulation of *PM19-A2* upon imbibition in Robigus (in contrast to Alchemy) is not specific to the *PM19-A2* copy because *PM19-A1, PM19-B1, PM19-B2* and *PM19-D2*, were also downregulated. Furthermore, there is a general pattern of higher expression of some *PM19* copies than other. This suggests that some copies of *PM19* are more active than other and therefore could have more dominant roles. Furthermore, this pattern of expression of the *PM19* genes shows a strong correlation to the sequences similarity between them as inferred from phylogeny analysis (Figure 4.19). That is, genes on the same clade on the phylogeny tree showed a similar pattern of expression. This might be indicative of the presence of different regulatory elements in the promoter sequences of the genes in the two clades.

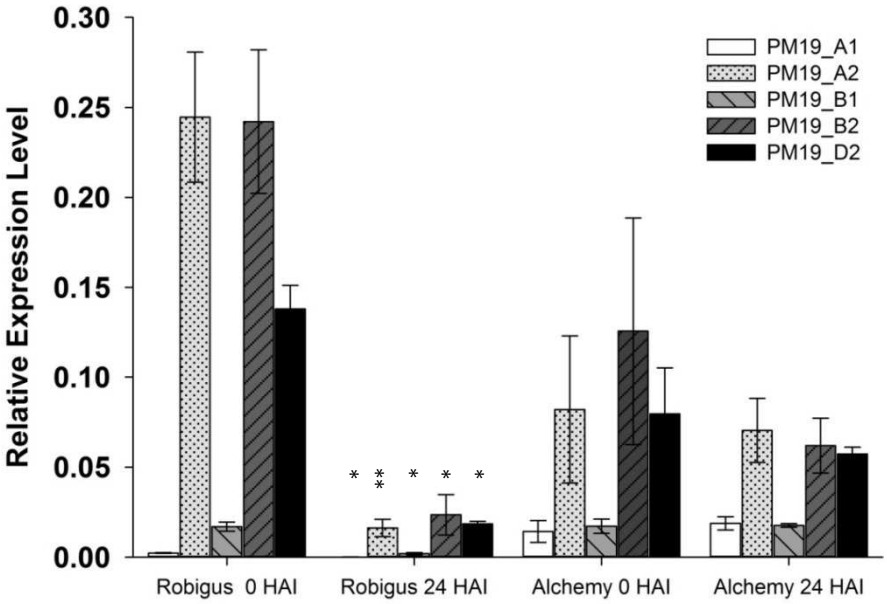


Figure 4.23: The *PM19* genes are differentially regulated. The figure shows the expression level of *PM19-A1, PM19-A2, PM19-B1, PM19-B2* and *PM19-D2* relative to the reference gene control in Robigus and Alchemy at 0 and 24 HAI. The error bars are SEM of at least 3 biological replicates and significance at P <0.001 (**) and at P < 0.05 (8) are indicated.

4.3 Discussion

Fine-mapping of Phs

In this chapter, characterisation, preliminary fine-mapping and physical map construction of a PHS QTL on wheat Chromosome 4AL (renamed *Phs*) has been undertaken. *Phs* plays a crucial role in the global effort to combat PHS. This is because it has been identified in different germplasm and across many geo-ecological zones around the world, including Australia (Li et al., 2004, Ogbonnaya et al., 2007), China (Chen et al., 2008), Japan (Torada et al., 2005), Canada (Cabral et al., 2014) and in Europe (Albrecht et al., 2015). In addition, *Phs* has also been reported to be effective in both white and red grain wheat population (Mares et al., 2005, Kulwal et al., 2012). This disperse and widespread identification of *Phs* across different agro-ecological zones, suggest that *Phs* might have originated from ancestral germplasm and has been introduced into modern germplasm through independent breeding programmes.

Despite its importance, the *Phs* locus has only been fine-mapped but not cloned. Flintham et al. (2002) first fine-mapped the locus to an approximate 8.3 cM interval between *Xwg622* and *Xpsr1316* in a Chinese Spring - Synthetic wheat population. Mares et al. (2005) subsequently mapped the locus to 2.9 cM interval between *Xgwm269* and *barc170*. This was also the same interval reported by Zhang et al. (2008). Ogbonnaya et al. (2008) placed the *Phs* locus at 3.7 cM to *xgwm894* (same as *wms894*) and 9.8 cM to *Xgwm269*. In the same year, Torada et al. (2008) reported the closest linkage to *Phs* at 0.3 cM from *xhbe03*.

In this study, two independent populations were used to fine-map *Phs*. The advantage of such trans-population fine-mapping strategy is that the populations have different recombination patterns and haplotype linkage disequilibrium around the *Phs* locus. These can be used in combination to achieve a higher fine-mapping resolution at the *Phs* locus than would have been possible with each individual population (Teo et al., 2010). Using these two independent populations, we have successful fine-mapped *Phs* to a 0.2 cM interval between *xhbe03* and *wms894*. This genetic interval is smaller than the 2.6 cM (proximal to distal flanking) interval reported by Torada et al. (2008) and as such represents the closest linkage to *Phs* reported, at least in European germplasm.

The physical map of the Phs interval

Importantly, we have been able to access the gene content information about this critical 0.2 cM interval through physical map construction. Until recently, producing an accurate whole genome physical map of hexaploid wheat using the BAC-to-BAC approach was difficult. This was because of the high amount of repetitive element and the high degree of conservation between the homoeologous genic sequence of the polyploid bread wheat genome. However, these complexities have been overcome by the application of latest advances in flow cytogenetic (Dolezel et al., 2011), as well as, the availability of chromosome arm genetic stocks for wheat.

Using these tools, individual chromosome arms of Chinese Spring bread wheat genome can be sorted by flow cytometry and then used to produce chromosome –arm-specific BAC libraries (Šafář et al., 2004, Luo et al., 2010, Fleury et al., 2010). This has resulted in the development of the physical maps of a number of bread wheat chromosome arms, with the physical map for many others being near completion (including the 4AL) (Paux et al., 2008, Raats et al., 2013, Akpinar et al., 2015, Breen et al., 2013, Kobayashi et al., 2015, Lucas et al., 2013, Philippe et al., 2013, Poursarebani et al., 2014). This represents a major breakthrough in wheat genomics and has ultimately enabled the characterisation of the *Phs* locus in this study.

The developed *Phs* physical map has so far revealed the presence of 17 high confidence genes within the 0.2 cM critical interval of *Phs* in wheat. However due to the presence of gaps in the physical map, this number could still grow. A comparative analysis of the *Phs* gene content with the gene contents of syntenic regions in barley and *Brachypodium* showed a level of gene conservation across the species except for an intervening non-syntenic block containing four to five genes (including Bradi1g00630 - Bradi1g00660) which have very low level of gene conservation between the grasses.

A similar analysis of synteny between *Brachypodium* and sorghum showed a break-down of micro-collinearity within the same interval. This, therefore, suggests that the *Phs* interval is contained in a region that has possibly undergone an extensive and ancient structural re-arrangement among grass species. Such chromosomal

rearrangements are not uncommon among grasses (Chantret et al., 2008, FEUILLET and KELLER, 2002) and it has sometimes be attributed to the accumulation of non-syntenic pseudogenes due to the activity of the transposable elements in the Triticeae (Wicker et al., 2011).

Although speculative, this rearrangement might even have huge implications for the *Phs* effect as it could be responsible for the polymorphism underlying the QTL effect (Saxena et al., 2014). In addition, this chromosomal rearrangement could be extensive and diverse among modern wheat varieties with different varieties having different haplotype variation around this locus. We recognize that the allelic variation present in the Chinese Spring variety used for the physical map construction might not be representative of the haplotype of the UK germplasm used in this study. However, the use of the Chinese Spring BAC-to-BAC physical map has enabled access to the gene content of the *Phs* interval. Similarly, Liu et al. (2013b) used the Chinese Spring BAC library to clone *TaMFT* as the causal gene underlying a PHS QTL - *Phs.pseru-3AS*. However, for a more accurate investigation of the causal polymorphisms underlying QTL effects, variety specific BAC library are commonly used in many a positional cloning project (Distelfeld et al., 2006, Fu et al., 2009).

For some of the genes identified in the physical map of *Phs*, there is no evidence to suggest their possible involvement in the *Phs* effect, while other genes have been implicated directly or indirectly in dormancy control and in the induction of germination and hence could be responsible for the *Phs* effect. Genes like *ACC Oxidase-1, ERF_B* and *ERC_C* are involved in ethylene biosynthesis and signaling (Chen et al., 2014a, Yoon and Kieber, 2013, John et al., 1999, Lürssen et al., 1979). Besides its many and varied roles in plant development and defense, ethylene has been shown to promote germination in many different species (Linkies and Leubner-Metzger, 2012, Matilla and Matilla-Vázquez, 2008, Corbineau et al., 2014). Ethylene action in seed germination includes interaction with GA to counter ABA biosynthesis and signaling (Beaudoin et al., 2000, Cheng et al., 2009) as well as the induction of enzymes for the weakening of the endosperm (Wang et al., 2013).

Recently, Chitnis et al. (2014) showed the upregulation of four ACC Oxidase probe sets during imbibition of after-ripened non–dormant wheat seeds, thereby implying that ethylene production in wheat seeds could be mediated by after-ripening. Since *Phs* effect is only observed in after-ripened dry seeds, ethylene could be directly or indirectly involved in this effect. Based on these evidences for the regulation of germination by ethylene, the *ACC Oxidase-1* and the *ERF* genes could be listed as possible candidate genes.

Besides *ACC Oxidase-1* and the *ERFs*, two other genes in the physical map could also have ethylene-related function. One of these is *MAPK-6* which has been shown in Arabidopsis to be part of the ethylene signaling cascade as it is strongly activated by Aminocyclopropane Carboxylic Acid (Ouaked et al., 2003, Yoo et al., 2008, Yoo and Sheen, 2008). Also, the gene encoding for a Hypothetical Protein identified in the physical map has not been characterised; it, however, showed high homology to *ACC Oxidase-1* suggesting that it might also be involved in ethylene biosynthesis.

Of all the genes identified in the physical map, the *PM19* genes (*PM19-A1* and *PM19-A2*) have the strongest evidence from literature in support of their possible involvement in the regulation of dormancy. The *PM19* genes encode for plasma membrane protein 19 which is a protein with four transmembrane domains. Their expression was found to be strongly correlated to freezing tolerance in wheat cell suspension (Koike et al., 1997a). In barley, a *PM19* transcript (now known to be *HvPM19-1*) was found in dry dormant and non-dormant barley seeds but was down-regulated only in non-dormant seeds after imbibition (Ranford et al., 2002). In both of these studies, *PM19* was found to be strongly induced by the dormancy promoting hormone, ABA.

Characterisation of the Phs effects

We undertook further genetic and physiological characterisation of the *Phs* effect in the hope that this will help establish the characteristic of the underlying causal gene of *Phs*. The study of the *Phs* effect in F_1 seeds obtained from reciprocal crosses between Claire and Option showed two important characteristics of the *Phs* effect. The first is that it originates from the zygotic tissue of the grain (embryo or endosperm). This is because the effect was observed in all the progeny obtained from the reciprocal crosses irrespective of the phenotype of the maternal and paternal parents.

This suggests that the *Phs* effect is not as a result of seed coat (testa or pericarp) imposed dormancy. This agrees with finding by Flintham (2000) who found the *Phs* effect to be independent of the Red grained gene (*R* gene) effect on seed dormancy manifested in the testa. Given these evidences, it can be concluded that the *Phs* effect is not mediated by biophysical or biochemical properties of the testa or pericarp which in part regulates dormancy in some species (MacGregor et al., 2015, Debeaujon et al., 2000, Chatterjee and Nagarajan, 2006, Leubner-Metzger, 2002)

Similarly, *TaMFT* - the gene underlying the major PHS resistance locus on chromosome 3A (*Phs.ocs-3A.1)* - was also found to be expressed in the embryo (Mori et al., 2005, Nakamura et al., 2011a, Liu et al., 2013b). Considering that *Phs.ocs-3A.1* and *Phs* have the largest and most stable effect on PHS resistance, it is tempting to speculate that zygotic-tissue-imposed seed dormancy play a bigger role in wheat seed dormancy than previously thought. *Phs* expression is also similar to the dormancy effect of *qSD12* QTL in wild rice which was found to control offspring tissue–imposed seed dormancy in after-ripened seeds. (Gu et al., 2008).

In addition, the reciprocal cross experiment also suggests the effect of *Phs* to be semi or partially dominant. This is because F_1 seeds from the reciprocal crosses were less dormant than the dormant parent but were still significantly more dormant than the non-dormant parent. This agrees in part with the study by Flintham (2000) in which F_1 seeds from reciprocal crosses between two *Phs* segregating UK varieties showed a midway GI phenotype relative to those of the susceptible and resistant variety.

Together, these observations support the notion by Noll et al. (1982) and Belderok (1963) that embryo dormancy displays a partial dominant phenotype.

Furthermore, our hormonal characterisation experiment showed an inhibition of the *Phs* effect by exogenous application of GA and fluridone (an ABA biosynthesis inhibitor). This was through the reduction of dormancy in the variety with the resistant *Phs* allele. Both treatments also showed additive inhibitory effect on *Phs* when used in combination. This is suggestive of an ABA-GA interaction in the control of *Phs* effect.

Unlike GA, fluridone also increases the dormancy of the non-dormant *Phs* parent while reducing the dormancy of the dormant parent. This is surprising given that fluridone is an inhibitor of ABA (an inhibitor of germination) and as such fluridone treatment is expected to increase and not decrease germination. One reason for this might be that concentration of fluridone used is too high or that the fluridone solvent used (ethanol) inhibits germination (MIYOSHI and SATO, 1997). However, no inhibitory effect of fluridone concentration or solvent was observed in the Option seeds. This experiment thus needs to be repeated with appropriate controls that will aid a more conclusive interpretation of the result.

In light of this potential dual effect of fluridone, we hypothesise that ABA biosynthesis could have a more dominant role in the ABA-GA interaction that mediates the *Phs* effect. Similar to this, the underlying gene of the *qSD12* locus in rice has been shown by high resolution fine-mapping and characterisation to promotes ABA accumulation in developing seeds (Gu et al., 2010). More experiment will, however, need to be done to unravel the ABA-GA interaction in the control of *Phs* effect in wheat.

Interestingly, treatment with ACC did not have any effect on the *Phs* phenotypes of the lines, as did treatments with Cobalt Chloride - ethylene biosynthesis inhibitor (result not shown). Although not conclusive, this potentially suggests that ethylene does not play a major role in the *Phs* effect. Also, it is important to note that the hormonal effect seen in these parental lines could emanate from other loci besides *Phs*. However, this is quite unlikely due to the predominant effect of *Phs* in after-ripened seed in the populations studied. Nevertheless, this experiment needs to be

repeated in isogenic recombinant line segregating only at the *Phs* locus so as to determine its linkage with *Phs*.

PM19-A2 as the major candidate gene for Phs

These genetic and hormonal experiments have helped to define the characteristic of the gene(s) underlying *Phs* effect. These characteristics, along with a spatio-temporal expression analysis of some of the identified genes, point equivocally to the *PM19* genes as the main candidate genes underlying *Phs* effect. Unlike the other genes, both *PM19-A1* and *PM19-A2* were specifically expressed in grains starting from the later stages of grain maturation in untreated plant samples. In addition, both genes appear to be highly polymorphic in contrasting *Phs* parents, with the most notable of these polymorphisms being a 188 bp indel in the promoter region of *PM19-A2*. This indel contains an ABA-responsive element (ABRE) which could be important for the regulation of the gene. Although not functional validated in wheat, similar ABRE and a drought responsive element (DRE) were identified in barley *HvPM19* (Sreenivasulu et al., 2006) and might account for the ABA responsiveness of *HvPM19*.

These finds are also supported by a very recent report by Barrero et al. (2015) in which *PM19-A1* and *PM19-A2* were identified as a candidate for a 4AL dormancy QTL in wheat (same as *Phs*). This work was only published within the last four months of this Ph.D. project (May 2015). Barrero and colleagues used bulked segregant RNAseq analysis and the high recombination frequencies afforded by a large multi-parental population to delimit the *Phs* effect to the two *PM19* genes. They also developed a physical map around the *PM19* loci on the A, B and D wheat genomes. The physical map developed by Barrero *et al*. and the physical map reported in this study identified the same set of genes in exactly the same order. This includes *LRR Kinase, ACC Oxidase Ubiquitin Conjugating Enzyme, Myosin J, PM19-A1, PM19-2 and YUCCA.* This further validates the accuracy of our physical map. However, our physical map covers a larger *Phs* interval than reported by Barrero *et al*.

In this study, we found the *PM19* genes on the A genome to be highly polymorphic between dormant and non-dormant parent. This is also supported by the Barrero *et al.* study except that the polymorphism they described differs in many ways to the polymorphisms found in this study. For instance, Barrero *et al*. reported three non–synonymous SNP out of the six SNPs in the exon of *PM19-A1* while in this study only one of the six SNPs identified was non-synonymous. They also did not identify the 12 bp polymorphic deletion which we detected in the coding sequence of *PM19-A1* (Table 4.9). Similarly, only one SNP was identified in the coding sequence of *PM19-A2* by Barrero and colleagues, while in this study we identified 9 SNPs of which three were non-synonymous. This clearly suggests differences in haplotype between Australian and UK germplasm at these loci, and this might have functional implication on the regulation and role of *PM19* in these germplasms.

Our expression analysis showed that *PM19-A2* (with the large promoter deletions) is differentially expressed in imbibed after-ripened seeds of *Phs* susceptible and resistant varieties. This supports the hypothesis that the promoter deletion has functional implication on *PM19-A2* expression and consequently on *Phs* effect. In contrast, Barrero et al. showed *PM19-A1* as the gene copy which is differentially regulated in dormant and non-dormant line. One potential reason for this difference is the timing of the expression analysis done in both studies. In this study, the *PM19* genes expression was studied in 3 - 4 weeks after-ripened grain unlike in the Barrero study where *PM19* expression was studied at 15, 20, 30 and 40 days post-anthesis. It is, therefore, possible that the *PM19* genes are differentially regulated depending on the stage of grain development and after-ripening.

Despite the differences in the polymorphism and expression pattern of *PM19-A1* and *PM19-A2* in this study and in the Barrero *et al.* study, both studies suggest a positive role of the *PM19* genes in the regulation of the dormancy. This was validated in transgenic RNAi plants with reduced expression of *PM19-A1* and *PM19-A2.* These RNAi plants showed a significant reduction in dormancy relative to the control (Barrero et al., 2015).

However, it is not yet clear how PM19 regulates dormancy in grains. One possible mechanism could be that, PM19, being a plasma membrane protein, could be involved in the maintenance of cellular membrane integrity under the physical stress imposed during seed desiccation (Golovina et al., 2001). *PM19* could also have a protective role against the damaging effect of reactive oxygen species (ROS) produced during the imbibition of after-ripening seeds (Leymarie et al., 2012, Gao et al., 2013a).

In support of this role of *PM19* in alleviating abiotic stress in grain, Li Y-C (2012) found *PM19* to induced in root under ABA treatment but also under drought, salt and heat stress. This suggest that the role of the PM19 is not just limited to seed dormancy but that it has a wider role in modulating ABA-dependent plant response to abiotic stress (Tuteja, 2007, Koike et al., 1997b). More work will need to be done to characterise fully the specific role of PM19 in seeds and the general role of PM19 in plants.

5 Chapter Five: Towards the positional cloning of *Phs*

Chapter Abstract

In the previous chapter, we undertook the construction of the physical map of the *Phs* locus and also identified potential candidate genes. In this chapter, we describe the high-resolution fine-mapping of *Phs* in three independent and diverse populations. This work allowed us to delimit *Phs* to a 10 kb genomic interval. Finally, we report on the comparative sequence analysis around this critical interval and we show the presence of some genomic lesions that could be critical for the *Phs* effect.

5.1 Introduction

5.1.1 Gene identification through positional cloning

Positional (or map-based) cloning is a gene identification method that involves the progressive mapping of a phenotype of interest to an ever narrower chromosomal interval until the lesions/mutations responsible for the phenotype of interest are found. Positional cloning can alternatively be seen as a way of excluding all non-causal part of the genome until a small region with the causative polymorphism(s) for a phenotype of interest is obtained (Lukowitz et al., 2000, Gallavotti and Whipple, 2015). Two levels of mapping generally precede a positional cloning effort. These include a preliminary mapping (rough mapping or QTL analysis) to delimit the target locus to a broad interval. This is followed by a fine-scaled mapping (fine-mapping) that further narrows down the causal locus. Depending on the resolution of mapping achieved, this can be a reiterative process until the causal locus is eventually delimited to a genetic distance that is within the resolution of a physical map and which contains a limited number of candidate genes (Chi et al., 2008).

Several factors affect the outcome of any positional cloning effort. Some of these include the nature of the phenotype of interest, the recombination frequencies around the causal locus and the availability of genetic marker /physical sequence map around the locus (Tanksley et al., 1995, Etherington et al., 2014). Low phenotype penetrance

or expressivity (Miko, 2008, Cooper et al., 2013, Lempe et al., 2013), makes fine-mapping of the phenotype of interest tricky and difficult given the need to unequivocally score recombinant plants. However for many Mendelian traits, this is less of an issue.

The use of meiotic recombination to break/disrupt linkage disequilibrium between physically linked genes underpins any positional cloning effort. Availability of recombination can therefore be the most important factor limiting the success of cloning. Genetic recombination occurs through the exchange of genetic material (crossovers, CO) between homologous chromosomes during meiosis. The rate of these CO events varies extensively along the length of the chromosome.

Several studies in different plants including Arabidopsis, rice, barley and wheat, have revealed a general pattern of higher frequencies of CO towards the telomere and a suppression of COs towards the centromere (Saintenac et al., 2009, Drouaud et al., 2006, Wu et al., 2003, International Barley Genome Sequencing Consortium, 2012). Moreover, some chromosomal region (called recombination hotspot) have been associated with higher levels of CO events (Mezard, 2006). In many cases, recombination hotspots are found in gene-rich regions of the genome (Faris et al., 2000, Sidhu and Gill, 2005). In addition, the repetitive element content of genomes also negatively affects the rate of recombination (Henderson, 2012).

Up until recently, attempting positional cloning of genes in wheat was cumbersome, time consuming and expensive. This was mainly due to the lack of genetic (markers) and genomic resources (genome sequences) necessary to achieve high-resolution mapping. However, with the advent of NGS technologies, there have been an explosion in the number of sequence (Chapman et al., 2015, International Wheat Genome Sequencing Consortium (IWGSC), 2014, Brenchley et al., 2012b) and marker (Cavanagh et al., 2013, Jordan et al., 2015, Wang et al., 2014) resources available to wheat researchers . This means that positional cloning projects in wheat are no longer held back by the availability of genetic markers or physical map sequences, although the suitability of the available physical maps for different positional cloning projects might still be an issue.

Instead, limited recombination frequencies and instability/complexity of phenotypes constitute the main limitations to positional cloning efforts in wheat. While the problem of phenotype instability and complexity can be partly solved by applying emerging high-throughput and precision phenotyping approaches (Araus and Cairns, 2014, Furbank and Tester, 2011), the challenges posed by the constraint of recombination still constitute a bottleneck for some positional cloning efforts. To circumvent this problem, alternative NGS-based approaches that do not depend exclusively on recombination events are being used to complement (or replace in some cases) map-based cloning method (Trick et al., 2012, Jupe et al., 2013, Mascher et al., 2014).

5.1.2 Aims

In the previous chapter, the fine-mapping of *Phs* to a 0.2 cM interval, as well as the construction of its physical map was described. This has provided us with sufficient genetic and genomic resources necessary for the positional cloning of the locus. Our aim in this chapter is to attempt the positional cloning of *Phs* by further high-resolution fine-mapping in diverse populations.

5.2 Results

5.2.1 Development of high-throughput SNP-based genetic markers within the *Phs* interval

Phs was fine-mapped in the previous chapter to a 0.2 cM interval between the SSR markers: *wms894* and *xhbe03*. Despite the close proximity of these markers, construction of the physical map between this interval show that there are at least 17 genes models spread over a physical distance of more than 1.25 Mb (minimum size of sequenced BAC cluster in Chapter Four, section 4.2.6). To detect critical recombination events within this large physical distance, it is important to develop additional markers across the physical map of *Phs*. Moreover, *wms894* and *xhbe03* are SSR markers which are not well adapted nor cost efficient for high-throughput genotyping (Ganal and Röder, 2007). In view of this, we undertook the development of more genetic markers within the *Phs* interval.

Development of genetic markers within the *Phs* interval was done at two stages. The first was synteny-assisted marker development in which corresponding parental sequences of the genes identified by synteny (Chapter Four, section 4.2.4) were obtained through Sanger sequencing. Polymorphisms identified between contrasting *Phs* parents were subsequently developed into SNP markers which can be easily assayed using the cost effective and highly-scalable KBioscience's Kompetitive Allele-Specific PCR amplification of target sequences (KASPar®; Smith and Maughan, 2015). Using this approach, three redundant markers were developed each for *PM19-A1, ACC Oxidase-1* and *Activating Signal Co-integrator*. This approach also yielded two redundant markers each for *PM19-A2* and *ERF_B,* while the *PP1-Like* and *Hypothetical protein* (homologous to Bradi1g00710) genes had a single SNP marker each (see Appendix 1 and 2 for marker sequence information).

The synteny-assisted marker development approach helped in the generation of new markers within the *Phs* interval, it however was limited by the lack of sequences information in the non-syntenic region of the *Phs* interval. Another drawback of this approach is the time taken to identify polymorphisms between contrasting *Phs* parents through individual gene amplification and Sanger sequencing. The generation of

physical sequence information for the *Phs* interval, as well as the sequencing of flow-sorted 4AL chromosome arm from Claire and Option helped to overcome these limitations and thereby accelerated the marker development process (Chapter four, section 4.2.7).

Using these new resources, new polymorphisms (SNPs and indels) were uncovered between resistant Option and susceptible Claire parents, especially in the region of the physical map with low gene conservation to barley and *Brachypodium* (see Chapter four, section 4.2.6 and Figure 5.1). Some of these polymorphisms were developed into KASPar markers for high-throughput genotyping. These include four redundant markers for *ERF_C*, two redundant markers each for *LRR Kinase-1* and *RPM1*; and a single marker each for the *Agmatine Coumaroyl Transferase* and the *Hypothetical Protein* (similar to *ACC- Oxidase 1*).

Markers *151_A19_T7_SNP9,* and *170_N01_T7_Del* and *031 _F05_T7_SNP1* were developed from the BAC end sequence of some of the BACs within the *Phs* interval. Other markers (*052_N17_SNP2 and 032_ctg1_SNP1)* were developed from the intergenic region of *Phs* physical sequence. In addition to these markers, two additional markers were developed from the iSelect SNP array. These include markers STW17 (BS00159699) and STW18 (IACX2890). These markers are located on wheat genes homologous to *BRADI1G00820* and *BRADI1G00730*, respectively.

Most of the markers developed behave in a co-dominant fashion. That is, they are able to detect both alleles of a SNP and are therefore able to distinguish heterozygous mutations from homozygous mutations (Allen et al., 2013) . This important characteristic of the markers makes them suitable not only for defining recombination points but also for screening for new recombination events in F_2 material with high heterozygosity.

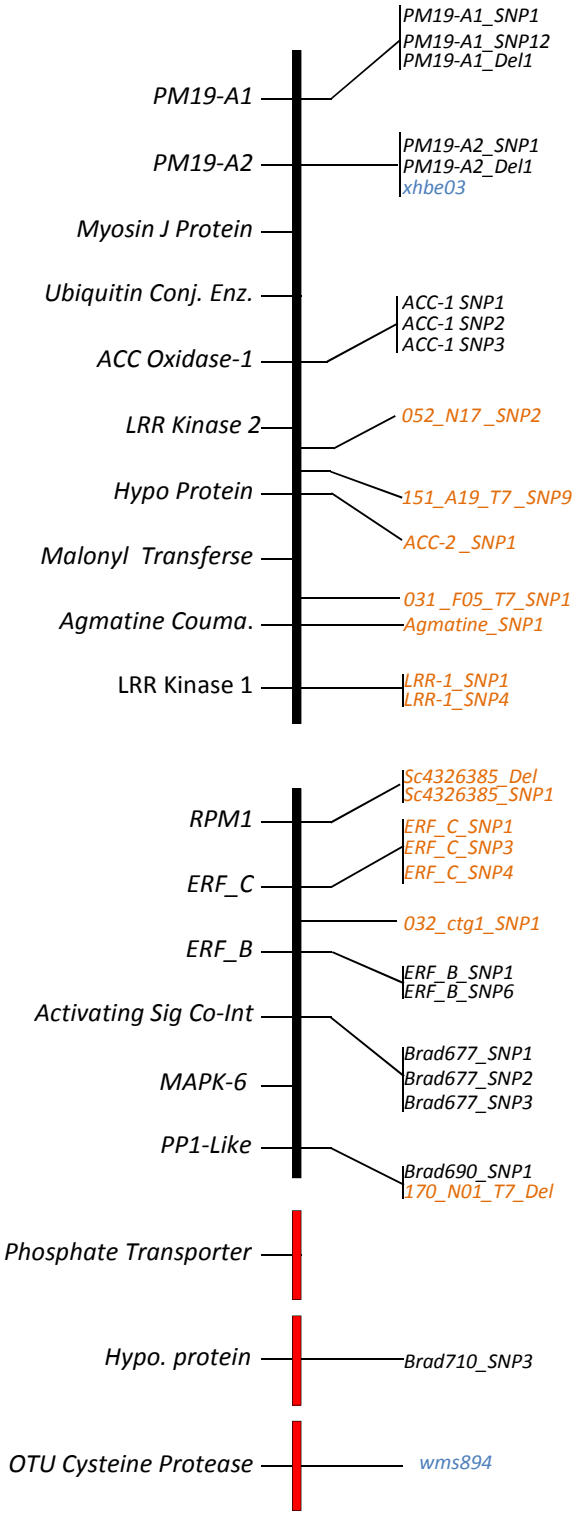


Figure 5.1: Development of high-throughput genetic markers across the *Phs* interval: Thirty redundant SNP-based genetic markers were developed across the *Phs* interval. Markers highlighted in amber were developed from the physical map sequences while the other markers were developed from synteny-derived IWGSC sequences. The SSR markers previously used for the fine-mapping of *Phs* (Chapter Four, section 4.2.3.2) are highlighted in blue

5.2.2 High-resolution fine-mapping in bi-parental populations delimits the *Phs* locus to 10 genes

The development of more genetic markers within the *Phs* interval made it possible to further fine-map *Phs* with higher resolution and possibly enable its delimitation to a single gene within the physical map. Using these markers, high-resolution fine-mapping was performed in 5 different experiments using three independent and diverse populations. These experiments are discussed in this section.

5.2.2.1 *Fine-mapping experiment-1 with the Option x Claire Population*

The previous chapter described the fine-mapping of *Phs* to a 0.2 cM interval between markers *wms894* and *xhbe03* in the Option x Claire population (OC Pop) with 131 lines which were derived from 27 independent F_2 recombinant lines. However, this previous experiment had low marker coverage and limited experimental replications. To address this, we selected a subset of 34 lines from this OC Pop and genotyped these lines using some of the newly developed SNP markers and the flanking SSR markers. At the time of this experiment the physical map sequence had not been obtained. Therefore only markers developed from syntenic genes were used. These include markers for *PM19-A1 (PM19-A1 _SNP1), PM19-A2 (xhbe03), ACC Oxidasse-1 (ACC1_SNP1), ERF_B (ERF_B_SNP9), Activation Signal Co-int. (Brad677_SNP1), PP1-Like (Brad690-SNP2)* and *Hypothetical Protein (Brad710_SNP1*).

Genotyping of the selected recombinant lines with these markers identified six recombination events within the *Phs* interval resulting in 8 different recombination haplotypes (Figure 5.2). These included recombination events between *xhbe03* and *ACC1_SNP1; ERF_B_SNP6* and *ACC1-SNP1* (two events); *Brad690_SNP2* and *Brad710_SNP1* and two recombination events between *Brad710_SNP1* and *wms894* (Figure 5.2). There were no recombination events between *ERF_B_SNP6, Brad677_SNP1* and *Brad690_SNP2* in all the lines examined.

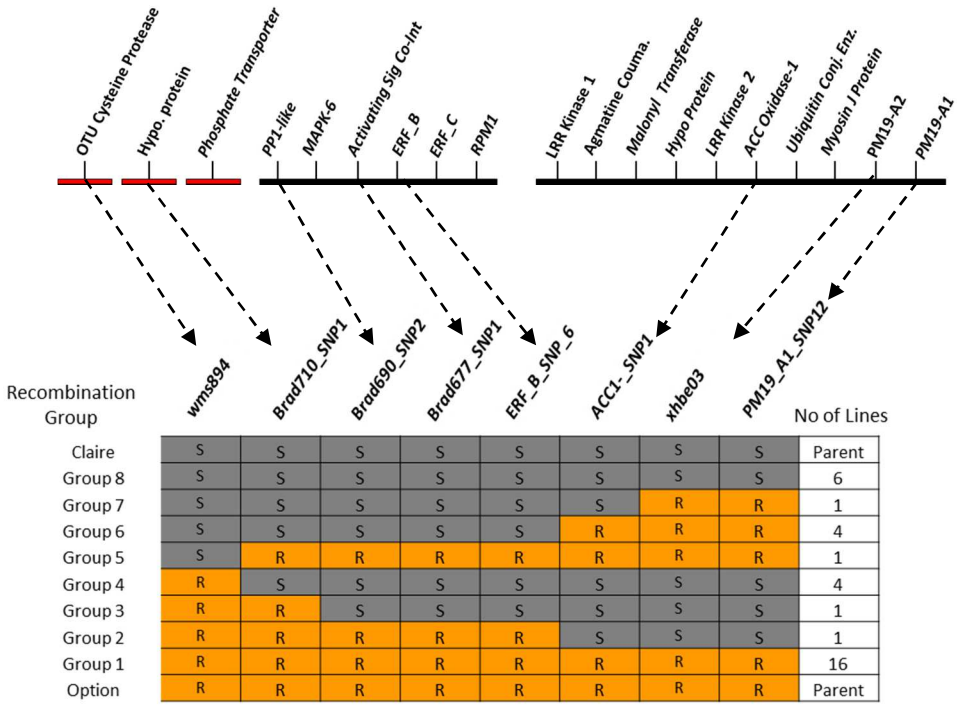


Figure 5.2: Recombination haplotypes in the OC population: The different recombination events present in the OC population were defined using genetic markers developed across the *Phs* interval. The positions of the markers in the physical map are indicated by the connecting arrows. Eight different recombination haplotypes were identified in this population. Claire and Option alleles are represented by C (grey) and O (amber), respectively. The number of lines containing the different recombination haplotypes that were used for the fine-mapping experiment are indicated in the rightmost column.

We examined the sprouting phenotype of glasshouse grown plants for each of the selected lines, as well as the Claire and Option parents. A large variation in sprouting percentage was observed in this population with sprouting scores ranging from 0 % to 88 %. Despite this variation, a bimodal distribution of the sprouting percentages was observed (Figure 5.3). Half of the lines examined (17 lines) as well as Option showed little signs of sprouting (0-10 %), while the other 17 lines and Claire showed moderate to high levels of sprouting (36 – 88 %).

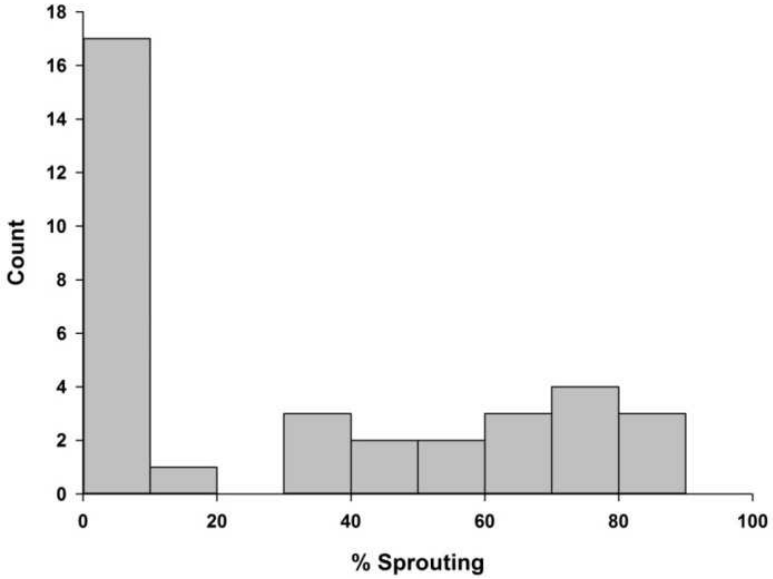


Figure 5.3: Bimodal distribution of the sprouting phenotype in the OC population. The frequency distribution of the sprouting percentage obtained from OC recombinant lines.

An ANOVA analysis done on Logit transformed percentage sprouting data showed significant differences between the sprouting phenotypes of the lines. Furthermore, the Dunnett's method of multiple comparisons was used to independently identify lines with significantly different transformed sprouting percentage relatively to those of Claire and Option parental controls. Based on this comparison, lines were designated as low sprouting only if they showed significantly lower sprouting percentage to Claire and also non-significantly different sprouting percentage to Option. Conversely, lines were designated as high sprouting if their sprouting percentages were significantly higher than Option's and non-significantly different to Claire's (Table 5.1).

Table 5.1: Dunnett multiple comparative analysis of the sprouting percentages obtained in the OC Population fine-mapping experiment 1

Line	plants per line (n)	% Sprouting	Sig (P*) against Option	Sig (P*) against Claire	*Phs* Score
OPTION	4	2	NA	0.000	Low
C x O 4/93/10	6	3	1.000	0.000	Low
C x O 20/51/9	2	2	1.000	0.000	Low
C x O 22/57/8	4	2	1.000	0.000	Low
C x O 26/37/1	6	9	0.069	0.000	Low
C x O 26/37/2	3	10	0.099	0.000	Low
C x O 26/37/6	6	4	0.612	0.000	Low
C x O 26/37/8	6	6	0.346	0.000	Low
C x O 26/48/4	6	1	1.000	0.000	Low
C x O 26/48/5	6	0	1.000	0.000	Low
C x O 26/48/8	5	0	1.000	0.000	Low
C x O 5/101/10	6	1	1.000	0.000	Low
C x O 5/206/10	4	6	0.644	0.000	Low
C x O 5/206/8	4	5	0.568	0.000	Low
C x O 24/47/12	6	3	0.982	0.000	Low
C x O 24/47/11	6	5	0.839	0.000	Low
C x O 21/46/8	5	5	0.332	0.000	Low
C x O 4/204/2	3	0	0.992	0.000	Low
C x O 4/95/12	2	75	0.000	1.000	High
C x O 20/24/5	6	63	0.000	1.000	High
C x O 24/28/2	6	75	0.000	1.000	High
C x O 25/58/8	6	36	0.000	0.134	High
C x O 20/24/3	6	57	0.000	0.999	High
C x O 22/57/1	6	74	0.000	1.000	High
C x O 20/42/10	6	49	0.000	0.832	High
C x O 21/19/8	6	34	0.000	0.094	High
C x O 26/8/9	6	37	0.000	0.159	High
C x O 5/189/8	4	88	0.000	0.640	High
C x O 26/48/7	6	54	0.000	0.991	High
C x O 4/95/11	4	64	0.000	1.000	High
C x O 5/206/12	5	82	0.000	0.985	High
C x O 5/43/7	6	88	0.000	0.511	High
C x O 4/53/12	4	65	0.000	1.000	High
C x O 25/58/1	6	41	0.000	0.327	High
CLAIRE	6	71	0.000	NA	High

* P represents the adjusted P values obtained from the Dunnett's multiple comparison analysis of Logit transformed sprouting percentage, using Claire or Options as control.

A correlation of the genotypic and phenotypic score for each line, showed a perfect linkage between *Phs* and the *ERF_B_SNP6, Brad677_SNP1* and *Brad690_SNP2* markers (Figure 5.4). The fine-mapping also dissociates the phenotype from markers *wms894, xhbe03 and PM19-A1_SNP1* and showed *Phs* to be flanked by markers *Brad710_SNP1* and *ACC1-SNP1.* However, due to the lack of physical map information at the time of this phenotyping, it was not possible to completely exclude the genes containing the new flanking markers (*Brad710_SNP1* and *ACC1-SNP1*). This is because, these two

markers are located in the promoter region of their respective genes, where recombination events are known to occur frequently (Xu et al., 1995, Hellsten et al., 2013). Thus depending on the site of recombination around these markers and the orientation of their respective genes, it is quite possible that these genes could still be in the linkage interval. This can, however, be clarified with information on the orientation of the genes from the physical map.

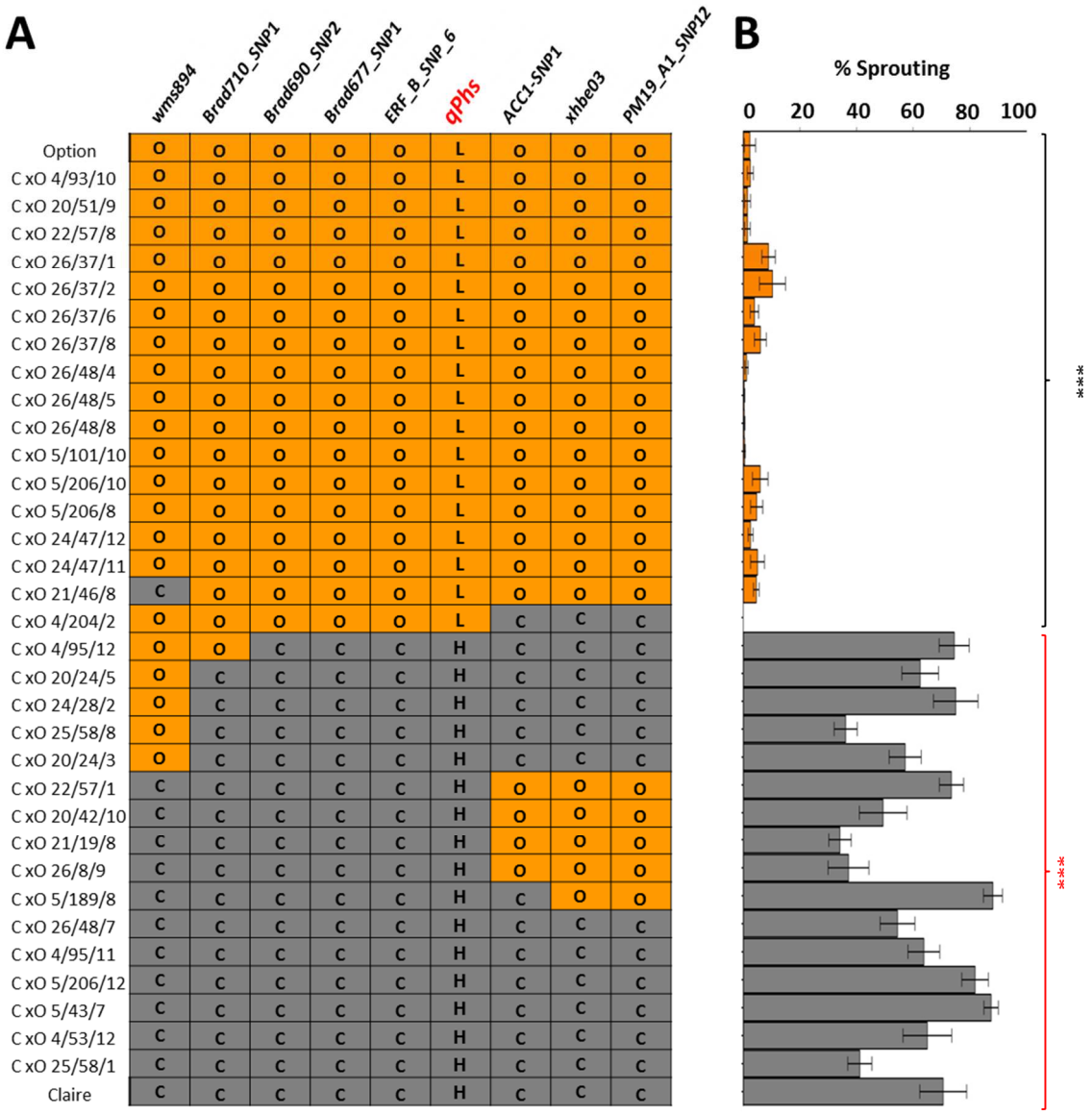


Figure 5.4: Fine-mapping of *Phs* in the OC population (experiment-1). Correlation of the marker genotypes of the selected recombinant lines from the OC population (A) with the sprouting percentages observed in these lines (B). Claire and Option alleles are represented by C (grey) and O (amber), respectively. Designation of the *Phs* phenotype of these lines as low (L) or high (H) is based on the statistical comparative analysis presented in Table 5.1. The error bar in the panel B represents standard errors of means (SEM) of biological replicates with n indicated in Table 5.1. Adjusted significance value ($P < 0.001$) from the Dunnett comparison of Logit transformed sprouting percentage against Claire (black asterisk) and Option (red asterisk) are indicated.

5.2.2.2 *Fine-mapping experiment-2 with Option x Claire population*

To further delimit *Phs*, we searched for more recombination events in the other 97 lines from the OC Population which were not phenotyped in experiment 1. The availability of the physical map sequence at the time of this experiment meant that it was possible to use more SNP markers to further define the recombination event in this population. These include markers *ERF_C_SNP1*, *052_N17_ctg1*, and *151_A19_T*7. The *STW18* SNP marker located on the wheat gene homologous to *Bradi1g00730* was used in place of the SSR *wms894*. This allowed for the identification of 12 more recombinant events: 10 between *052_N17_ctg1* and *ACC1_SNP1 and* two *between STW18* and *Brad710_SNP1.* However, there was no recombination between any of the genes in the *Brad690_SNP2* - *052_N17_ctg1* interval. These new recombinants along with the critical recombinants identified in the previous experiment were grown in a Controlled Environment Room (CER) and harvested spikes from these were phenotyped as in experiment-1.

Similar to experiment-1, a large variation in sprouting percentages was observed in this experiment, with sprouting scores ranging from 1 % to 98 %. However, unlike the result of the previous experiment, the distribution of the phenotype was not bimodal. In addition, a higher level of sprouting was observed in this experiment. With the exception of Option, all the lines had sprouting score greater than 20 %. One possible reason for this higher level of sprouting might be due to difference in the environmental temperature in which the plants in these two experiments were grown (CER and glasshouse; see Chapter Two, section 2.2).

A similar approach as in experiment-1 was used to assign sprouting phenotype to lines (ANOVA and Dunnett's comparison on Logit transformed percentages). This identified high sprouting lines with significantly higher sprouting percentage than Option but not Claire (Table 5.2). However, we could not confidently call lines as low sprouting because all the lines examined had significantly higher sprouting percentage than Option. Despite this high level of sprouting than Option, these recombinants still showed significantly lower levels of sprouting compared to Claire and these were therefore described as having moderate level of sprouting (Table 5.2).

Table 5.2: Dunnett multiple comparative analysis of the sprouting percentages obtained in Option x Claire fine-mapping experiment-2

Line	Plants per line (n)	% Sprouting	Sig (P*) against Option	Sig (P*) against Claire	*Phs* Score
Option	6	1	NA	0	Low
C x O 4/204/10	6	22	0	0	Moderate⁺
C x O 27/8/12	6	27	0	0	Moderate⁺
C x O 27/8/8	6	39	0	0	Moderate⁺
C x O 4/204/1	6	39	0	0	Moderate⁺
C x O 21/46/8	6	50	0	0	Moderate⁺
C x O 26/37/2	6	69	0	0.155	High
C x O 21/19/8	6	89	0	1	High
C x O 26/8/9	6	90	0	1	High
C x O 21/19/7	6	91	0	1	High
C x O 26/8/5	6	92	0	1	High
C x O 21/19/1	6	93	0	1	High
C x O 22/57/7	6	94	0	0.999	High
C x O 22/57/1	6	95	0	0.19	High
C x O 20/42/9	6	96	0	0.666	High
C x O 20/42/10	6	99	0	0.001	High⁺
C x O 20/42/8	6	97	0	0.201	High
C x O 20/24/3	6	93	0	1	High
C x O 20/24/5	6	93	0	1	High
C x O 24/28/2	6	94	0	0.995	High
C x O 25/58/8	4	97	0	0.466	High
C x O 24/28/11	6	96	0	0.606	High
C x O 4/95/12	6	96	0	0.16	High
Claire	6	90	0	NA	High

*P represents the adjusted P values obtained from the Dunnett multiple comparison analysis of Logit transformed sprouting percentage, using Claire or Options as control.

⁺ Lines with statistically different sprouting values compared to both Option and Claire are described as having a moderate sprouting phenotype except for line C x O 20/42/10 which was designated as high because of it higher sprouting percentage than Claire.

When analysed together, the genotype and phenotypes information suggests that *Phs* is linked to the six markers located between and including *Brad690_SNP2 and 052_N17_ctg1* in all but one of the recombinants designated as high sprouting (Figure 5.5). The only exception to this was the recombinant line C x O 26/37/2 which showed a lower but non-significantly different sprouting score to Claire. Surprisingly, in experiment 1, this recombinant line showed a more resistant sprouting phenotype similar to Option (10%). This aberrant behaviour of C x O 26/37/2 could not be attributed to any genetic marker in the *Phs* interval and therefore could be due to the confounding effect of the interaction of few segregating background loci with the environment. Furthermore, in support of the linkage of *Phs* to the six markers located between *Brad690_SNP2 and 052_N17_ctg1,* all the lines with moderate sprouting

phenotype had the *Phs* resistance allele across this interval. This agrees with the moderate level of resistance to sprouting observed in these lines.

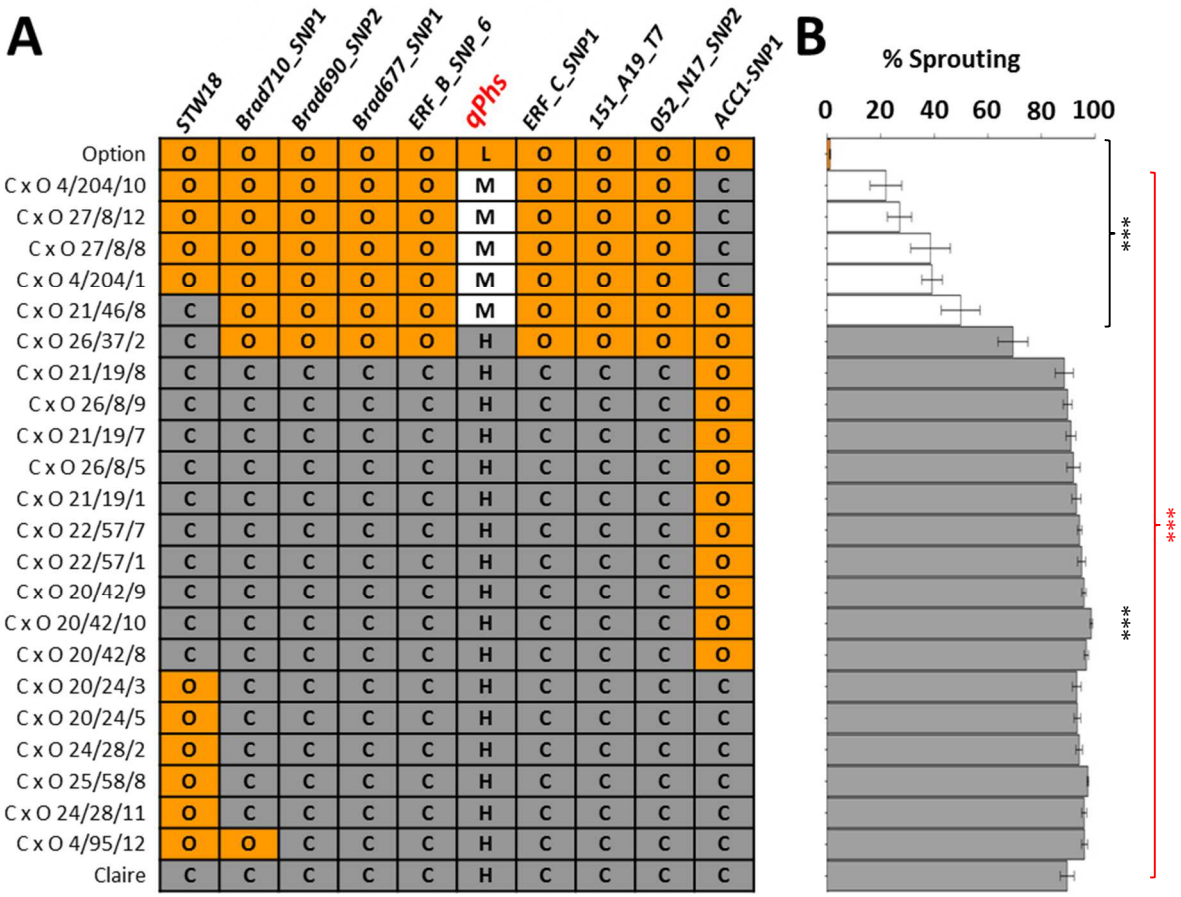


Figure 5.5: Fine-mapping of *Phs* in the OC population (experiment-2). Correlation of the marker genotypes of the selected OC recombinant lines (A) with the sprouting percentages observed in these lines (B). Claire and Option alleles are represented by C (grey) and O (amber), respectively. Designation of the *Phs* phenotype of these lines as low (L), moderate (M) or high (H) is based on the statistical comparative analysis presented in Table 5.2. The error bars in the panel B represents SEM of biological replicates with n indicated in Table 5.2. Adjusted significance value ($P < 0.001$) from the Dunnett comparison of Logit transformed sprouting percentage against Claire (black asterisk) and Option (red asterisk) are indicated.

Taken together, both fine-mapping experiments in the OC Population showed that the *Phs* phenotype is linked to the interval delimited by flanking markers *Brad710_SNP1* and *ACC1_SNP1*. Anchorage of these markers to the *Phs* physical map sequences revealed that there are 10 high confidence genes in this interval (Figure 5.6). The physical map sequences also suggest that the recombinations between markers *052_N17_SNP2* and *ACC1-SNP1* excludes *ACC Oxidase-1, Ubiquitin Conjugating Enzyme, Myosin-J protein, PM19-A2* and *PM19-A1* from the *Phs* effect.

This is especially important for the *PM19* genes, which were identified as major candidates for the *Phs* phenotype in the previous chapter of this study and in the study by Barrero et al. (2015). Evidence for the non-linkage of the *PM19* genes to the *Phs* phenotype was found in at least 12 independent recombinant lines across both experiments (excluding lines with moderate phenotype). This, therefore, suggests that the *PM19* genes are not the causal genes of the *Phs* effect in the UK mapping populations being studied.

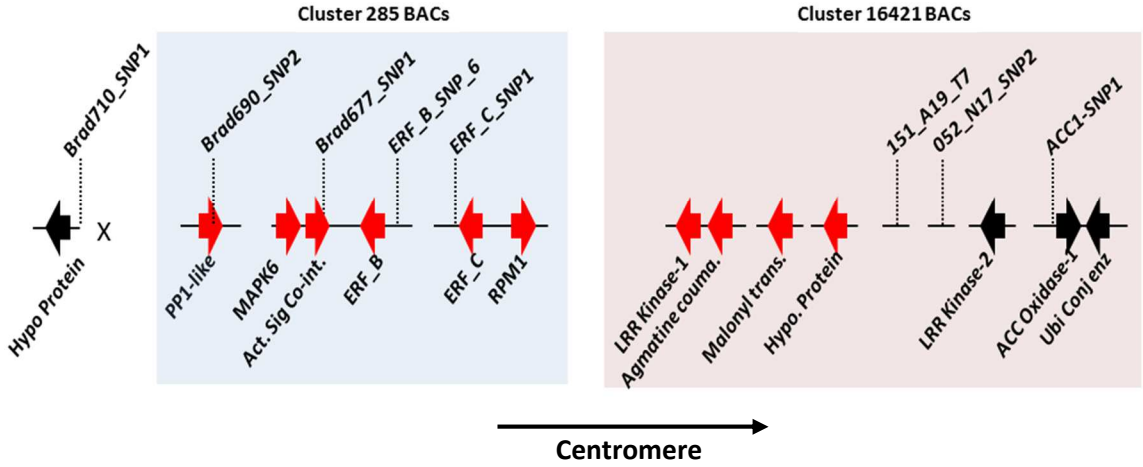


Figure 5.6: *Phs* is delimited to a minimum of 10 genes in the OC population. Schematic representation of the genes (filled arrow) found in the physical map and the locations of the genetic markers (dotted line) developed around them. The orientations of the genes are indicated by the direction of the filled arrow. The 10 genes linked to the *Phs* effect are filled in red. The chromosome orientation of the map is indicated by the horizontal arrow.

5.2.2.3 *Fine-mapping experiment-3 with the Alchemy x Robigus population*

To validate the results obtained from the OC Pop., an independent fine-mapping experiment was performed in the Alchemy x Robigus Population (AR Pop.). This AR Pop comprises of 79 independent recombinant lines selected from a recombination screen described in 5.2.2.4. The recombination events identified in these lines are predominantly between marker *STW17* and *Brad690_SNP2* and markers *052_N17* and *ACC1-SNP1*. Due to this redundancy of recombination haplotypes, only 22 of these recombinants, as well as Robigus and Alchemy, were phenotyped in this fine-mapping experiment.

The phenotyping results showed large contrasts in the sprouting phenotypes of these lines, with percentages of sprouting ranging from 3 % to 94 %. Similar to the OC Pop fine-mapping experiment-1 result, the distribution of the sprouting percentages was bimodal, suggesting a clustering of lines into two phenotypic groups. In one group was Alchemy along with five recombinant lines which showed lower sprouting percentages ranging from 3 % -18 %. The other group comprised of Robigus and the other 17 recombinant lines with high sprouting percentages ranging from 77 % to 94 %. By conducting an ANOVA with Dunnett Comparison analysis (Table 5.3), it was possible to designate lines as either low sprouting or high sprouting.

Furthermore, comparison of the phenotype of these recombinants with their genotypes (Figure 5.7) allowed a fine-mapping of *Phs* to the same marker interval as in the OC Population. This therefore further validates the previous fine-mapping experiment and also confirms that the same *Phs* locus is segregating in both the OC and AR Populations.

Table 5.3: Dunnett's multiple comparative analysis of the sprouting percentages obtained in fine-mapping experiment-3

Line	Plants per line (n)	% Sprouting	Sig (P*) against Alchemy	Sig (P*) against Robigus	*Phs* Score
Alchemy	3	18	NA	0	Low
RRA45/3/101/6/48	3	3	0.92	0	Low
RRA45/3/172/10/92	3	11	0.986	0	Low
RRA45/3/148/2/4	3	5	1	0	Low
RRA45/3/101/6/31	3	9	1	0.001	Low
RRA45/3/70/3/27	3	16	0.982	0.005	Low
RRA45/3/101/9/23	3	78	0.001	1	High
RRA45/3/101/9/73	3	90	0	1	High
RRA45/3/124/6/81	3	84	0	1	High
RRA45/3/158/1/82	3	92	0	0.992	High
RRA45/3/175/1/60	3	90	0	1	High
RRA45/3/63/1/22	3	86	0	1	High
RRA45/3/70/2/94	3	77	0.001	1	High
RRA45/3/72/3/54	3	77	0.001	1	High
RRA45/3/72/4/41	3	88	0	1	High
RRA45/3/158/2/67	3	93	0	0.984	High
RRA45/3/175/1/82	3	94	0	0.987	High
RRA45/3/149/9/55	3	90	0	0.87	High
RRA45/3/158/1/38	3	88	0	1	High
RRA45/3/175/1/50	3	89	0	1	High
RRA45/3/106/9/2	3	82	0	1	High
RRA45/3/124/6/24	3	81	0.001	1	High
RRA45/3/124/6/74	3	94	0	0.617	High
Robigus	3	84	0	NA	High

*P represents the adjusted P values obtained from the Dunnett multiple comparison analysis of Logit transformed sprouting percentage, using Robigus or Alchemy as control.

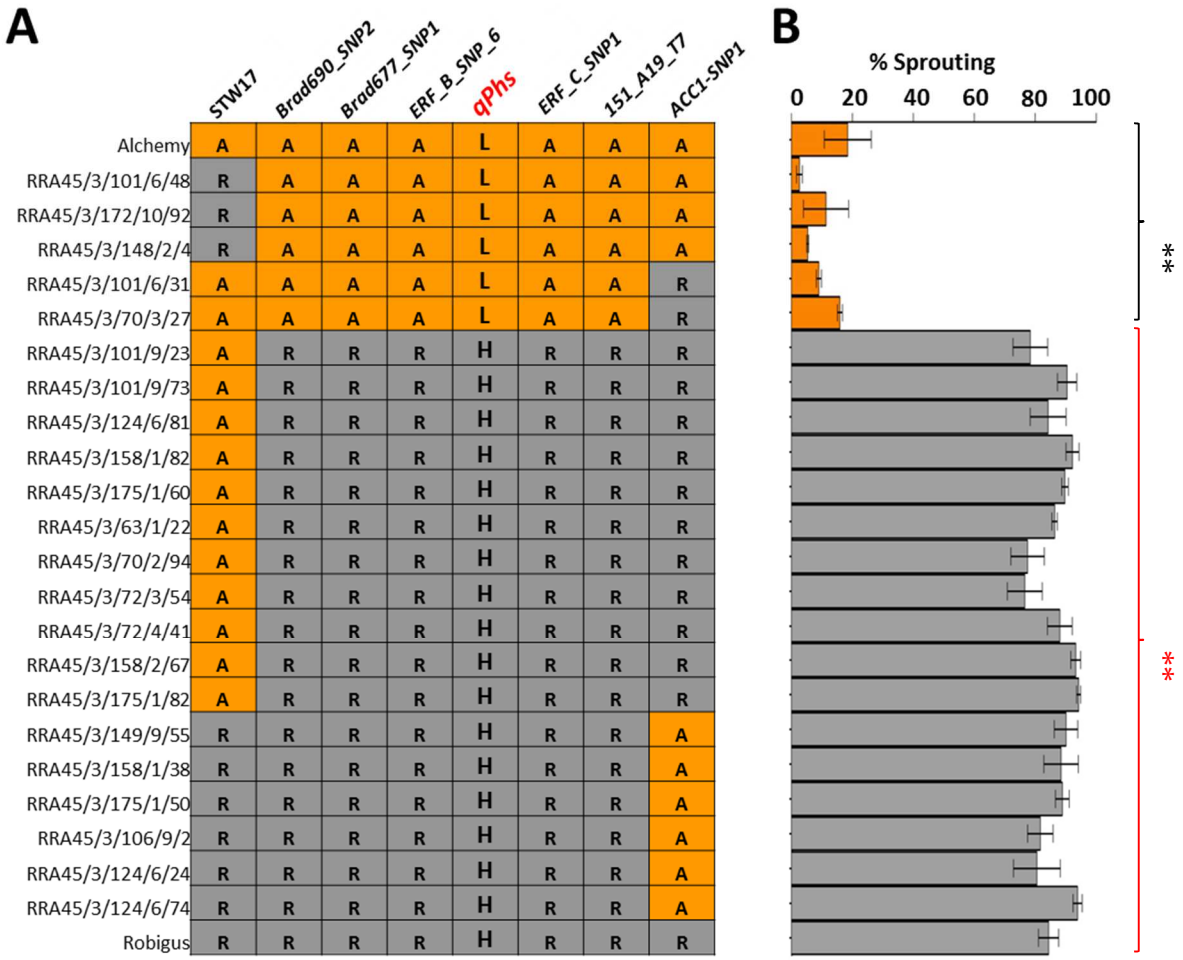


Figure 5.7: Fine-mapping of *Phs* in the AR Population (experiment 3). Correlation of the marker genotypes of the selected recombinant lines from the AR Population (A) with the sprouting percentages observed in these lines (B). Robigus and Alchemy alleles are represented by R (grey) and A (amber), respectively. Designation of the *Phs* phenotype of these lines as low (L) or high (H) is based on the statistical comparative analysis presented in Table 5.3. The error bar in panel B represents SEM of biological replicates with n = 3. Adjusted significance value ($P < 0.01$) from the Dunnett comparison of Logit transformed sprouting percentage against Robigus (black asterisk) and Alchemy (red asterisk) are indicated

5.2.2.4 *Evidence of suppressed recombination within the Phs interval in the AR and OC Populations*

The linkage of the sprouting phenotype in the OC and AR Populations to at least 10 genes on the physical map meant that *Phs* could not be delimited to a single gene locus in these populations. In order to break the linkage between these genes, we screened F_2 progeny for new recombination events. For OC Pop., F_2 progeny of F_1 plants obtained from a cross between Claire and one of the original OC Pop. RIL (C x O 26/48/4) were used, while the F_2 progeny of the original BC_3F_1 generation were screened for the AR Pop. For the recombination screen in the AR Pop., 2,590 F_2 plants were screened while 2,770 plants were screened in the OC Pop. Together, a total of 10,720 gametes were screened across both populations. Despite this large number of gamete screened, we could not detect any recombination event between any of the linked genes in the *Phs* region. This represents a minimum of 0.7 Mb of sequences. All the recombination events detected in these populations occurred around the flanking markers but not in the critical *Phs* interval. This, therefore, suggests that there is a suppression of recombination in the critical *Phs* interval.

5.2.3 High-resolution fine-mapping in multi-parental breeding lines further delimits *Phs*

5.2.3.1 *Recombination screen identifies historic recombination events in diverse breeding population*

To overcome the limitation posed by the lack of recombination events between the genes linked to *Phs*, we collaborated with our breeding partners from Limagrain UK to screen for recombination events between these genes in their diverse breeding population. This breeding population was composed of 6,500 F_5 lines originally developed from independent bi-parental or multi-parental crosses involving many different UK wheat varieties. Due to the high number of lines involves (6,500 lines, 13,000 gametes), an initial screen of the population was first done using five SNP markers: *ACC1_SNP1, ERF_B_SNP6, Brad677_SNP1, Brad690_SNP2 and STW17*. This

initial screen identified 29 lines with possible recombination events in the *Phs* interval. These are afterwards referred to as Limagrain recombinants.

The 29 lines identified from this initial screen have very diverse pedigree. They were developed from 20 independent bi-parental or tri-parental crosses involving about 31 UK wheat varieties. Subsequently genotyping of these lines with additional markers (*ERF_C_SNP1, 151_A19* and *052_N17_SNP2*) showed a series of recombination events in the target *Phs* interval and this gave rise to 10 different haplotype groups (Figure 5.8). With these recombination events, it was possible to break the linkage between many of the genes in the critical *Phs* region especially between those on the BAC Cluster285. Furthermore, subsequent genotyping of the parental varieties from which these selected recombinant lines were developed showed that some of them also contain the same recombination detected in the breeding lines. This implies that most of the recombination events detected in the recombinants are historic and did not arise from recent meiosis events.

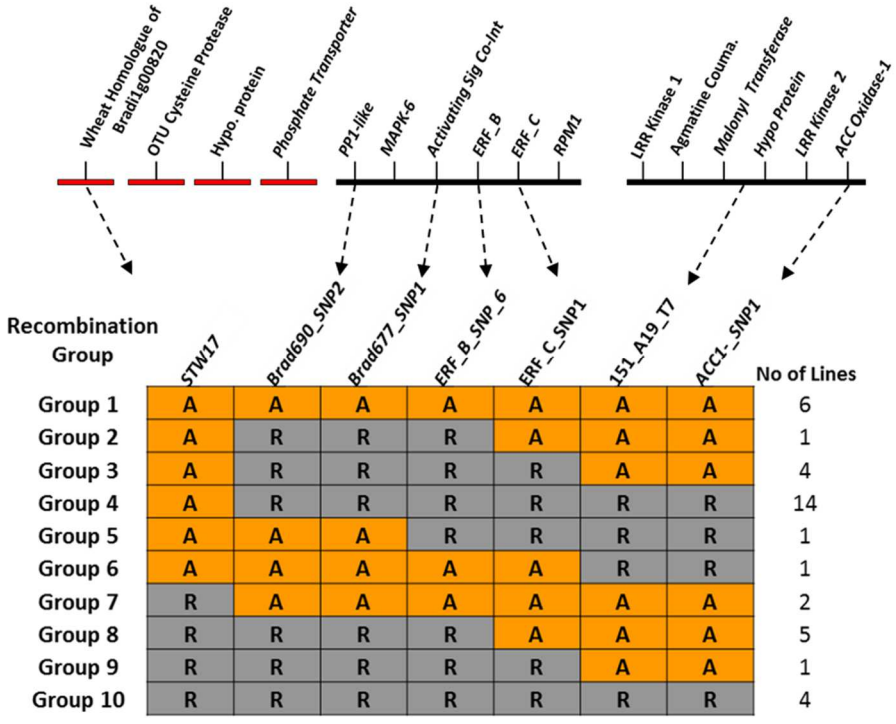


Figure 5.8: Recombination haplotypes in the Limagrain population. The different recombination events present in the Limagrain population were defined using genetic markers developed across the *Phs* interval. Since the recombinants are derived from diverse parents, the susceptible Robigus (R) and resistant Alchemy (A) alleles are used to represent the genotype of the recombinants. The position of the markers in the physical map is indicated by the connecting arrows. Eight different recombination haplotypes were identified in this population. The number of lines per recombination haplotype is indicated in the rightmost column.

5.2.3.2 *Fine-mapping experiment-4 with the Limagrain Recombinants*

To assess the sprouting phenotype of these recombinants, we harvested and phenotyped spikes from plants of the selected breeding lines. These plants were grown in the field site of Limagrain at Thriplow, UK. In addition to these recombinants, lines with homozygous resistant or susceptible *Phs* alleles across the target interval were also phenotyped as controls.

A larger variation in sprouting percentage was observed among the selected recombinant lines and control, and this ranged from 4 % to 96 %. Despite this variation, frequency distribution of the sprouting scores obtained from these lines showed two distinct groups (Figure 5.9). The first group had lines with varying levels of sprouting resistance (4 - 47 %) while the other group had lines with higher level of sprouting (63 - 96 %). This is indicative of a *Phs* effect despite the very diverse pedigree and heterogeneous genetic background of these lines.

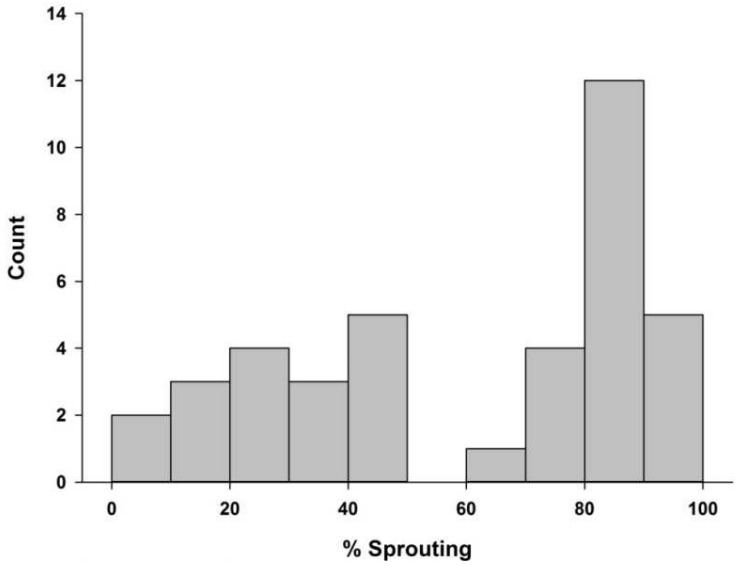


Figure 5.9: Bimodal distribution of the sprouting phenotype of Limagrain recombinants. The frequency distribution of the sprouting percentage obtained from selected recombinants and control lines from the Limagrain breeding population

For a more accurate and statistically robust assignment of *Phs* sprouting phenotype to each line, we employed the Tukey's Honestly Significant Difference(HSD) test (Tukey, 1949) which allows for pairwise comparisons of treatments in the absence of appropriate controls. Using this test on Logit transformed sprouting percentage data allowed us to designate lines as either low or high sprouting (Table 5.4). Lines in the same significant group as the least sprouting lines were designated as low sprouting. Similarly, lines were designated as high sprouting if they were in the same significance group with the highest sprouting line. Lines that could not be assigned to either group based on this Tukey grouping information were assigned a moderate sprouting phenotype.

Comparing the genotypic and phenotypic score for these breeding lines show that *Phs* is linked to *ERF_C_SNP1* except in two anomalous lines - 3318 and 5196 (Figure 5.10). This linkage of *ERF_SNP_1* to *Phs* represents the highest correlation of *Phs* to any of the markers on the genetic map. In at least six independent recombination lines, *Phs* was found not to be linked with the *ERF_B_SNP6* marker, neither was it linked to *Brad677_SNP1* nor the *Brad690_SNP2* markers. Similarly, the phenotype of six different recombinant lines suggested that *Phs* is not linked to the *151_A19_T7* marker. These, therefore, delimit the *Phs* effect to the interval between *ERF_B_SNP6 and 151_A17*.

The sprouting phenotype of lines 3318 and 5196 could not be correlated with any of the markers on the genetic map. Although the cause of the aberrant phenotype of these two lines is not yet clear, it could however be caused by effects arising from loci segregating in the heterogeneous genetic background of this diverse lines. In addition, differences in the days to flowering time of these field-grown recombinant lines could also contribute to the anomalous phenotypes. Such flowering time difference would not have been detected at harvest ripeness when spikes were harvested and might consequently have an effect on the onset of after-ripening in these lines.

Table 5.4: Tukey multiple comparative analyses of the sprouting percentages obtained in fine-mapping experiment-4

Line	Sprouting (%)	Grouping Information Using the Tukey method of pairwise comparison and 95% Confidence*												*Phs* Score
9130	5												L	Low
3321	10											K	L	Low
7507	11										J	K	L	Low
1414	12										J	K	L	Low
1450	18										J	K	L	Low
8190	21										J	K	L	Low
7169	24									I	J	K	L	Low
8062	23								H	I	J	K	L	Low
3384	25									I	J	K	L	Low
8439	32							G	H	I	J	K	L	Low
5009	33						F	G	H	I	J	K	L	Low
5196	38					E	F	G	H	I	J	K	L	Low
4160	41					E	F	G	H	I	J	K	L	Low
3387	40					E	F	G	H	I	J	K		Moderate
3588	42				D	E	F	G	H	I	J	K		Moderate
7398	42				D	E	F	G	H	I	J	K		Moderate
2004	47			C	D	E	F	G	H	I	J			Moderate
3318	64		B	C	D	E	F	G	H	I				Moderate
7510	74	A	B	C	D	E	F	G	H					High
7233	73	A	B	C	D	E	F	G						High
6007	76	A	B	C	D	E	F	G						High
8395	79	A	B	C	D	E	F							High
3291	80	A	B	C	D	E	F							High
7200	84	A	B	C	D	E	F							High
8394	82	A	B	C	D	E								High
3306	83	A	B	C	D	E								High
3101	83	A	B	C	D	E								High
3309	84	A	B	C	D	E								High
8534	84	A	B	C	D	E								High
8269	86	A	B	C	D	E								High
3305	87	A	B	C	D	E								High
8538	87	A	B	C	D	E								High
8247	88	A	B	C	D									High
10259	89	A	B	C	D									High
8028	90	A	B	C										High
9301	90	A	B											High
8240	92	A	B											High
8574	95	A												High
7393	95	A												High

* Lines that do not share a letter have significantly different Logit transformed percentages

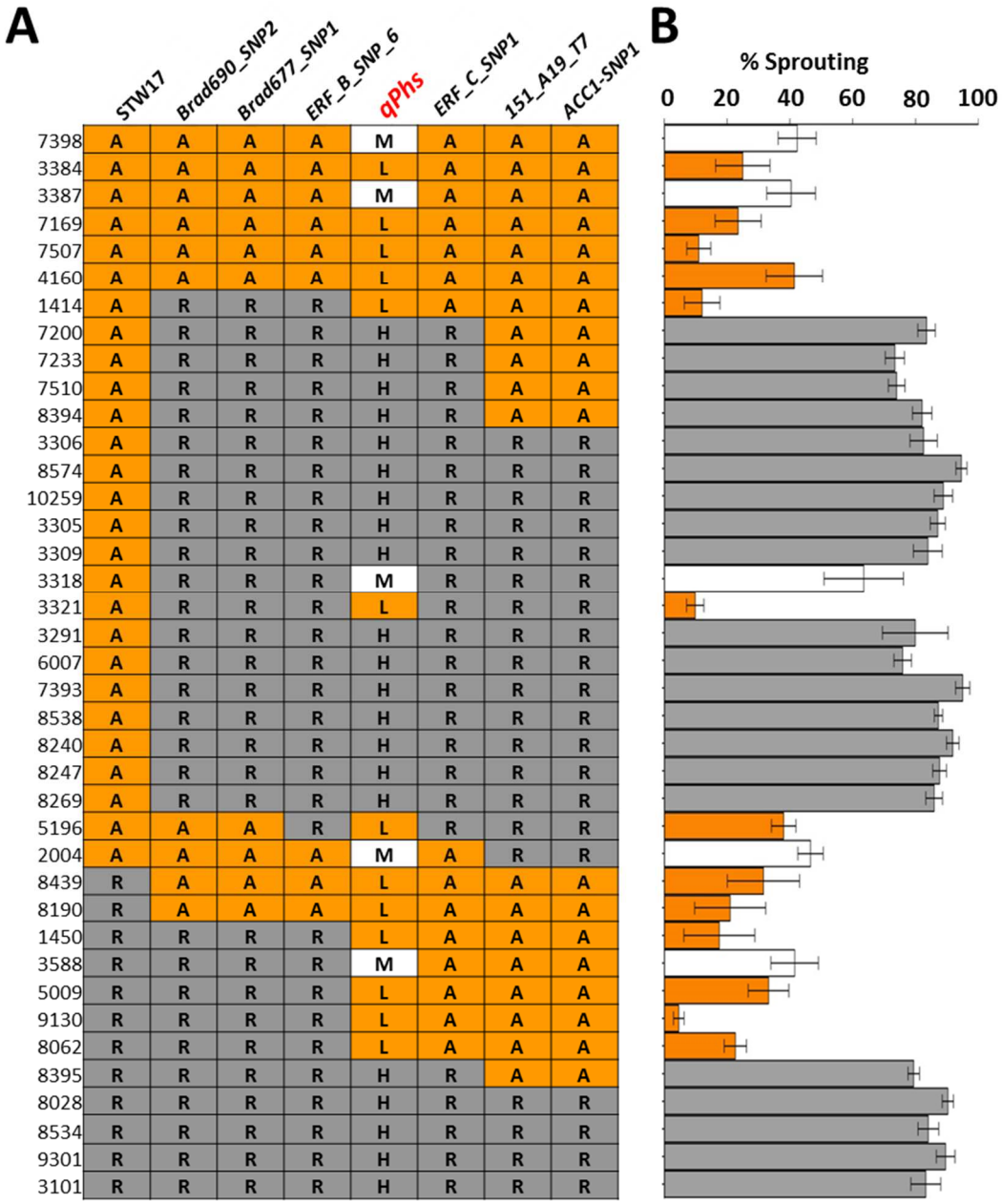


Figure 5.10: Fine-mapping of *Phs* in the Limagrain population (experiment 4). Correlation of the marker genotypes of selected recombinant lines from the Limagrain population (A) with the sprouting percentages observed in these lines (B). The alleles of the recombinants at each marker locus are presented either as Robigus-type (R, grey) or Alchemy-type (A, amber). Designation of the *Phs* phenotype of these lines as low (L), Moderate (M) or high (H) is based on the statistical comparative analysis presented in Table 5.4. The error bars in the panel B represents SEM of biological replicates with n >= 4.

5.2.3.3 *Fine-mapping experiment-5 with Limagrain recombinants*

To validate the result obtained from the previous Limagrain recombinant fine-mapping experiment, a subset of these recombinant lines (including those with critical recombination events around the *Phs* locus) were grown in the glasshouse for further phenotyping. In addition, Robigus, Alchemy, Claire and Option - the four contrasting *Phs* parents of the OC and AR Population - were also included in this experiment. So as to adjust for differences in the flowering time of these lines, spikes were tagged at anthesis and were harvested at approximately equal days post anthesis. Harvested spikes were allowed to after-ripen for 2 - 3 weeks and were subsequently phenotyped in the sprouting chamber with an artificial misting system.

Similar to the results of the previous Limagrain recombinant fine-mapping experiment, a large variation in sprouting phenotype was observed in this experiment. The percentage of sprouting in this lines ranged from the 2 % to 95 %. However, unlike the result of the previous experiment, the distribution of the sprouting percentages was not bimodal. This made it difficult to delineate (with high confidence) lines with low sprouting phenotypes from those with high sprouting phenotypes.

To overcome this, we used the Dunnett's comparison test to compare the Logit transformed sprouting percentages of the individual lines with both Robigus and Alchemy (Table 5.5). Lines were therefore assigned as low sprouting if they show statistically lower percentages to Robigus but were not significantly different to Alchemy. Conversely, lines designated as low sprouting have statistically higher percentages than Alchemy but not significantly different percentages to Robigus. However line 3387 showed an in-between phenotype which is not statistically different to both Robigus and Alchemy.

Table 5.5: Dunnett's multiple comparative analyses of the sprouting percentages obtained in fine-mapping experiment-5

Line	Plants per line (n)	% Sprouting	Sig (P*) against Alchemy	Sig (P*) against Robigus	*Phs* Score
Option	6	2	0	0	Low
9130	6	8	0.005	0	Low*
8439	6	18	1	0	Low
Alchemy	6	24	NA	0	Low
1450	6	24	1	0	Low
5009	6	25	1	0	Low
1414	6	27	1	0	Low
3588	6	31	1	0	Low
3384	6	39	0.763	0.001	Low
4160	6	39	0.978	0	Low
8190	6	43	0.809	0.001	Low
7398	6	56	0.167	0.023	Low
8062	6	59	0.112	0.038	Low
3387	5	61	0.077	0.115	Moderate^{+}
2004	6	71	0	0.903	High
5196	6	77	0.001	0.205	High
7233	6	81	0	0.999	High
Robigus	6	83	0	NA	High
10259	6	86	0	1	High
8395	6	86	0	1	High
7510	6	88	0	1	High
3291	6	88	0	1	High
8028	6	90	0	1	High
3318	6	92	0	0.774	High
3101	6	92	0	1	High
Claire	6	94	0	0.031	High^{+}
7200	6	95	0	0.538	High
8534	6	95	0	0.032	High^{+}

*P represents the adjusted P values obtained from the Dunnett multiple comparison analysis of Logit transformed sprouting percentage, using Robigus or Alchemy as control.

$^{+}$Lines with statistically different sprouting values compared to the both Robigus and Alchemy are described as having a moderate sprouting phenotype. The exception to this were lines 9130 and Option which had lower sprouting score than Alchemy and line 8534 and Claire with higher sprouting score than Robigus.

As was observed in the previous Limagrain Recombinant fine-mapping experiment-4, *Phs* was found to be linked with the *ERF_C_SNP1* marker in all but one of the recombinant lines (Figure 5.11). The exception to this was line 2004 that showed a high sprouting phenotype instead of the low sprouting phenotype suggested by its *ERF_C_SNP1* marker genotype. Although we do not know the reason for this, we hypothesise that this line has other background genetic loci that are masking the expression of *Phs* effect. In support of this hypothesis, line 2004 showed a moderate level of sprouting in the previous fine-mapping experiment 4 (Figure 5.10, B) and has the largest variance of all the high sprouting lines in this experiment. Furthermore, line 5196 which did not support *Phs* linkage to *ERF_C_SNP1* in the previous experiment showed a soundly high phenotype in this experiment and this agrees with *Phs*-*ERF_C_SNP1* linkage.

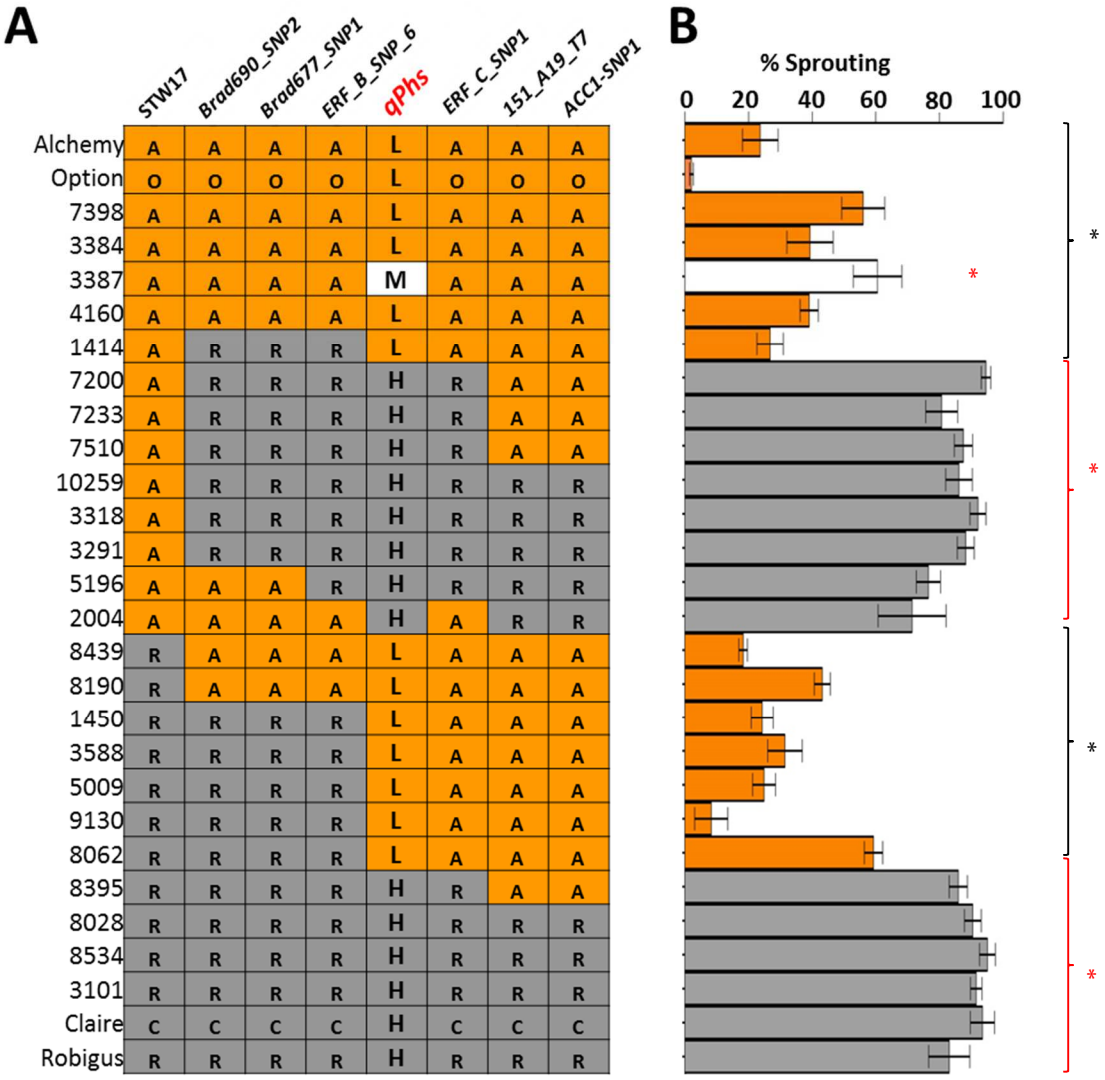


Figure 5.11: Fine-mapping of *Phs* in the Limagrain population (Experiment 5). Correlation of the marker genotypes of selected recombinant lines from the Limagrain population (A) with their sprouting percentages (B). At each marker locus, R and C represents the susceptible (grey) Robigus and Claire parental alleles, respectively; while A and O represent the resistant (amber) Alchemy and Option alleles. The genotypes of the recombinants are presented using the Robigus and Alchemy alleles designation. Designation of the *Phs* phenotype of these lines as low (L), moderate (M) or high (H) is based on the statistical comparative analysis presented in Table 5.5. The error bars in the panel B represents SEM of biological replication with n >= 5. Adjusted significance value ($P < 0.05$) from the Dunnett comparison of Logit transformed sprouting percentage against Robigus (black asterisk) and Alchemy (red asterisk) are indicated.

5.2.3.4 *Summary of the fine-mapping experiments with Limagrain recombinants*

The data from both fine-mapping experiments with the Limagrain population were combined together to enable a more robust scoring of the *Phs* phenotypes of the recombinant (Figure 5.12, B). The Dunnett's Comparison test was once again used to assign *Phs* phenotypes to these lines relative to the phenotypes of Robigus and Alchemy. The result of this combined analysis further confirms that *Phs* is linked to the *ERF_C_SNP1* marker.

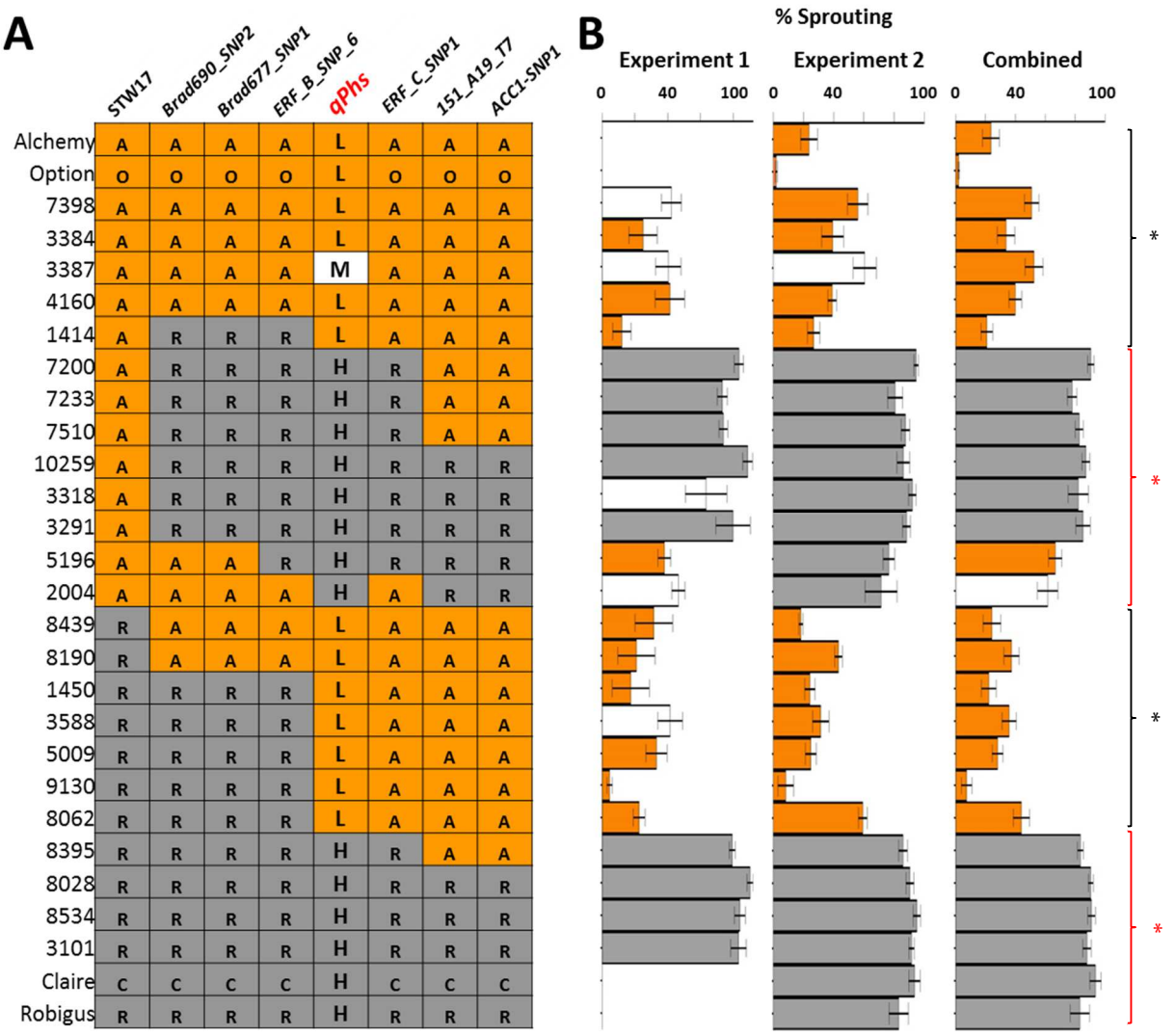


Figure 5.12: Summary of the fine-mapping experiments in the Limagrain population. Correlation of the marker genotypes of selected recombinant lines from the Limagrain population (A) with the individual and combined sprouting percentages observed in experiment 1 and 2 (B). At each marker locus, R and C represents the susceptible (grey) Robigus and Claire parental alleles, respectively; while A and O represent the resistant (amber) Alchemy and Option alleles. The genotypes of the recombinants are presented using the Robigus and Alchemy alleles designation. Designation of the *Phs* phenotype of these lines as low (L), Moderate (M) or high (H) is based on the statistical comparative analysis with the combined sprouting data. The error bars in panel B represents SEM of biological replicates with n >= 4. Adjusted significance value ($P < 0.05$) from the Dunnett comparison of Logit transformed sprouting percentage of combined data against Robigus (black asterisk) and Alchemy (red asterisk) are indicated

The only exception to this *Phs-ERF_C_SNP_1* linkage in this combined data was line 5196 which was assigned a low sprouting phenotype (based on the combined experiment data) as opposed to a high sprouting phenotype suggested by it *ERF_C_SNP1* genotype. Line 5196 phenotype varied under the different environmental conditions of the two phenotyping experiments (field and glass-house conditions). It is, therefore, possible that the expression of the *Phs* effect in this line is being masked by its environmental sensitivity, and this could originate from loci other than the target 4AL interval. To reduce such masking effect, line 5196 has been backcrossed to both Claire and Option so as to create Near Isogenic lines (NILs) which will retain the unique recombination event present in this line but will have genetic background similar to either Claire or Option. The F_1 progeny of these crosses are currently being grown for subsequent screening in the F_2 generation.

In summary, the two fine-mapping experiments with the Limagrain recombinants has delimited *Phs* to the *ERF_C_SNP1* marker locus. Through the recombination events between *ERF_B_SNP6* and the critical *ERF_C_SNP1* marker, we can dissociate the *Phs* phenotype from all the markers proximal to and including *ERF_B-SNP6*. Considering the position of the *ERF_B _SNP6* marker on the physical map sequence (Figure 5.13), this mean that four genes (*ERF_B, Activating Signal Cointegrator, MAPK-6* and *PP1-Like*), which previously were linked to *Phs* in the CO and AR Pops, can now be excluded from the *Phs* effect. In addition, the linkage of the *151_A19_T7* with *Phs* was broken in these fine-mapping experiments. This implies that all the genes distal to this marker loci on the physical map can also be excluded from the *Phs* effect. This include *LRR kinase 2, ACC Oxidase1, Ubiquitin Conjugating Enzyme, Myosin J protein* and the *PM19* genes. Importantly, the mapping results in this population further confirms that the *PM19* genes are not the cause of the *Phs* effect in UK material.

Consequently, the *Phs* effect can now be narrowed down to six genes located between the *ERF_B_SNP6* and *151_A19_T7*. This includes *ERF_C, RPM1, LRR kinase-1, Agmatine Coumarorl, Malonyl transferase* and the *Hypothetical Protein*. However, this depends on the position of the recombination events between the *ERF_C _SNP1* and *151_A19_T7* markers. It is also important to note that there is still a gap in the physical

map of *Phs* in this interval, and this might contain more genes that could underly the *Phs* effect.

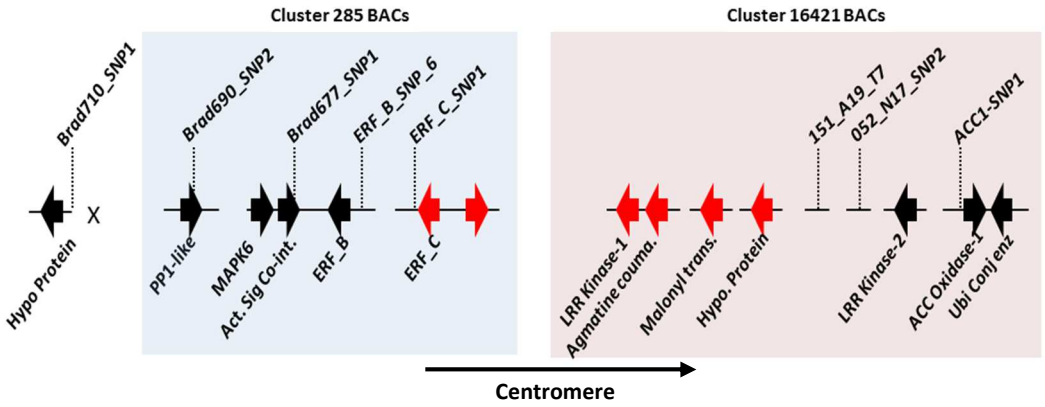


Figure 5.13: *Phs* is delimited to six genes in the Limagrain pop. Schematic representation of the genes (filled arrow) found in the physical map and the locations of the genetic marker (dotted line) developed around them. The orientations of the genes are indicated by the direction of the filled arrow. The six genes linked to the *Phs* phenotype are filled in red. The chromosome orientation of the map is indicated by the horizontal arrow.

5.2.4 High-resolution genetic mapping of the critical *Phs* interval reveals discordance in the genetic and physical map position of *ERF_C* markers

To further define the critical recombination events associated with the *Phs* effect (especially between *ERF_C_SNP1* and *151_A19_T7*), the Limagrain recombinants phenotyped in the experiment-5 were further genotyped with additional genetic markers. These include *ERF_C_SNP3* located in the exon of the *ERF_C* gene; *Sc4326385_SNP1* located in the 3' untranslated region of *RPM1*; *LRR-1_SNP1* located in the exon of *LRR Kinase-1* , *Agmatine_SNP1* marker located in the promoter region of the *Agmatine Coumarorl* gene and *ACC-2_SNP1* which is located on the exon of the *Hypotethical Protein* (similar to *ACC Oxidase-1 like*) (Figure 5.14).

The genetic linkage map obtained from this genotyping showed a rather surprising discordance between the position of the two *ERF_C* markers (*ERF_C_SNP1* and *ERF_C_SNP3*) on the genetic map (Figure 5.14). This is because *ERF_C _SNP1* genetically mapped between *LRR_SNP1* and *Agmatine SNP1* while *ERF_C_SNP3* was

linked to *LRR-SNP1 , Sc436385_SNP1* and *ERF_B_SNP6* in all the recombinants examined (Figure 5.14). The map position was verified in 12 independent plants per recombinant line confirming the robustness of the genotypic score.

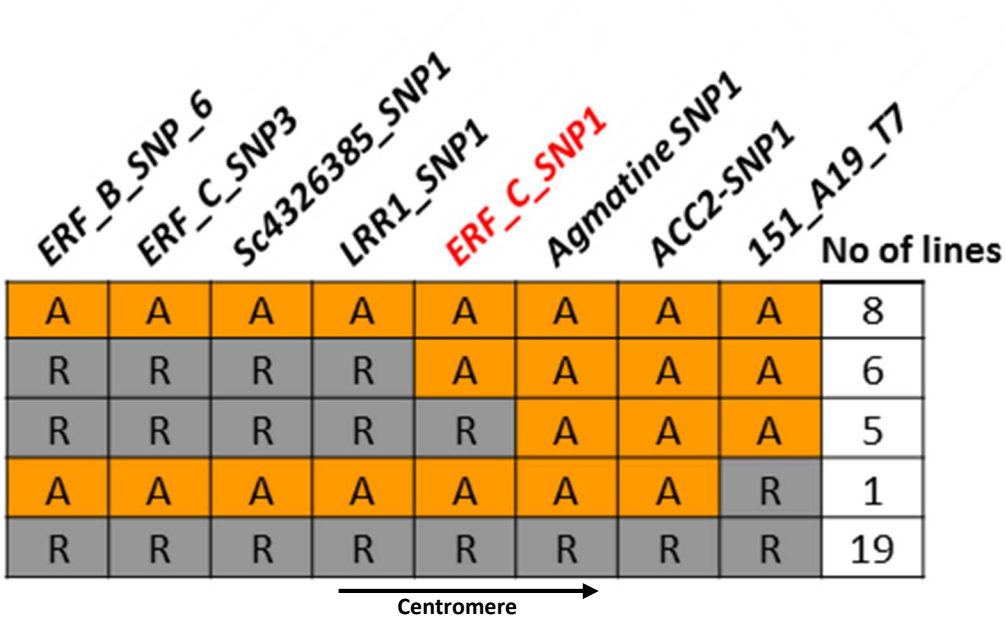


Figure 5.14: High-resolution genetic linkage map of the Limagrain recombinant population. The different recombination events present in the Limagrain recombinant lines were further defined using more genetic markers developed across the *Phs* interval. This identified discordance in the genetic position of the *ERF_C_SNP1* marker (highlighted in red). The susceptible Robigus (R) and resistant Alchemy (A) alleles are used to represent the genotypes of the recombinants. The number of lines having the different recombination haplotypes is indicated in the rightmost column

To resolve this discrepancy, the recombinants were further genotyped with more genetic markers designed from the sequences around *ERF_C* and *LRR Kinase 1*. These include *ERF_C_SNP4* and *ERF_C_SNP*5 markers, both located in the promoter region of *ERF_C*; and *LRR1_SNP4* also located in the promoter region of *LRR kinase-1*. The genetic positions of these new markers agree with the positions of *ERF_C_SNP3* and *LRR1-SNP1* previously used and further highlights the discrepancy in the genetic and physical map position of *ERF_C_SNP1* (Figure 5.15).

The positions of *ERF_C_SNP3, 4 and 5* on the genetic map are supported by the location of *ERF_C* in the physical map. However, this is not true for the *ERF_C_SNP1*

which was also designed to mark the same gene. Possible reasons for this discordance could be that the primer sequences for the *ERF_C_SNP1* marker are not gene specific and are therefore amplifying random non-target sequences elsewhere in the genome.

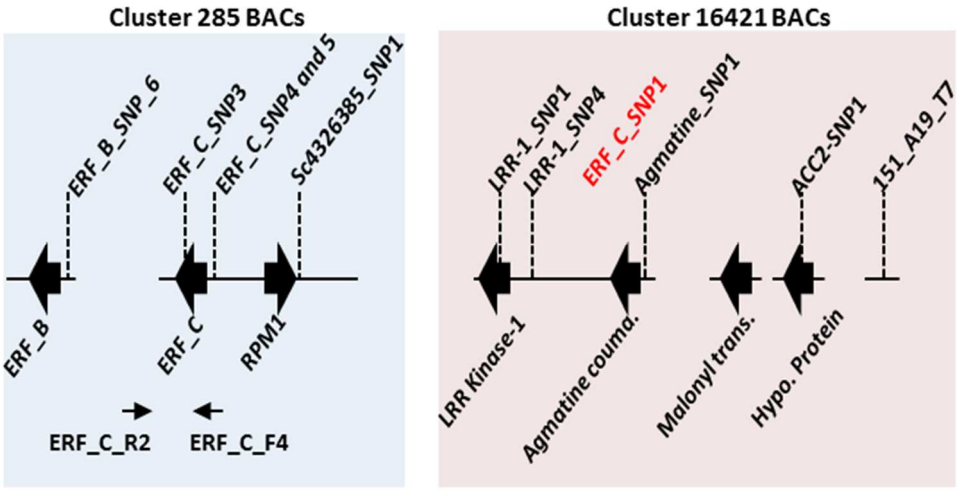


Figure 5.15: High-resolution genetic mapping highlights discordant positions of the *ERF_C* markers. Filled arrows represent the genes located in the fine-mapped interval in the Limagrain recombinant lines. The dotted lines represent the genetic marker around each gene locus and are arranged according to the order of the linkage map in Figure 5.14 (filled arrow). The discordant *ERF_C_SNP1* marker is highlighted in red. Unfilled arrow indicates the primer pair used for the initial PCR amplification in the nested genotyping assay.

To test this hypothesis, we repeated the genotyping of the recombinant lines using a nested PCR assay. Nested PCR assays are known to improve the specificity of amplification of target sequences (Soleimani et al., 2003). The assay employed involved an initial PCR amplification of a 1127 bp sequence of *ERF_C* in the recombinant lines using primers ERF_C_F4 and R2 (Figure 14). The amplified sequence encompasses the sequences from which both the *ERF_C_SNP1* and *ERF_C_SNP3* markers were designed. The amplicon obtained from this PCR was then subsequently used for genotyping the recombinants with the SNP markers. The result obtained from the nested PCR genotyping assay was not different from the score obtained from the conventional genotyping approach. This further validates the genetic position of these markers and suggest that primers for the *ERF_C_SNP1* marker are not just amplifying spurious sequences in the genome.

Given the certainty of the genetic map position of *ERF_C_SNP1*, the discordance in its genetic-physical map position could alternatively arise from a potential duplication of *ERF_C*, with the other copy located between *LRR kinase-1* and *Agmatine Coumarorl*. Such a duplication event will agree with the genetic position of *ERF_C_SNP1* being between *LRR1_SNP4* and *Agmatine_SNP1*. We, however, could not find any physical sequence corresponding to the *ERF_C_SNP1* marker sequences between these two genes.

Although speculative, the lack of a physical sequence evidence for this duplication, could be due to the absence of the duplication (if at all existent) in the Chinese Spring cultivar from which the physical map was developed. Alternatively, Chinese Spring could also have also undergone a deletion event in this interval, and this would have included the sequences of *ERF_C_SNP1*. However, more experimental work needs to be done to characterise further this interval.

5.2.5 *Phs* maps to a 10 kb interval on the Chinese Spring physical map

The previous fine-mapping information obtained from the Limagrain recombinant lines suggests a tight linkage between *Phs* and *ERF_C_SNP1*. This, along with the high-resolution genetic map information obtained from the same population (Figure 5.16), shows that this *Phs-ERF_C_SNP1* linkage is delimited to the physical interval between *LRR1_SNP4* and *Agmatine_SNP1* (Figure 5.14). Interestingly, the SNP loci for the *LRR1_SNP4* and *Agmatine_SNP1* markers are 9,822 bp apart on the same contig in three independent Chinese Spring BAC clones sequences (TaaCsp4AL031F05, TaaCsp4AL034K23, TaaCsp4AL121J20; see Chapter Four, section 4.2.6.2.1). This interval includes one of the exons of *Agmatine Coumaroyl Transferase* (hereafter referred to as *ACT*) but does not include the coding sequence of *LRR kinase-1* (Figure 5.15).

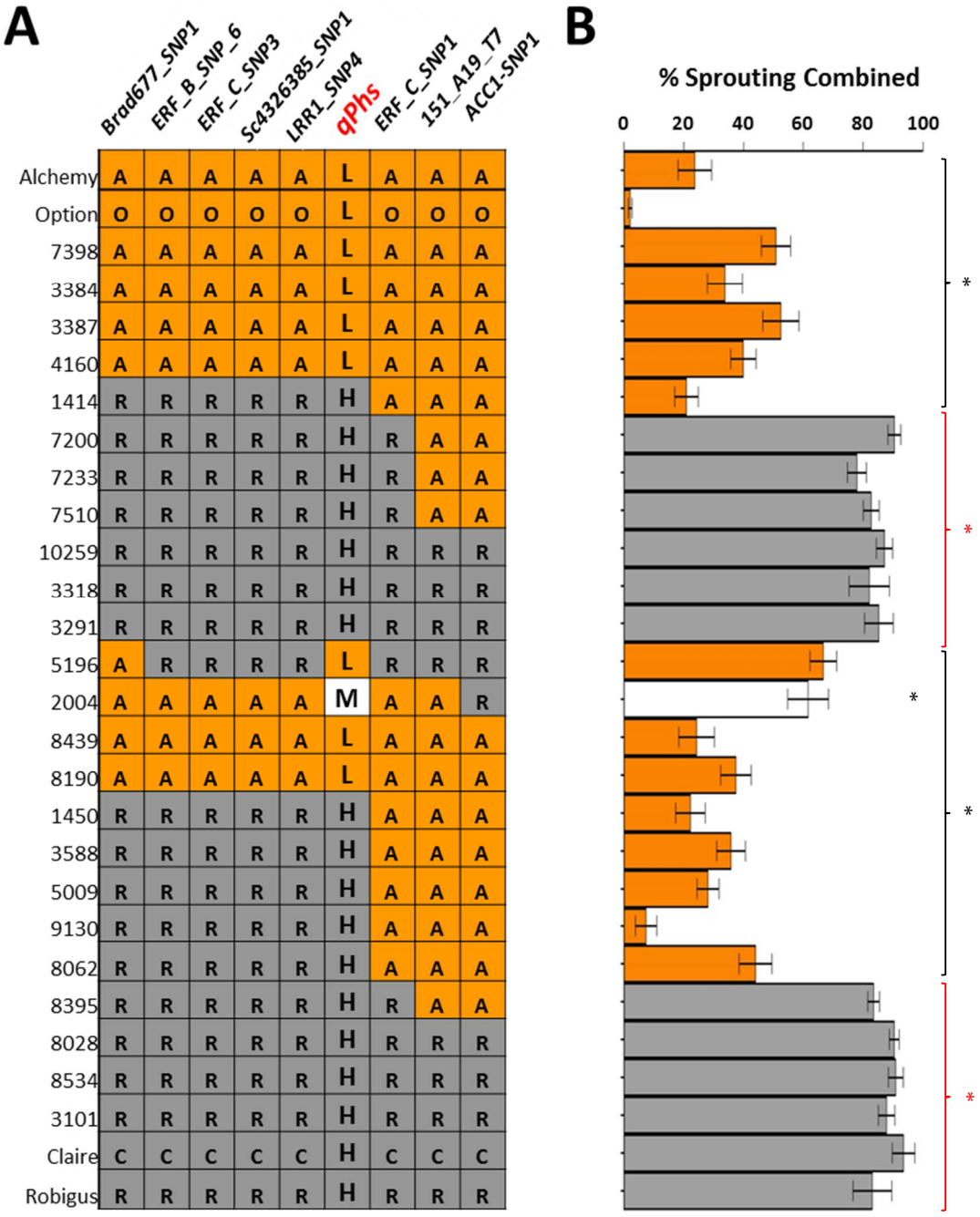


Figure 5.16: Summary of fine-mapping of *Phs* in the Limagrain recombinant population. Correlation of the marker genotypes of selected recombinant lines from the Limagrain population (A) with the combined sprouting percentages observed in experiment 4 and 5 (B). At each marker locus, R and C represents the susceptible (grey) Robigus and Claire parental alleles, respectively; while A and O represent the resistant (amber) Alchemy and Option alleles. The genotypes of the recombinants are presented using the Robigus and Alchemy alleles designation. Designation of the *Phs* phenotype of these lines as low (L), moderate (M) or high (H) is based on the statistical comparative analysis with the combined sprouting data. The error bars in the panel B represents SEM of biological replicates with n >= 4. Adjusted significance value ($P < 0.05$) from the Dunnett comparison of Logit transformed sprouting percentage of combined data against Robigus (black asterisk) and Alchemy (red asterisk) are indicated

To further characterise this genomic interval, we undertook a comparative sequence analysis of the Chinese Spring 4AL physical map sequence with other assemblies of hexaploid and progenitor wheat species. These included the assemblies of W7984 synthetic hexaploid wheat (Chapman et al., 2015), as well as those of diploid *Triticum urartu* and *Aegilops tauschii* assemblies (Ling et al., 2013, Akhunov et al., 2005, Slageren, 1994).

A BLAST analysis using the critical genomic interval sequences against the W7984 assembly on the CerealsDB website (http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/DOC_BLAST.php) returned a 22.5 kb scaffold (Scaffold173511) with 98 % sequence identity. A similar search against the *T. urartu* database on the Ensembl Plant database returned a 12.7 kb scaffold (scaffold70788) with 99 % sequence identity. Corresponding scaffold sequence for *Aegilops tauschii* was obtained from a BLAST analysis on NCBI. This returned a 256 Kb whole genome shotgun assembly scaffold (Scaffold14338, Genbank no: KD513537.1). Interestingly, this *Ae. tauschii* scaffold also contains truncated sequences for *ERF_C* and *RPM1* as well as sequences for *Malonyl Co Enzyme-A, Ubiquitin Conjugating enzyme, Myosin –J protein, PM19-A2* and *PM19-A1*. This suggests that the sequence of Scaffold14338 is truly homologous to the target *Phs* interval sequences. As is typical of many scaffolds made from whole genome shotgun assembly, the *Ae. tauschii* scaffold contains some gaps and ambiguous sequences.

The *T. urartu* scaffold contains a 1,995 bp gene model for *ACT*, and this consists of four exons with the last two exons being the smallest (112 and 114 bp; Figure 5.17). The scaffold also contains the *LRR kinase-1* genes with two exons. Interestingly, there is an overlap between the third exon of *ACT* and the second exon of *LRR Kinase-1* suggesting the occurrence of gene overprinting (Rancurel et al., 2009). While most of these genic sequences are also represented in the *Ae. tauschii* scaffold, we could not find the second exon of ACT in this assembly. Both hexaploid wheat assemblies (Chinese Spring and W7984) show high genic and intergenic sequences similarity except for the presence of a few polymorphic indel in the intergenic region. Also the, 5' sequence containing the transcription start site of *LRR Kinase-1* was found to be deleted in these assemblies. This is not evident in the *T. urartu* or *Ae. tauschii* scaffolds neither is it

evident in the Chinese Spring 5DL contig (5DL_4501029) which is homoeologous to the Chinese Spring 4AL BAC sequence.

Examination of the critical intergenic and genic region between the promoter region of *LRR kinase 1* and *ACT* in these assemblies revealed evidence of potential sequence expansion and contraction across the different genomes (Figure 5.17). While the physical distance of this interval is approximately the same in both the Chinese Spring and W7984 scaffold (10,321 and 10,106 bp respectively), the corresponding distance in the *T. urartu* scaffold was relatively shorter (6,900 bp). This is mainly due to a truncation of a retrotransposon sequence located in this interval in the *T. urartu* scaffold. However in the *tauschii* scaffold, the same interval covers a minimum physical distance of 30 kbp (including ambiguous gap sequences).

This expanded genomic sequence of *A. tauschii* also contains additional gene models for two hypothetical proteins (*F775_02800* and *F775_02801*). These gene models are not present in the other genomes examined. The highest sequences identity obtained when the protein sequences of these genes were used in a BLASTP analysis on Ensembl Plant and NBCI was to *Bradi1g00640*. A similar reciprocal BLASTP analysis with Bradi1g00640 sequence against the entire available grass database on Ensembl Plant returned back hits to the same *Ae. tauschii* genes. This suggests that these genes could be truly homologous to *Bradi1g00640* in the Triticeae genomes (See Chapter 4, section 4.2.2). However, F775_02800 and F775_02801 only share very low protein sequence identities to Bradi1g00640 (39% and 38% respectively).

Besides the comparative sequence analysis of this critical *Phs* interval across species, we also compared sequences obtained from the assembly of flow-sorted 4AL chromosome arm of Claire and Option (Chapter Four, section 4.2.9). We obtained 11 contigs each from Claire and Option assemblies that map across the *LRR Kinase-1-ACT* interval (Figure 5.18). Although these contigs do not overlap to produce a contiguous sequence that spans across the *Phs* interval, all of the contigs obtained showed very high sequence similarity to the Chinese Spring sequence (identity >=99%). There were however contigs that do not align correctly to the Chinese Spring scaffold. These include Claire contig 5959768 and Option contig 80786.

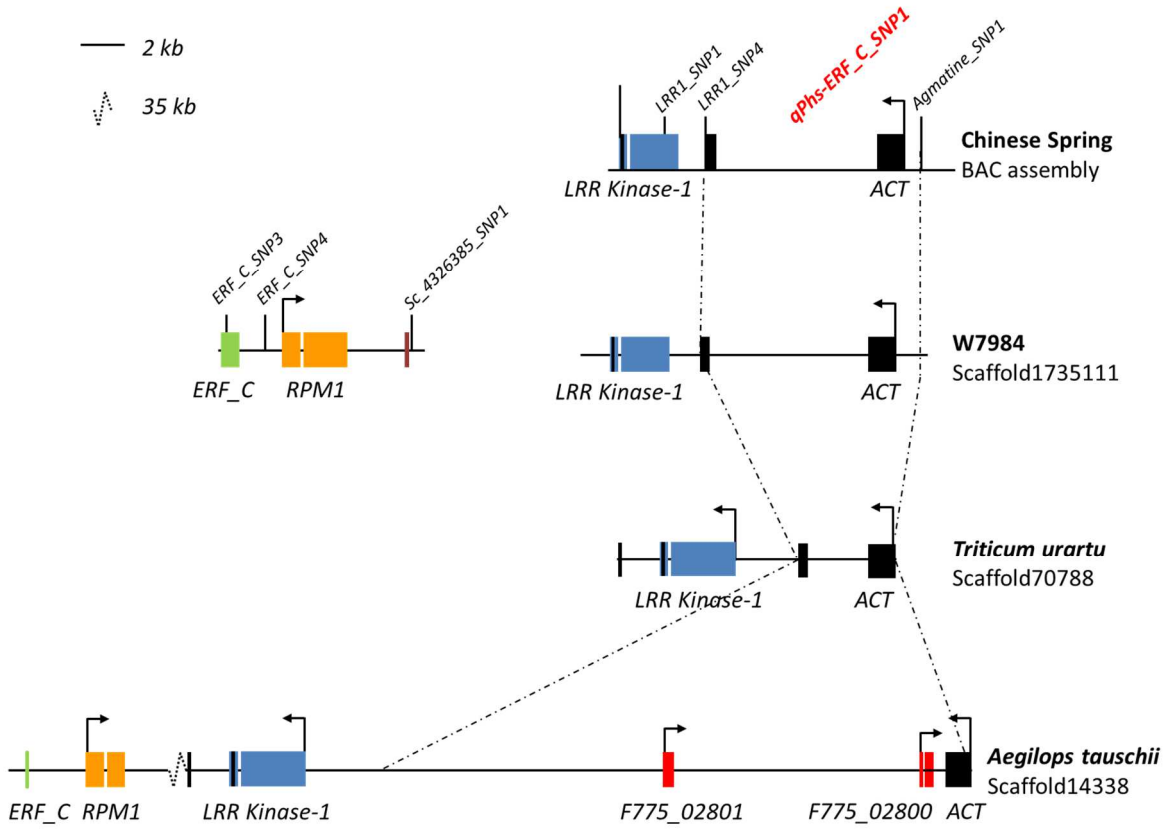


Figure 5.17: Comparative sequence analysis of the critical 10 kb *Phs* interval across different species. Scaffold sequences corresponding to the critical 10 kb *Phs* interval were obtained from W7984, *Triticum urartu and Aegilops tauschii* assemblies. Gene models for *ERF_C* (green box), *RPM1* (amber boxes), *LRR Kinase-1* (blue boxes), ACT (black boxes), F775_02800 and F775_02801 (red boxes) are indicated with each individual box representing an exon. The orientation and transcription start site for each gene is indicated by the arrows above each gene model. Vertical lines around each gene model represent the SNP position of the genetic marker designed in this interval. All the components of the figure are drawn to scale with the scale bar presented at the top left corner of the figure. Due to the relatively longer length of the *Ae. tauschii* assembly, a 35 kb interval between the *RPM1* and *LRR kinase 1* gene was shortened.

For Claire contig 5959768 (1790 bp), different parts of this contig aligned to two different regions of the Chinese Spring assembly. The first 1,424 bp sequence of the contig aligned well to sequences just downstream of the first exon of ACT genes. However, the last 360 bp aligned 6,108 bp further downstream away from the rest of the contig. This split alignment of contig 5959768 might have functional implication on the transcription of *ACT* as the displaced 3' end of the contig could extend into first exon sequence of *ACT*. Similarly, the alignment of the sequence of Option contig 80786 (1280 bp) was split into two. The first 762 bp sequence of this contig aligned to a

similar position as for contig5959768. However, the last 518 bp sequence of this contig could only be aligned with the Chinese Spring sequence in the reverse orientation.

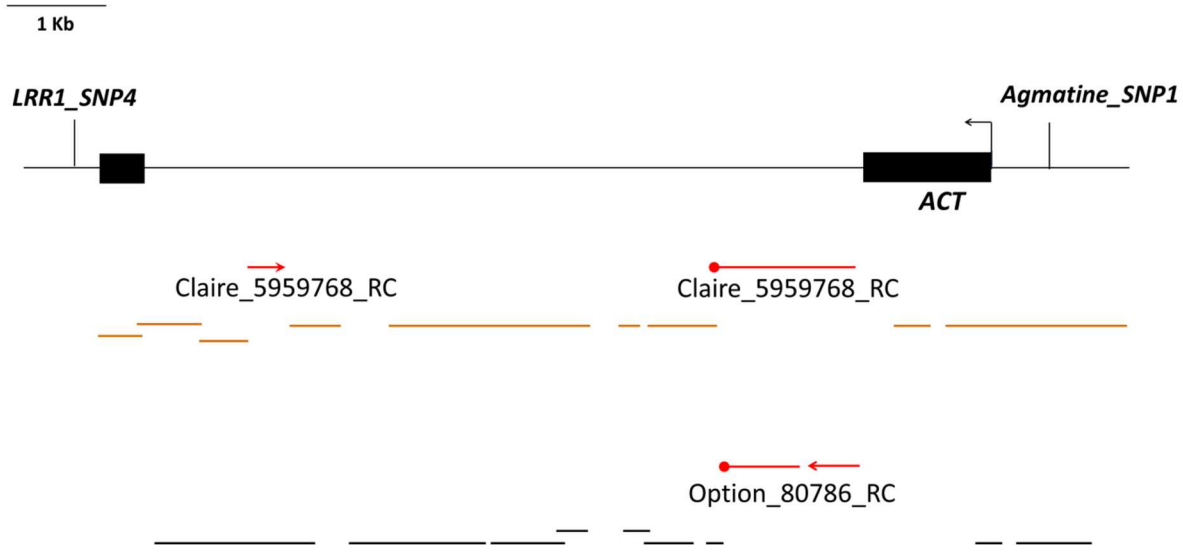


Figure 5.18: Alignment of Claire and Option contigs to the critical 10 kb *Phs* interval. Contigs from Claire (amber lines) and Option were (black lines) aligned to the 10 kb interval between LRR Kinase - 1 and ACT genes. Contigs from both assemblies with split alignment are indicated with the red lines with the dot indicating the beginning of each contig. All the components of the figure are drawn to scale with the scale bar presented at the top left corner of the figure.

Taken together, the comparative sequence analysis across the hexaploid and ancestral diploid wheat genomes highlights a series of genomic aberration within the physical sequence of the *Phs* interval. This is also evident in the misalignment of the contigs obtained from flow-sorted chromosome arm sequence of Claire and Option parent. We are however yet to find the causal polymorphism underlying *Phs* effect in this interval.

5.3 Discussion

In the previous chapter, we report on the fine-mapping of *Phs* to a 0.2 cM interval containing at least 17 high-confidence gene models based on physical map information. While such a small genetic interval is suitable for the deployment of *Phs* into elite cultivar through marker-assisted breeding, it is still of paramount importance to gain a mechanistic understanding of how *Phs* exerts its effect at the molecular level. For this to be fully achieved, it is important to delimit the *Phs* interval to a single gene locus thereby identifying the causal gene underlying the *Phs* effect. To this end, we described in this chapter the efforts toward the positional cloning of the *Phs* locus.

Development of high-throughput genetic markers in the Phs interval

As was discussed in the introduction of this chapter, the success of many a positional cloning project depends, among many other things, on the presence of sufficient recombination frequencies within the target locus and also the availability of genetic markers by which these recombination events can be defined. To address this, we took advantage of the sequence resources provided by syntenic comparison and the *Phs* physical map construction to develop a total of 30 high-throughput genetic markers which are evenly distributed across the 17 genic loci identified in the *Phs* interval. This by far represents the highest marker density ever reported around the *Phs* locus (Cabral et al., 2014, Lin et al., 2015, Ogbonnaya et al., 2007). This is particularly important for the region of the *Phs* interval which has very low gene conservation across the grass genomes (see Chapter 4, section 4.2.8). All of the 14 genetic markers developed within this low conservation region would not have been possible without the availability of the physical sequence for this region. This further highlights the advantage of such a physical map resources and also the disadvantage of sole dependence on synteny-based genetic mapping approach.

The high marker density achieved within the *Phs* interval was by no doubt facilitated by the generation of flow-sorted 4AL chromosome arm sequences from the parental Claire and Option lines. This allowed for a more rapid and targeted mining for polymorphisms in the *Phs* interval. Given the efficacy of this flow–sorted chromosome arm sequences, we envisage that this approach can easily complement already

available high-density SNP genotyping arrays that are currently being used in wheat while also providing the advantage of population-specific SNP mining (Jordan et al., 2015, Wang et al., 2014)

Fine-scaled mapping of the Phs locus

To maximise the chances of capturing recombination events that will help break the gene linkage in the *Phs* interval, we undertook the positional cloning effort of *Phs* in three independent populations and under different environmental conditions. The positional cloning effort in the two bi-parental population (OC and AR Pop) across three different experiments delimited the *Phs* locus to a genetic interval containing 10 genes including *PP1-Like*, *MAPK-6*, *Activating Signal Co-integrator*, *ERF_B, ERF_C, RPM1, LRR1-Kinase, ACT, Malonyl Transferase* and a gene encoding for a Hypothetical Protein. These experiments also exclude *LRR kinase-2, ACC oxidase-1, Ubiquitin Conjugating enzymes, Myosin-J Protein and the PM19 genes* (*PM19-A1* and *PM19-A2)*- from the *Phs* effect.

This fine-mapping result disagrees with result reported by (Barrero et al., 2015) in which the *PM19* genes were identified as the main candidate genes for a 4AL dormancy QTL (same as *Phs*). They also show through a transgenic approach that the *PM19* genes are positive regulator of seed dormancy in wheat. In addition, an expression analysis of genes in the *Phs* interval in the previous chapter supports the idea of the *PM19* genes underlying the *Phs* effect. However, we showed non-linkage of the sprouting phenotype to the *PM19* gene loci in at least 20 recombinant lines across both the OC and AR Populations. The phenotype of these 20 recombinant lines provides conclusive evidence to suggest that the *PM19* genes are not the main cause of the *Phs* effect, at least in UK bread wheat germplasm.

We cannot however rule out an indirect involvement of the *PM19* genes in the expression of the *Phs* effect. It is possible that the *PM19* genes, being positive regulators of primary dormancy in wheat, acts downstream of the main gene(s) underlying the *Phs* effect. Hence, effects arising from the main *Phs* causal gene will

have downstream consequences on the regulation of the *PM19* genes. This could account for the differential expression of *PM19-A1* and *PM19-A2* in contrasting *Phs* parents even when they are not the causal gene from which the *Phs* effect originates. In support of this hypothesis, other *PM19* genes from the B and D genomes (*PM19_B1, PM19_B2* and *PM19_D2*) were also differentially expressed in dormant and non-dormant parents. This further suggest that the *Phs* effect is not limited to and does not originate from the A genome copies of the *PM19* genes alone.

While it is very unlikely, it is also possible that the 4AL PHS QTL identified in the UK and Australian wheat germplasm are distinct and are under different genetic control. This could account for the different results obtained from the mapping population used by Barrero et al. and the mapping population used in this study. In line with this, we previously show that the parental lines used in these two studies have different haplotypes of *PM19-A1* and *PM19-A2* and that this might have functional implication on the regulation and role of the *PM19* genes in these germplasms. Due to the fact that the study by Barrero *et al*. was only published in the last few months of this Ph.D. project, we have not been able to characterise this haplotype difference in the *PM19* genes and the implication they have on the *Phs* effect. More work therefore need to be done on this in the future.

The fine-mapping experiments in the AR and OC populations also highlight evidence of possible suppression of recombination in the critical *Phs* interval in this study. Despite the large number of gametes screened (10,720 gametes, 5,360 plants) we could not identify any recombination event between the 10 genes linked to the *Phs* phenotypes in these populations. A possible reason for the suppressed recombination could be a high level of polymorphism (SNPs and indel) between the parental genotypes in this region. Such polymorphisms are known to have an inhibitory effect on the rate of crossovers in both plant and animal genomes (Jeffreys and Neumann, 2005, Dooner, 1986, Yao and Schnable, 2005). In addition, the occurrence of large chromosomal deletions or inversion within the target region of any one of the parental genotypes used in this study can also contribute to the reduced recombination frequencies observed (Jiang et al., 2007, Lowry and Willis, 2010). While the suppression of

recombination in these populations constitute a bottleneck in the fine–mapping effort, it nevertheless provides some indication of the genomic features of the *Phs* interval.

We overcame the limitation posed by the lack of recombination events at the *Phs* interval by screening 6,500 lines (13,000 gametes) from the diverse multi-parental breeding populations of our industrial partner - Limagrain UK. Such multi-parental population are associated with higher meiotic recombination frequencies (Huang et al., 2012). However many of the recombination event identified in these lines did not arise from recent crossover events. This is because, they were also found in the parental genotype from which the breeding lines were developed. Nevertheless, the valuable recombination events present in these lines afforded us the opportunity to delimit further the *Phs* interval.

Despite the very diverse pedigree and genetic architecture of the newly identified Limagrain recombinant lines, we still observed a strong effect arising from the segregation of *Phs* in these lines. This implies that, in after-ripened seeds, the *Phs* effect is dominant over the effect of any other dormancy controlling loci that could be segregating in the background of these lines. This dominant effect of *Phs* allowed us to classify lines as low or high sprouting. Nevertheless, there was still variation in the level of sprouting resistance observed in the low sprouting lines. This is probably due to the interaction of the *Phs* loci with other background loci and with the environment. Therefore, in other to optimally characterise the effect of *Phs*, it is important to study the recombination haplotype present in these line in a common and isogenic genetic background. To address this, we are currently undertaking a crossing program which will transfer the unique and critical recombination haplotypes present in five of the Limagrain recombinant lines into a common Claire or Option background.

Genetic discordance and genomic aberration in the Phs interval

Importantly, the two fine-mapping experiments in the Limagrain recombinant concordantly show the linkage of *Phs* to the *ERF_C_SNP1* genetic marker in almost all of the recombinants examined. High-resolution genetic mapping around this critical linkage interval placed the *ERF_C_SNP1* marker in the physical interval located between *LRR Kinase-1* and *ACT*. In addition, this also suggests that the intervening gap identified in our *Phs* physical map is not relevant for the *Phs* effect.

However, the new genetic placement of the *ERF_C_SNP1* marker does not agree with the physical position of the *ERF_C* gene from which the marker sequence was obtained. Although not widely reported, it is possible that such discordance in the physical and genetic map placement of genes/marker are common in plant and animal genomes and have not hitherto be uncovered due to inadequate genetic map resolution and lack of physical reference map for many species. However, with the advent of high-throughput genotyping technology and the availability of contiguous physical sequences, we envisage that more of this discordance will be uncovered in many genomes.

For instance, a study by Felcher et al. (2012) on the integration of two potato linkage maps with the potato reference genome sequence, identified 71 discordant markers which have differences in their genetic and physical map position. This represents about 1 % of the total number of markers examined in this study. Such parology of genetic marker seems to be common to polyploid species (Griffin et al., 2011, Korbecka et al., 2010).

We could not find any physical sequence representation for the *ERF_C_SNP1* between the *LRR Kinase-1* and *ACT* genes interval. We hypothesise that this is caused by structural lesions (large deletions or inversion) within this genomic interval in the Chinese Spring cultivar. The strongest evidence for this hypothesis comes from sequence analysis of a homologous *Aegilops tauschii* assembly scaffold which has two additional gene models in this interval - F775_02800 and F775_02801. Although the functions of these new genes are not known, a reciprocal BLAST analysis shows that they could be the previously unfound homologue of *Bradi1g00640* in the Triticeae.

However, the very low sequences conservation of these genes to *Bradi1g00640* might suggest the occurrence of gene degeneration and sub-functionalisation events (Innan and Kondrashov, 2010) within this genomic region. This might also account for the inability to identify homologues for *Bradi1g00630, Bradi1g00650* and *Bradi1g00670* in Triticeae genomes.

Additional evidence for the presence of genomic aberration within this critical *Phs* region is obtained from the misalignment of contigs obtained from the assembly of flow-sorted chromosome arm sequences of Claire and Option. Such sequence misalignment could arise from a fault in the assembly of the Chinese Spring BAC used as a reference or even that of the Claire and Option sequences. However contig/scaffold sequences obtained from independent IWGSC CSS chromosome arm assembly (International Wheat Genome Sequencing Consortium (IWGSC), 2014) and the W7984 whole genome shotgun assembly (Chapman et al., 2015) validate the Chinese Spring sequence assembly used in this study to be correct. This, therefore, implies that Claire and Option varieties might have some irregular sequence structure that affect the correct assembly of sequence reads in this region.

On the other hand, if these contigs misalignments are real, this would have functional implications on the exon structure and transcription of the adjacent *ACT* gene. Although the *ACT* genes are expressed in germinated cereal grains, there is currently no evidence for the involvement of *ACT* in the control of dormancy or germination. Rather they are involved in the biosynthesis of the antifungal hydroxycinnamoylagmatine derivatives. (Burhenne et al., 2003, BIRD and SMITH, 1984, Nomura et al., 2007)

In conclusion, the genetic evidence obtained from the many different fine-mapping experiments in three different populations along with physical map evidence suggest that the causal polymorphism(s) from the *Phs* locus is located between the *LRR Kinase-1* and *ACT*. We are however yet to identify the causal polymorphism(s) underlying the *Phs* locus. Based on all the evidence from comparative sequence analysis of this interval, we hypothesis that the genomic lesions discovered in this region may be more extensive in the genomes of the contrasting *Phs* parents than what is observed in the

Chinese Spring cultivar and this may in turn give rise to polymorphisms that are linked to the *Phs* phenotype. This could also be the cause of suppressed recombination frequencies in the biparental populations. Many of these genomic aberrations were only discovered towards the end of this Ph.D. project, and this did not leave enough time to characterise fully this critical interval in contrasting *Phs* parents. We, however, acknoweledge that more work will need to be done to identify the functional polymorphism(s) underlying the *Phs* effect.

6 Chapter 6: General Discussion

6.1 Implication of this study for PHS/PMA resistance breeding in the UK

In this project, we undertook the characterisation of six QTL for resistance to PHS and PMA. These included PHS QTL on wheat chromosomes 1A, 2D, 3A, and 4A; and a PMA QTL on chromosome 7B. These QTL were first identified along with many other QTL for PHS/PMA resistance in DH lines made from multiple bi-parental populations comprised entirely of UK wheat varieties. This shows that there are large variations for resistance to PHS and PMA in UK germplasm, and these can be harnessed for breeding new PHS and PMA resistant varieties. However, due to the strong environmental influence on PHS and PMA expression, many of these QTL are not stable and cannot always guarantee effective resistance to PHS/PMA under the UK climatic conditions. The six QTL chosen in this study showed consistent resistance effect across multiple years of field trials (2005-2008) in different field locations in the UK, suggesting their suitability for deploying stable PHS/PMA resistance in the UK. In this section, I will discuss how the findings from this study can be used for this purpose.

6.1.1 PHS/PMA QTL characterisation informs breeding strategies

To maximise PHS/PMA resistance within a target environment, it is important to characterise the QTL detected in the target environment as this will highlight the mode and timing of action of the QTL. Unfortunately, this is not commonly the case in many QTL studies for PHS and PMA resistance. However, the physiological characterisation done in this study has begun to shed new light on the genetic and physiological regulation of some of these QTL as well as how environmental elements like temperature affect their expression.

Using F_3 lines, we show that most of the QTL (1A, 2D, $3A_{SR}$ and 4A) exert their resistance by modulating the expression of dormancy in seeds albeit at different stages during grain maturation and after-ripening. In addition, the magnitude of the effect of these QTL also varies. The 2D QTL showed effect on grain dormancy during the late stages of physiological maturity and at harvest ripeness. While other QTL, like the 1A,

$3A_{SR}$ and 4A QTL showed later effects after harvest ripeness. The 7B QTL however showed no effect on PHS resistance but affects the induction of PMA in grains.

It should, however, be noted, that besides the validation of these QTL in F_3 population, we could not further confirm some of these QTL effects in BC_3F_2 NILs. This was mainly due to residual heterogeneity in the NILs and also differences in the experimental environments (CER vs. field). However, this does not constitute a major problem for the translation of the research output of this study because some of the QTL showed effect in the field which is the most relevant environment for breeding and farming purposes.

This highlights the importance of choice of environment for experiments in crop research. While the CER is an excellent facility for studying the genetic effects of a locus without the modulating effects of the interaction between different environmental elements, some complex genetic loci still require environmental dynamism before their effect can be observed. Furthermore, for a direct translation of experimental finding into breeding and agricultural uses, it is important that the experimental set-up be similar in many respects to the end-use set-up. I, therefore, suggest that the CER and other artificial experimental growth facilities should be used only to complement field experimentation in field crop research.

6.1.2 Combinatorial breeding strategies for PHS/PMA resistance in the UK

The variety of effects of the PHS/PMA QTL presents a unique opportunity to achieve stable PHS/PMA resistance which is expressed for a longer period during grain maturation and after-ripening under the UK climatic conditions. This can be achieved by combining the early PHS resistance effect of the 2D QTL with the resistance afforded by the 1A, $3A_{SR}$ or the 4A (*Phs*) QTL, thereby extending the resistance from harvest ripeness to the early stages of after-ripening. Such a longer expression of PHS resistance could be beneficial under the ever-changing climatic environment which is increasingly having an impact on the timing of harvest in the UK farming system (Olesen et al., 2011). Furthermore by combining these QTL with the 7B QTL for PMA

resistance, breeders could achieve a robust and effective solution to the problem of low HFN in the UK.

The choice of germplasm into which to combine these QTL is also very important. Many of the QTL were originally identified in populations made from relatively old UK wheat varieties. While the resistance that some of these varieties afford is very useful and relevant, many of these varieties themselves might not be competitive in the UK wheat market. It is, therefore, important that the resistance identified be combined into currently utilised varieties. To this end, we are currently combining the QTL together into a common and relevant UK wheat variety as part of a follow-on DEFRA-LINK project (HFNII LINK Project) aimed at addressing the problem of low HFN in the UK.

Furthermore, the agronomic characterisation of some of these QTL showed that the QTL do not negatively affect yield. This means that introduction of these QTL into modern varieties will not be detrimental. However, we have not examined the effect of these QTL on other important traits like disease resistance or grain quality trait like protein content. Neither have we examined the effect of the QTL on the rheological properties of bread. Rheological properties are particularly important because the resistances studied in this project are targeted for use in bread-making varieties. I, therefore, suggest that more agronomic and grain quality characterisation of the QTL effects be done.

In addition, by combining the QTL into a common genetic background, it will be possible to study the epistatic relationship between the QTL. While it is likely that QTL with different expression pattern function through different molecular pathways, there could still be some indirect cross talks in their expression. For instance, Mori et al. (2005) looked at the effect of combining the *Phs*.ocs-3A QTL on chromosome 3AS with *Phs* on 4AL and a 4BL QTL, all of which were identified in recombinant Inbreed lines (RIL)s made from a cross of Chinese Spring with the dormant Zen variety. They found that RILs which have the Zen alleles at the three QTL had enhanced resistance to PHS than lines with just the individual QTL. This suggested that the QTL have an additive effect on one another.

Kulwal et al. (2004) also identified 8 main QTL for PHS and found that 5 of the main QTL were involved in digenic epistatic interactions. Some of these interactions had an overall negative effect on PHS resistance. Given this precedence, it is possible that some of the QTL identified in the UK germplasm could display epistatic relationships.

In addition to the physiological and agronomic characterisation we did, we also identified more genetic markers within each of the QTL interval. While only the fine mapping of *Phs* is described in this project, we also are currently fine-mapping some of the other QTL as part of the wider HFNII LINK Project. Although not as advanced as the fine mapping of *Phs*, these fine-mapping efforts have led to the identification of markers that that are linked to the 1A, 2D and 7B QTL

The fine-mapping of *Phs* to a less than 0.2 cM in UK germplasm represents one of the biggest achievements of this project. This is because *Phs* has important relevance to PHS resistance in the UK. From our interaction with our partners from the breeding industry, we gather that many of the elite varieties currently grown in the UK have the susceptible 4AL parents (Robigus or Claire) in their pedigree. The mean that many of these lines potentially have the susceptible *Phs* allele.

In support of this, haplotype analysis of the 35 varieties (excluding new varieties) on the Home Grown Cereal Association (HGCA) 2015/2016 recommended winter wheat list, showed that 63% of the listed varieties have the Robigus/Claire type *Phs* susceptibility haplotype across the 0.2 cM *Phs* interval. Although this is not always reflected in the HFN score ascribed to these lines due to the effect of other PHS and PMA QTL, we believe that many of these varieties could be prone to PHS losses if they are not timely harvested before grain after-ripening. The markers we have generated within the *Phs* interval will therefore prove valuable for breeding for PHS resistance in UK varieties. We have passed on the *Phs* markers information as well as those of other QTL markers to our breeding partners, and this information are currently being used to develop new varieties with stable and extended PHS resistance.

6.1.3 Importance of collaborative research with industrial partners

This project, as part of a larger HFNII LINK project, showcases the benefit of collaborative engagements between research institutions and industrial partners. Without the contributions of our breeding partners, we would not have been able to achieve some of the milestones reached in this project. I presume the most notable contribution of our breeding partners is their assistance in screening for and identifying recombination events between the 0.2 cM *Phs* interval, at a time when we could not find any recombination events in our biparental populations. This finding made such a significant impact on our effort to delimit *Phs* to an approx. 10 kb locus. Similarly, the results generated in this study have greatly benefited our breeding partners. These have been valuable in informing their strategy for breeding for PHS/PMA resistance in the UK.

6.2 Implication of this work for global PHS/PMA breeding and wheat genetic studies

6.2.1 Targeted breeding for PHS resistance

I will, first of all, like to highlight the need for targeted breeding of PHS/PMA resistance to the specific agro-climatic system in which it is to be deployed. This is because some QTL seem to be important in certain agro-climatic zones and not others. For instance, we could not detect the segregation of *Phs*.ocs-3A QTL (the 3AS QTL regulated by *TaMFT*) in UK germplasm even though this QTL is widely utilised in Japan (Chono et al., 2015, Nakamura et al., 2011a) and the United States (Lei et al., 2013, Liu et al., 2013b).

We analysed the sequence of *TaMFT* in varieties in which we identified QTL on chromosome 3A (Rialto, Savannah, Avalon and Cadenza). We could not however detect the causative promoter SNP reported by Nakamura et al. (2011a) neither could we find the splice site SNPs reported by Liu et al. (2013b). All of the varieties we sequenced have the susceptible Chinese Spring allele at the Nakamura et al promoter SNP position and also the dormant Rio Blanc SNP allele at the exon splice site reported

by Liu et al. This mean that these alleles might be have been fixed in UK germplasm. A more extensive sequence analysis of more UK wheat varieties will confirm this hypothesis.

On the other hand, *Phs* which we show to have large effect on PHS resistance in UK germplasm and also segregates in Australian (Barrero et al., 2015), Canadian (Cabral et al., 2014), Japanese (Torada et al., 2008) and Chinese (Chen et al., 2008) germplasm, seems not to be important for PHS resistance in the US (Scott Haley, Colorado State University, personal communication). This might be due to the effect of temperature on *Phs* as demonstrated by us and by Barrero et al. (2015). While we showed that *Phs* is still effective even when plant were grown under low ambient temperature, Barrero et al. showed that high temperature during the later stages of grain development diminishes the effect of the 4AL QTL. So the temperature regime during grain development within each agro-climatic zone matters when considering which QTL to deploy for PHS/PMA resistance.

6.2.2 *Phs* mode of action and genetic linkage

Furthermore, our characterisation of *Phs* will prove to be valuable for the study and deployment of this 4AL QTL in different parts of the world. Some studies have reported the inability to consistently detect *Phs* effect across different years of studies. We think one of the reasons for this is that the phenotyping scheme employed might not have taken into account the timing of *Phs* effect. And indeed many PHS researchers are not familiar with this timing of *Phs* effect. We have however shown in this study that *Phs* functions during the stages of grain after-ripening (2-4 weeks after harvest ripeness). Using this knowledge, we have consistently detected the effect of *Phs* in our parental lines and experimental populations. We believe this knowledge of the timing of *Phs* effect will help researchers around the world to consistently detect and study this QTL in their various germplasm. We also anticipate that using this knowledge, the *Phs* effect will be detected in many new germplasm and populations.

Besides our physiological characterisation of *Phs*, we have delimited *Phs* to a 10 kb physical distance interval in the Chinese Spring BAC assembly. This is by far the closest

linkage to *Phs* in European germplasm that has ever been reported. Although we are yet to find the causal polymorphism responsible for the *Phs* effect, we believe the sequences of the markers (*LRR1-SNP4* and *Agmatine_SNP1*) flanking this 10 kb interval can be used to select for *Phs* resistance across other germplasm as well.

6.2.3 Caution on the use of candidate gene approach

Our work also shows that caution should be taken when employing a candidate gene approach for the analysis of QTL in wheat. While this is a quick and easy method for potentially cloning genes underlying QTL effect, it can sometimes lead to misleading results. Irrespective of the criteria by which such candidate genes are chosen, either by empirical transcriptomic analysis, association evidence or through inferred function by comparison with other homologues or related genes, this approach is still liable to false positive identification.

Our experience with the *PM19* genes demonstrates this argument. We identified the *PM19* genes as candidates for the *Phs* effect based on their genetic association with the initial 0.2 cM *Phs* locus linked to the sprouting phenotype and also their expression pattern. However, our more detailed genetic analysis dissociates the *Phs* effect from the *PM19* genes in at least 26 recombinant lines from two bi-parental populations and a multi-parental population, tested across 5 different experiments. We, therefore, recommend that results obtained using the candidate gene approach should be verified by more reliable genetic analyses. We, however, recognise that genetic validation of candidate genes might not always be possible due to the limitation of recombination events in some QTL interval. In such cases, we recommend that multiple validation approaches, like transgenic complementation or knock down or the use of mutagenised populations, be employed to validate candidate genes.

6.3 Recommendation for further studies

We recognise that a 3-year Ph.D. project (excluding my rotation Ph.D. year) only allows for a limited amount of experimental work to be done. This is particularly true for a Ph.D. programme in crop research which requires substantial time for germplasm development and field validation. In addition, the PHS and PMA traits studied in this project could only be examined at the very late stages of the wheat plant development. This often results in a 6-month cycle from the planning of experiments to final execution. Due to this limitation of time, we have not been able to look into some aspects of this studies that are vital. Some of these are discussed briefly below.

6.3.1 Detailed sequence analysis of the 10 kb genomic *Phs* interval in parents with contrasting of sprouting phenotype

In chapter five, we described the fine mapping of *Phs* to a 10 kb interval and we reported on the preliminary comparative analysis of this interval between contrasting *Phs* parents (Claire and Option) using flow-sorted chromosome contig sequences. However due to the fragmented nature of the contigs used, we could not compare a contiguous sequence assembly across the 10 kb interval between the parents. To address this, we will take one of two approaches. First, using the Chinese Spring BAC sequence as a reference, multiple PCR primers pairs have been designed across the entire 10 kb interval for the amplification of overlapping fragments from contrasting *Phs* parents. We hypothesise that the presence of a genomic lesion (deletion or inversion) in one of these parents will result in null amplification or an amplicon with an unexpected size.

One potential problem however with the PCR approach is that the Chinese Spring sequences might not be a suitable reference for the sequences of any of the varieties used in this study. A more comprehensive approach will therefore be the development of an ungridded BAC library of contrasting *Phs* parents. This will allow for the screening, isolation and comparison of corresponding BAC sequence of the resistant and susceptible parents across and beyond the 10 kb interval. One advantage of this approach is that large genomic deletions or inversion can be better characterised.

6.3.2 Analysis of functional effects of the *PM19* gene editing in barley

Using the RNA-guided Cas9 genome editing system, we successfully developed transgene-free barley plants with mutations in *HvPM19-1*. However due to time limitation, as well as the heterozygosity of many of the mutations created, we could not functional characterise the *PM19* mutations. For *HvPM19-1*, only one of two T_1 lines with homozygous mutation produced progeny with mutations in the T_2 generation and in the absence of the Cas9 transgene. The progeny all have frameshift deletions leading to a premature stop codon towards the 3' end of the gene. For *HvPM19-3,* we could not identify any homozygous mutant T_2 line in the absence of the Cas9 transgene. However, we identified a line with heterozygous mutations in both *HvPM19-1* and *HvPM19-3*. This line is currently being grown for the selection of homozygous mutant plants. Using the homozygous *hvpm19-1* mutant and homozygous *hvpm19-1hvpm19-3* double mutant, we will examine the effect of these mutations on seed germination potential in barley and thereby elucidate on *PM19's* role in germination in barley.

6.3.3 Allelic diversity study using diverse world germplasm segregating for *Phs*

We currently have a varietal panel consisting of most of the germplasms that have been used in studying the *Phs* QTL across the world. These include nine varieties reported to have the dormant *Phs* QTL allele and six *Phs* susceptible varieties. We hope to add the varieties used by (Barrero et al., 2015)to this varietal panel. Using all the available markers across the *Phs* 0.2 cM interval, we will characterise the haplotype structure present in these varieties. This will not only shed more light on the nature of the *Phs* interval in these diverse germplasm but will also identify markers that are polymorphic across the various germplasm and could therefore be used as global diagnostic markers for the *Phs* effect. In addition, we will investigate in more details the haplotype diversity within the *Phs* interval between the Australian germplasm used by Barrero et al. and the UK germplasm used in this study. This could potential reveal whether the gene(s) underlying the *Phs* effect is the same in these populations or are

two closely linked genes that have been selected differently in UK and Australia breeding programmes.

6.3.4 Further hormonal characterisation of the *Phs* effect

The physiological characterisation of the *Phs* effect in Chapter 4 highlights a possible ABA-GA interaction in the control of the *Phs* effect. This was because GA and fluridone treatment significantly reduced the resistance effect of *Phs*. However, these experiments were only done with two contrasting *Phs* parents and have not been fine-mapped to the *Phs* locus. We will therefore repeat these treatment on some of the critical recombinant lines. In addition, the experiment could also be modified to include more treatments that will shed more light on the ABA-GA interaction. Furthermore, examination of the role of the endogenous content of these hormones in the expression of the *Phs* effect will also be important.

6.4 My personal development through the Ph.D. programme

As a young undergraduate student in Nigeria, I desired to contribute towards agricultural productivity in Africa by engaging in genetic research that will address pressing food security challenges of the continent and by training the new generation of African scientists to address the critical scientific capability shortages in Africa. To actualise these desires, I embarked on a journey of personal development which eventually brought me to the UK for my MSc and Ph.D. degree. I, therefore, see my Ph.D. degree as a phase in this journey of self-actualization.

In the course of this Ph.D., I have gained both academic and research skills that are valuable for my career development. As a rotation Ph.D. student, I first explored and gathered knowledge from studies in various biological systems ranging from microbial system (*Streptomyces*) to a model plant (Arabidopsis) and also a crop plant (wheat). Working on these various systems broadened my knowledge of the molecular regulation of many biological processes. While I thoroughly enjoyed and gained a lot

from working on these rotation projects, the most important learnings of my degree were gained in the course of my main Ph.D. project (this study).

Embarking on this project has helped me to acquire skills in the study of the quantitative and qualitative genetics of wheat and indeed of cereals. I have learnt how to use conventional genetic approaches (fine-mapping approach) as well as various NGS-based approaches to unravel the genetic control of important traits in crops. While doing this, I gathered several molecular biology skills including DNA and RNA extractions, gene-specific primer design, genomic PCR, qRT-PCR, gene cloning, Sanger sequencing, SNP mining, marker development and high-throughput genotyping.

In addition, I also gained many skills on NGS sample preparation and data management as well as on the use of web-based and command line (Linux) based bioinformatics tools. These include ability to perform BAC extraction, prepare NGS library for sequencing, perform quality check on NGS data, assembly read into scaffolds, align NGS reads to reference sequence, SNP calling, BLAST analysis on custom dataset, sequences and text file manipulation and the ability to interprete and evaluate the findings of NGS analysis. Central to all of these learning process however, are the skills I gained in the generation and testing of hypothesis through the reiterative process of experimental planning and execution as well as analysis and critical review of experimental data.

I have also gathered numerous transferable skills necessary for my professional development. These include the ability to work in a team, take responsibility for tasks, manage projects, and also to self-organise. I have as well been privileged to supervise an MSc. student, and this helped me to develop my leadership skill. In addition, I have developed my sciences communication skills over the years. This has helped me to win numerous prizes and awards, the most notable being an invitation to the SET for Britain Competition at the UK house of Commons in 2015.

I am by no means implying that my Ph.D. journey has been without its challenges nor that I have been a perfect student; I nevertheless consider the strides I have made as important steps towards self-actualisation. Importantly I am thankful to everyone, particularly my supervisor – Dr Cristobal Uauy- who contributed towards my

development in one way or the other. Now that I am at the tail end of my Ph.D., I do not consider this to be the end of my personal development but rather I see this as the beginning of a new phase of my development. In this new phase, I intend to build on my research excellence and work towards the conception and development of a career defining research programme.

7 References

Addisu, M., Snape, J. W., Simmonds, J. R. & Gooding, M. J. 2009. Reduced height (Rht) and photoperiod insensitivity (Ppd) allele associations with establishment and early growth of wheat in contrasting production systems. *Euphytica,* 166**,** 249-267.

Ahdb 2014. AHDB Cereals & Oilseeds Quality Calculator.

Ahdb. 2015a. *AHDB Cereals & Oilseeds Variety Survey* [Online]. Available: http://cereals.ahdb.org.uk/markets/survey-results.aspx [Accessed 20/09/2015.

Ahdb 2015b. Market Data Centre, Corn Returns Ex-farm Regional Spot & Forward Prices - Weekly Prices

Akhunov, E., Akhunova, A. & Dvorak, J. 2005. BAC libraries of Triticum urartu, Aegilops speltoides and Ae. tauschii, the diploid ancestors of polyploid wheat. *Theor Appl Genet,* 111**,** 1617 - 1622.

Akhunov, E., Nicolet, C. & Dvorak, J. 2009. Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *Theoretical and Applied Genetics,* 119**,** 507-517.

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., Dubcovsky, J., Gill, B. S., Gu, Y. Q., Hadam, J., Heo, H., Huo, N., Lazo, G. R., Luo, M.-C., Ma, Y. Q., Matthews, D. E., Mcguire, P. E., Morrell, P. L., Qualset, C. O., Renfro, J., Tabanao, D., Talbert, L. E., Tian, C., Toleno, D. M., Warburton, M. L., You, F. M., Zhang, W. & Dvorak, J. 2010. Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. *BMC Genomics,* 11**,** 702-702.

Akpinar, B. A., Magni, F., Yuce, M., Lucas, S. J., Simkova, H., Safar, J., Vautrin, S., Berges, H., Cattonaro, F., Dolezel, J. & Budak, H. 2015. The physical map of wheat chromosome 5DS revealed gene duplications and small rearrangements. *BMC Genomics,* 16**,** 453.

Alani, A., Bruzau, F., Raymond, P., Saintges, V., Leblanc, J. M. & Pradet, A. 1985. Germination, Respiration, and Adenylate Energy-Charge of Seeds at Various Oxygen Partial Pressures. *Plant Physiology,* 79**,** 885-890.

Albrecht, T., Oberforster, M., Kempf, H., Ramgraber, L., Schacht, J., Kazman, E., Zechner, E., Neumayer, A., Hartl, L. & Mohler, V. 2015. Genome-wide association mapping of preharvest sprouting resistance in a diversity panel of European winter wheats. *Journal of Applied Genetics***,** 1-9.

Ali-Rachedi, S., Bouinot, D., Wagner, M. H., Bonnet, M., Sotta, B., Grappin, P. & Jullien, M. 2004. Changes in endogenous abscisic acid levels during dormancy release and maintenance of mature seeds: studies with the Cape Verde Islands ecotype, the dormant model of Arabidopsis thaliana. *Planta,* 219**,** 479-88.

Allen, A. M., Barker, G. L., Berry, S. T., Coghill, J. A., Gwilliam, R., Kirby, S., Robinson, P., Brenchley, R. C., D'amore, R., Mckenzie, N., Waite, D., Hall, A., Bevan, M., Hall, N. & Edwards, K. J. 2011. Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (Triticum aestivum L.). *Plant Biotechnol J,* 9**,** 1086-99.

Allen, A. M., Barker, G. L. A., Wilkinson, P., Burridge, A., Winfield, M., Coghill, J., Uauy, C., Griffiths, S., Jack, P., Berry, S., Werner, P., Melichar, J. P. E., Mcdougall, J., Gwilliam, R., Robinson, P. & Edwards, K. J. 2013. Discovery and development of exome-based, co-dominant single nucleotide polymorphism

markers in hexaploid wheat (Triticum aestivum L.). *Plant Biotechnology Journal,* 11**,** 279-295.

Allen, R. C. 1997. Agriculture and the Origins of the State in Ancient Egypt. *Explorations in Economic History,* 34**,** 135-154.

Anderberg, R. J. & Walker-Simmons, M. K. 1992. Isolation of a wheat cDNA clone for an abscisic acid-inducible transcript with homology to protein kinases. *Proceedings of the National Academy of Sciences of the United States of America,* 89**,** 10183-10187.

Araus, J. L. & Cairns, J. E. 2014. Field high-throughput phenotyping: the new crop breeding frontier. *Trends in Plant Science,* 19**,** 52-61.

Barnard, A. & Smith, M. F. 2009. The effect of rainfall and temperature on the preharvest sprouting tolerance of winter wheat in the dryland production areas of the Free State Province. *Field Crops Research,* 112**,** 158-164.

Barrero, J. M., Cavanagh, C., Verbyla, K. L., Tibbits, J. F., Verbyla, A. P., Huang, B. E., Rosewarne, G. M., Stephen, S., Wang, P., Whan, A., Rigault, P., Hayden, M. J. & Gubler, F. 2015. Transcriptomic analysis of wheat near-isogenic lines identifies PM19-A1 and A2 as candidates for a major dormancy QTL. *Genome Biol,* 16**,** 93.

Barrero, J. M., Downie, A. B., Xu, Q. & Gubler, F. 2014. A Role for Barley CRYPTOCHROME1 in Light Regulation of Grain Dormancy and Germination. *The Plant Cell,* 26**,** 1094-1104.

Barrero, J. M., Jacobsen, J. V., Talbot, M. J., White, R. G., Swain, S. M., Garvin, D. F. & Gubler, F. 2012. Grain dormancy and light quality effects on germination in the model grass Brachypodium distachyon. *New Phytol,* 193**,** 376-86.

Barrero, J. M., Mrva, K., Talbot, M. J., White, R. G., Taylor, J., Gubler, F. & Mares, D. J. 2013. Genetic, Hormonal, and Physiological Analysis of Late Maturity α-Amylase in Wheat. *Plant Physiology,* 161**,** 1265-1277.

Baskin, C. C. & Baskin, J. M. 2001. *Seeds: ecology, biogeography, and evolution of dormancy and germination*, Elsevier.

Baskin, J. M. & Baskin, C. C. 2004. A classification system for seed dormancy. *Seed Science Research,* 14**,** 1-16.

Baulcombe, D. C., Huttly, A. K., Martienssen, R. A., Barker, R. F. & Jarvis, M. G. 1987. A novel wheat alpha-amylase gene (alpha-Amy3). *Mol Gen Genet,* 209**,** 33-40.

Bazin, J., Langlade, N., Vincourt, P., Arribat, S., Balzergue, S., El-Maarouf-Bouteau, H. & Bailly, C. 2011. Targeted mRNA Oxidation Regulates Sunflower Seed Dormancy Alleviation during Dry After-Ripening. *The Plant Cell,* 23**,** 2196-2208.

Beaudoin, N., Serizet, C., Gosti, F. & Giraudat, J. 2000. Interactions between abscisic acid and ethylene signaling cascades. *Plant Cell,* 12**,** 1103-15.

Becraft, P. W. & Yi, G. 2010. Regulation of aleurone development in cereal grains. *Journal of Experimental Botany*.

Belderok, B. 1963. Report on the investigation of on dormancy and sprouting in the ear carried out in 1961. *Plant Breeding Abstract,* 33**,** 371.

Benech-Arnold, R. L., Gualano, N., Leymarie, J., Côme, D. & Corbineau, F. 2006. Hypoxia interferes with ABA metabolism and increases ABA sensitivity in embryos of dormant barley grains. *Journal of Experimental Botany,* 57**,** 1423-1430.

Bewley, J. D. 1997. Seed Germination and Dormancy. *The Plant Cell,* 9**,** 1055-1066.

Bird, C. R. & Smith, T. A. 1984. Agmatine Metabolism and Hordatine Formation in Barley Seedlings. *Annals of Botany,* 53**,** 483-488.

Bolot, S., Abrouk, M., Masood-Quraishi, U., Stein, N., Messing, J., Feuillet, C. & Salse, J. 2009. The 'inner circle' of the cereal genomes. *Current Opinion in Plant Biology,* 12**,** 119-125.

Borojevic, K. & Borojevic, K. 2005. The Transfer and History of "Reduced Height Genes" (Rht) in Wheat from Japan to Europe. *Journal of Heredity,* 96**,** 455-459.

Borrill, P., Adamski, N. & Uauy, C. 2015a. Genomics as the key to unlocking the polyploid potential of wheat. *New Phytologist***,** n/a-n/a.

Borrill, P., Fahy, B., Smith, A. M. & Uauy, C. 2015b. Wheat Grain Filling Is Limited by Grain Filling Capacity rather than the Duration of Flag Leaf Photosynthesis: A Case Study Using NAM RNAi Plants. *PLoS One,* 10**,** e0134947.

Bourbousse, C., Ahmed, I., Roudier, F., Zabulon, G., Blondet, E., Balzergue, S., Colot, V., Bowler, C. & Barneche, F. 2012. Histone H2B monoubiquitination facilitates the rapid modulation of gene expression during Arabidopsis photomorphogenesis. *PLoS Genet,* 8**,** e1002825.

Bowden, W. M. 1959. THE TAXONOMY AND NOMENCLATURE OF THE WHEATS, BARLEYS, AND RYES AND THEIR WILD RELATIVES. *Canadian Journal of Botany,* 37**,** 657-684.

Box, M. S., Coustham, V., Dean, C. & Mylne, J. S. 2011. Protocol: A simple phenol-based method for 96-well extraction of high quality RNA from Arabidopsis. *Plant Methods,* 7**,** 7.

Boyes, D. C., Nam, J. & Dangl, J. L. 1998. The Arabidopsis thaliana RPM1 disease resistance gene product is a peripheral plasma membrane protein that is degraded coincident with the hypersensitive response. *Proceedings of the National Academy of Sciences of the United States of America,* 95**,** 15849-15854.

Breen, J., Wicker, T., Shatalina, M., Frenkel, Z., Bertin, I., Philippe, R., Spielmeyer, W., Simkova, H., Safar, J., Cattonaro, F., Scalabrin, S., Magni, F., Vautrin, S., Berges, H., Paux, E., Fahima, T., Dolezel, J., Korol, A., Feuillet, C. & Keller, B. 2013. A physical map of the short arm of wheat chromosome 1A. *PLoS One,* 8**,** e80272.

Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G. L., D'amore, R., Allen, A. M., Mckenzie, N., Kramer, M., Kerhornou, A., Bolser, D., Kay, S., Waite, D., Trick, M., Bancroft, I., Gu, Y., Huo, N., Luo, M. C., Sehgal, S., Gill, B., Kianian, S., Anderson, O., Kersey, P., Dvorak, J., Mccombie, W. R., Hall, A., Mayer, K. F., Edwards, K. J., Bevan, M. W. & Hall, N. 2012a. Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature,* 491**,** 705-10.

Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G. L. A., D'amore, R., Allen, A. M., Mckenzie, N., Kramer, M., Kerhornou, A., Bolser, D., Kay, S., Waite, D., Trick, M., Bancroft, I., Gu, Y., Huo, N., Luo, M. C., Sehgal, S., Gill, B., Kianian, S., Anderson, O., Kersey, P., Dvorak, J., Mccombie, W. R., Hall, A., Mayer, K. F. X., Edwards, K. J., Bevan, M. W. & Hall, N. 2012b. Analysis of the breadwheat genome using whole-genome shotgun sequencing. *Nature,* 491**,** 705-710.

Brkljacic, J., Grotewold, E., Scholl, R., Mockler, T., Garvin, D. F., Vain, P., Brutnell, T., Sibout, R., Bevan, M., Budak, H., Caicedo, A. L., Gao, C., Gu, Y., Hazen, S. P., Holt, B. F., Hong, S.-Y., Jordan, M., Manzaneda, A. J., Mitchell-Olds, T., Mochida, K., Mur, L. a. J., Park, C.-M., Sedbrook, J., Watt, M., Zheng, S. J. & Vogel, J. P. 2011. Brachypodium as a Model for the Grasses: Today and the Future. *Plant Physiology,* 157**,** 3-13.

Buckler, E. S. T., Thornsberry, J. M. & Kresovich, S. 2001. Molecular diversity, structure and domestication of grasses. *Genet Res,* 77**,** 213-8.

Buraas, T. & Skinnes, H. 1985. Development of Seed Dormancy in Barley, Wheat and Triticale under Controlled Conditions. *Acta Agriculturae Scandinavica,* 35**,** 233-244.

Burhenne, K., Kristensen, B. K. & Rasmussen, S. K. 2003. A New Class of N-Hydroxycinnamoyltransferases: PURIFICATION, CLONING, AND EXPRESSION OF A BARLEY AGMATINE COUMAROYLTRANSFERASE (EC 2.3.1.64). *Journal of Biological Chemistry,* 278**,** 13919-13927.

Cabral, A. L., Jordan, M. C., Mccartney, C. A., You, F. M., Humphreys, D. G., Maclachlan, R. & Pozniak, C. J. 2014. Identification of candidate genes, regions and markers for pre-harvest sprouting resistance in wheat (Triticum aestivum L.). *BMC Plant Biology,* 14**,** 340.

Cavanagh, C. R., Chao, S. M., Wang, S. C., Huang, B. E., Stephen, S., Kiani, S., Forrest, K., Saintenac, C., Brown-Guedira, G. L., Akhunova, A., See, D., Bai, G. H., Pumphrey, M., Tomar, L., Wong, D. B., Kong, S., Reynolds, M., Da Silva, M. L., Bockelman, H., Talbert, L., Anderson, J. A., Dreisigacker, S., Baenziger, S., Carter, A., Korzun, V., Morrell, P. L., Dubcovsky, J., Morell, M. K., Sorrells, M. E., Hayden, M. J. & Akhunov, E. 2013. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proceedings of the National Academy of Sciences of the United States of America,* 110**,** 8057-8062.

Chantret, N., Salse, J., Sabot, F., Bellec, A., Laubin, B., Dubois, I., Dossat, C., Sourdille, P., Joudrier, P., Gautier, M. F., Cattolico, L., Beckert, M., Aubourg, S., Weissenbach, J., Caboche, M., Leroy, P., Bernard, M. & Chalhoub, B. 2008. Contrasted microcolinearity and gene evolution within a homoeologous region of wheat and barley species. *J Mol Evol,* 66**,** 138-50.

Chapman, J. A., Mascher, M., Buluc, A., Barry, K., Georganas, E., Session, A., Strnadova, V., Jenkins, J., Sehgal, S., Oliker, L., Schmutz, J., Yelick, K. A., Scholz, U., Waugh, R., Poland, J. A., Muehlbauer, G. J., Stein, N. & Rokhsar, D. S. 2015. A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biol,* 16**,** 26.

Chatterjee, N. & Nagarajan, S. 2006. Evaluation of water binding, seed coat permeability and germination characteristics of wheat seeds equilibrated at different relative humidities. *Indian J Biochem Biophys,* 43**,** 233-8.

Chen, C. X., Cai, S. B. & Bai, G. H. 2008. A major QTL controlling seed dormancy and pre-harvest sprouting resistance on chromosome 4A in a Chinese wheat landrace. *Molecular Breeding,* 21**,** 351-358.

Chen, D., Ma, X., Li, C., Zhang, W., Xia, G. & Wang, M. 2014a. A wheat aminocyclopropane-1-carboxylate oxidase gene, TaACO1, negatively regulates salinity stress in Arabidopsis thaliana. *Plant Cell Rep,* 33**,** 1815-27.

Chen, M., Macgregor, D. R., Dave, A., Florance, H., Moore, K., Paszkiewicz, K., Smirnoff, N., Graham, I. A. & Penfield, S. 2014b. Maternal temperature history activates Flowering Locus T in fruits to control progeny dormancy according to time of year. *Proceedings of the National Academy of Sciences,* 111**,** 18787-18792.

Chen, Z. Y., Guo, X. J., Chen, Z. X., Chen, W. Y., Liu, D. C., Zheng, Y. L., Liu, Y. X., Wei, Y. M. & Wang, J. R. 2015. Genome-wide characterization of developmental stage- and tissue-specific transcription factors in wheat. *BMC Genomics,* 16**,** 125.

Cheng, C.-R., Oldach, K., Mrva, K. & Mares, D. 2014. Analysis of high pI α-Amy-1 gene family members expressed in late maturity α-amylase in wheat (Triticum aestivum L.). *Molecular Breeding,* 33**,** 519-529.

Cheng, W. H., Chiang, M. H., Hwang, S. G. & Lin, P. C. 2009. Antagonism between abscisic acid and ethylene in Arabidopsis acts in parallel with the reciprocal regulation of their metabolism and signaling pathways. *Plant Mol Biol,* 71**,** 61-80.

Cheng, Y., Dai, X. & Zhao, Y. 2006. Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in Arabidopsis. *Genes & Development,* 20**,** 1790-1799.

Chi, X.-F., Lou, X.-Y. & Shu, Q.-Y. 2008. Progressive fine mapping in experimental populations: An improved strategy toward positional cloning. *Journal of Theoretical Biology,* 253**,** 817-823.

Chiang, G. C., Bartsch, M., Barua, D., Nakabayashi, K., Debieu, M., Kronholm, I., Koornneef, M., Soppe, W. J., Donohue, K. & De Meaux, J. 2011. DOG1 expression is predicted by the seed-maturation environment and contributes to geographical variation in germination in Arabidopsis thaliana. *Mol Ecol,* 20**,** 3336-49.

Chitnis, V. R., Gao, F., Yao, Z., Jordan, M. C., Park, S. & Ayele, B. T. 2014. After-Ripening Induced Transcriptional Changes of Hormonal Genes in Wheat Seeds: The Cases of Brassinosteroids, Ethylene, Cytokinin and Salicylic Acid. *PLoS ONE,* 9**,** e87543.

Chono, M., Matsunaka, H., Seki, M., Fujita, M., Kiribuchi-Otobe, C., Oda, S., Kojima, H. & Nakamura, S. 2015. Molecular and genealogical analysis of grain dormancy in Japanese wheat varieties, with specific focus on MOTHER OF FT AND TFL1 on chromosome 3A. *Breed Sci,* 65**,** 103-9.

Choulet, F., Alberti, A., Theil, S., Glover, N., Barbe, V., Daron, J., Pingault, L., Sourdille, P., Couloux, A., Paux, E., Leroy, P., Mangenot, S., Guilhot, N., Le Gouis, J., Balfourier, F., Alaux, M., Jamilloux, V., Poulain, J., Durand, C., Bellec, A., Gaspin, C., Safar, J., Dolezel, J., Rogers, J., Vandepoele, K., Aury, J. M., Mayer, K., Berges, H., Quesneville, H., Wincker, P. & Feuillet, C. 2014. Structural and functional partitioning of bread wheat chromosome 3B. *Science,* 345**,** 1249721.

Choulet, F., Wicker, T., Rustenholz, C., Paux, E., Salse, J., Leroy, P., Schlub, S., Le Paslier, M. C., Magdelenat, G., Gonthier, C., Couloux, A., Budak, H., Breen, J., Pumphrey, M., Liu, S., Kong, X., Jia, J., Gut, M., Brunel, D., Anderson, J. A., Gill, B. S., Appels, R., Keller, B. & Feuillet, C. 2010. Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. *Plant Cell,* 22**,** 1686-701.

Cockram, J., Jones, H., Leigh, F. J., O'sullivan, D., Powell, W., Laurie, D. A. & Greenland, A. J. 2007. Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. *Journal of Experimental Botany,* 58**,** 1231-1244.

Colasuonno, P., Maria, M. A., Blanco, A. & Gadaleta, A. 2013. Description of durum wheat linkage map and comparative sequence analysis of wheat mapped DArT markers with rice and Brachypodium genomes. *BMC Genet,* 14**,** 114.

Comai, L. 2005. The advantages and disadvantages of being polyploid. *Nat Rev Genet,* 6**,** 836-846.

Consortium, T. I. W. G. S. 2014. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science,* 345.

Cooper, D., Krawczak, M., Polychronakos, C., Tyler-Smith, C. & Kehrer-Sawatzki, H. 2013. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Human Genetics,* 132**,** 1077-1130.

Corbineau, F., Xia, Q., Bailly, C. & El-Maarouf-Bouteau, H. 2014. Ethylene, a key factor in the regulation of seed dormancy. *Front Plant Sci,* 5**,** 539.

Cristinarodriguez, M., Petersen, M. & Mundy, J. 2010. Mitogen-Activated Protein Kinase Signaling in Plants. *Annual Review of Plant Biology,* 61**,** 621-649.

Cvikova, K., Cattonaro, F., Alaux, M., Stein, N., Mayer, K. F., Dolezel, J. & Bartos, J. 2015. High-throughput physical map anchoring via BAC-pool sequencing. *BMC Plant Biol,* 15**,** 99.

De Laethauwer, S., Reheul, D., De Riek, J. & Haesaert, G. 2012. Vp1 expression profiles during kernel development in six genotypes of wheat, triticale and rye. *Euphytica,* 188**,** 61-70.

Debeaujon, I., Léon-Kloosterziel, K. M. & Koornneef, M. 2000. Influence of the Testa on Seed Dormancy, Germination, and Longevity in Arabidopsis. *Plant Physiology,* 122**,** 403-414.

Defra 2014. Farming Statistics – 2014 wheat and barley production, UK. *In:* AFFAIRS, D. F. E. F. A. R. (ed.).

Defra 2015. Farming Statistics Final crop area and cattle, sheep and pig populations At 1 June 2015 - England. *In:* AGRICULTURE, D. F. E. F. A. R. (ed.).

Del Carmen Rodríguez-Gacio, M., Matilla-Vázquez, M. A. & Matilla, A. J. 2009. Seed dormancy and ABA signaling: The breakthrough goes on. *Plant Signaling & Behavior,* 4**,** 1035-1048.

Dennis, E. S. & Peacock, W. J. 2009. Vernalization in cereals. *Journal of Biology,* 8**,** 57-57.

Devos, K. M., Dubcovsky, J., Dvořák, J., Chinoy, C. N. & Gale, M. D. 1995. Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination. *Theoretical and Applied Genetics,* 91**,** 282-288.

Distelfeld, A., Uauy, C., Fahima, T. & Dubcovsky, J. 2006. Physical map of the wheat high-grain protein content gene Gpc-B1 and development of a high-throughput molecular marker. *New Phytologist,* 169**,** 753-763.

Dodson, J. R., Li, X., Zhou, X., Zhao, K., Sun, N. & Atahan, P. 2013. Origin and spread of wheat in China. *Quaternary Science Reviews,* 72**,** 108-111.

Dolezel, J., Kubalakova, M., Cihalikova, J., Suchankova, P. & Simkova, H. 2011. Chromosome analysis and sorting using flow cytometry. *Methods Mol Biol,* 701**,** 221-38.

Dooner, H. K. 1986. Genetic Fine Structure of the BRONZE Locus in Maize. *Genetics,* 113**,** 1021-36.

Drouaud, J., Camilleri, C., Bourguignon, P. Y., Canaguier, A., Berard, A., Vezon, D., Giancola, S., Brunel, D., Colot, V., Prum, B., Quesneville, H. & Mezard, C. 2006. Variation in crossing-over rates across chromosome 4 of Arabidopsis thaliana reveals the presence of meiotic recombination "hot spots". *Genome Res,* 16**,** 106-14.

Dubcovsky, J. & Dvorak, J. 2007. Genome Plasticity a Key Factor in the Success of Polyploid Wheat Under Domestication. *Science,* 316**,** 1862-1866.

Eckardt, N. A. 2010. Evolution of Domesticated Bread Wheat. *The Plant Cell,* 22**,** 993-993.

Edgerton, M. D. 2009. Increasing Crop Productivity to Meet Global Needs for Feed, Food, and Fuel. *Plant Physiology,* 149**,** 7-13.

Ekblom, R. & Wolf, J. B. W. 2014. A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications,* 7**,** 1026-1042.
El-Maarouf-Bouteau, H., Meimoun, P., Job, C., Job, D. & Bailly, C. 2013. Role of protein and mRNA oxidation in seed dormancy and germination. *Frontiers in Plant Science,* 4**,** 77.
Etherington, G. J., Monaghan, J., Zipfel, C. & Maclean, D. 2014. Mapping mutations in plant genomes with the user-friendly web application CandiSNP. *Plant Methods,* 10**,** 41.
Evers, T. & Millar, S. 2002. Cereal Grain Structure and Development: Some Implications for Quality. *Journal of Cereal Science,* 36**,** 261-284.
Eversole, K., Feuillet, C., Mayer, K. F. X. & Rogers, J. 2014. Slicing the wheat genome. *Science,* 345**,** 285-287.
Fang, J., Chai, C., Qian, Q., Li, C., Tang, J., Sun, L., Huang, Z., Guo, X., Sun, C., Liu, M., Zhang, Y., Lu, Q., Wang, Y., Lu, C., Han, B., Chen, F., Cheng, Z. & Chu, C. 2008. Mutations of genes in synthesis of the carotenoid precursors of ABA lead to pre-harvest sprouting and photo-oxidation in rice. *The Plant Journal,* 54**,** 177-189.
Fao. 2014. *FAOSTAT* [Online]. Available: http://faostat3.fao.org [Accessed 09/09/2015.
Fao. 2015. *FAO Cereal Supply and Demand Brief* [Online]. Available: http://www.fao.org/worldfoodsituation/csdb/en/ [Accessed 21/09/2015.
Faris, J. D., Fellers, J. P., Brooks, S. A. & Gill, B. S. 2003. A bacterial artificial chromosome contig spanning the major domestication locus Q in wheat and identification of a candidate gene. *Genetics,* 164**,** 311-21.
Faris, J. D., Haen, K. M. & Gill, B. S. 2000. Saturation Mapping of a Gene-Rich Recombination Hot Spot Region in Wheat. *Genetics,* 154**,** 823-835.
Faris, J. D., Zhang, Z. & Chao, S. 2014. Map-based analysis of the tenacious glume gene Tg-B1 of wild emmer and its role in wheat domestication. *Gene,* 542**,** 198-208.
Farrell, A. D. & Kettlewell, P. S. 2008. The Effect of Temperature Shock and Grain Morphology on Alpha-amylase in Developing Wheat Grain. *Annals of Botany,* 102**,** 287-293.
Felcher, K. J., Coombs, J. J., Massa, A. N., Hansey, C. N., Hamilton, J. P., Veilleux, R. E., Buell, C. R. & Douches, D. S. 2012. Integration of Two Diploid Potato Linkage Maps with the Potato Genome Sequence. *PLoS ONE,* 7**,** e36347.
Feldman, M. & Levy, A. A. 2012. Genome evolution due to allopolyploidization in wheat. *Genetics,* 192**,** 763-74.
Feuillet, C. & Keller, B. 2002. Comparative Genomics in the Grass Family: Molecular Characterization of Grass Genome Structure and Evolution. *Annals of Botany,* 89**,** 3-10.
Feuillet, C., Travella, S., Stein, N., Albar, L., Nublat, A. & Keller, B. 2003. Map-based isolation of the leaf rust disease resistance gene Lr10 from the hexaploid wheat (Triticum aestivum L.) genome. *Proceedings of the National Academy of Sciences,* 100**,** 15253-15258.
Finch-Savage, W. E. & Leubner-Metzger, G. 2006. Seed dormancy and the control of germination. *New Phytol,* 171**,** 501-23.
Fleury, D., Luo, M. C., Dvorak, J., Ramsay, L., Gill, B. S., Anderson, O. D., You, F. M., Shoaei, Z., Deal, K. R. & Langridge, P. 2010. Physical mapping of a large plant genome using global high-information-content-fingerprinting: the distal region of the wheat ancestor Aegilops tauschii chromosome 3DS. *BMC Genomics,* 11**,** 382.

Flintham, J., Adlam, R., Bassoi, M., Holdsworth, M. & Gale, M. D. 2002. Mapping genes for resistance to sprouting damage in wheat. *Euphytica,* 126**,** 39-45.

Flintham, J. E. 2000. Different genetic components control coat-imposed and embryo-imposed dormancy in wheat. *Seed Science Research,* 10**,** 43-50.

Flintham, J. E. & Gale, M. D. 1982. The Tom Thumb dwarfing gene, Rht3 in wheat, I. Reduced pre-harvest damage to breadmaking quality. *Theor Appl Genet,* 62**,** 121-6.

Flintham, J. E., Holdsworth, M. J., Jack, P., Kettlewell, P. S. & Phillips, A. L. unpublished. An integrated approach to stabilising HFN in wheat: screens, genes and understanding.

Fofana, B., Humphreys, D. G., Rasul, G., Cloutier, S., Brûlé-Babel, A., Woods, S., Lukow, O. M. & Somers, D. J. 2009. Mapping quantitative trait loci controlling pre-harvest sprouting resistance in a red × white seeded spring wheat cross. *Euphytica,* 165**,** 509-521.

Food & Agriculture Organization of the United, N. 2014. FAOSTAT. Rome, Italy: FAO.

Frenkel, Z., Paux, E., Mester, D., Feuillet, C. & Korol, A. 2010. LTC: a novel algorithm to improve the efficiency of contig assembly for physical mapping in complex genomes. *BMC Bioinformatics,* 11**,** 584.

Friedman, W. E. 1998. The evolution of double fertilization and endosperm: an "historical" perspective. *Sexual Plant Reproduction,* 11**,** 6-16.

Fu, D., Uauy, C., Distelfeld, A., Blechl, A., Epstein, L., Chen, X., Sela, H., Fahima, T. & Dubcovsky, J. 2009. A kinase-START gene confers temperature-dependent resistance to wheat stripe rust. *Science,* 323**,** 1357-60.

Fujii, H. & Zhu, J.-K. 2009. Arabidopsis mutant deficient in 3 abscisic acid-activated protein kinases reveals critical roles in growth, reproduction, and stress. *Proceedings of the National Academy of Sciences,* 106**,** 8380-8385.

Furbank, R. T. & Tester, M. 2011. Phenomics – technologies to relieve the phenotyping bottleneck. *Trends in Plant Science,* 16**,** 635-644.

Gale, M. D. & Devos, K. M. 1998. Comparative genetics in the grasses. *Proceedings of the National Academy of Sciences,* 95**,** 1971-1974.

Gallavotti, A. & Whipple, C. J. 2015. Positional cloning in maize (Zea mays subsp. mays, Poaceae). *Applications in Plant Sciences,* 3**,** apps.1400092.

Gamble, P. E. & Mullet, J. E. 1986. Inhibition of carotenoid accumulation and abscisic acid biosynthesis in fluridone-treated dark-grown barley. *Eur J Biochem,* 160**,** 117-21.

Ganal, M. & Röder, M. 2007. Microsatellite and SNP Markers in Wheat Breeding. *In:* VARSHNEY, R. & TUBEROSA, R. (eds.) *Genomics-Assisted Crop Improvement.* Springer Netherlands.

Gao, F., Rampitsch, C., Chitnis, V. R., Humphreys, G. D., Jordan, M. C. & Ayele, B. T. 2013a. Integrated analysis of seed proteome and mRNA oxidation reveals distinct post-transcriptional features regulating dormancy in wheat (Triticum aestivum L.). *Plant Biotechnology Journal,* 11**,** 921-932.

Gao, X., Hu, C. H., H.Z., L., Yao, Y. J., Meng, M., Dong, L., Zhao, W. C., Chen, Q. J. & Li, X. Y. 2013b. FACTORS AFFECTING PRE-HARVEST SPROUTING RESISTANCE IN WHEAT (TRITICUM AESTIVUM L.): A REVIEW. *The Journal of Animal and Plant Sciences* 23**,** 556-565.

Gasperini, D., Greenland, A., Hedden, P., Dreos, R., Harwood, W. & Griffiths, S. 2012. Genetic and physiological analysis of Rht8 in bread wheat: an alternative source

of semi-dwarfism with a reduced sensitivity to brassinosteroids. *Journal of Experimental Botany*.

Gatford, K. T., Eastwood, R. F. & Halloran, G. M. 2002. Germination inhibitors in bracts surrounding the grain of Triticum tauschii. *Functional Plant Biology,* 29**,** 881-890.

Gerjets, T., Scholefield, D., Foulkes, M. J., Lenton, J. R. & Holdsworth, M. J. 2010. An analysis of dormancy, ABA responsiveness, after-ripening and pre-harvest sprouting in hexaploid wheat (Triticum aestivum L.) caryopses. *Journal of Experimental Botany*, 61**,** 597-607.

Giorgi, D., Farina, A., Grosso, V., Gennaro, A., Ceoloni, C. & Lucretti, S. 2013. FISHIS: Fluorescence In Situ Hybridization in Suspension and Chromosome Flow Sorting Made Easy. *PLoS ONE,* 8**,** e57994.

Giulini, A., La Rocca, N., Durantini, D., Malgioglio, A., Dalla Vecchia, F., Manzotti, P., Consonni, G., Rascio, N. & Gavazzi, G. 2011. A viviparous mutant of maize exhibiting permanent water stress symptoms. *Plant Growth Regulation,* 64**,** 99-108.

Golan, G., Oksenberg, A. & Peleg, Z. 2015. Genetic evidence for differential selection of grain and embryo weight during wheat evolution under domestication. *Journal of Experimental Botany*.

Golovina, E. A., Hoekstra, F. A. & Van Aelst, A. C. 2001. The competence to acquire cellular desiccation tolerance is independent of seed morphological development. *Journal of Experimental Botany*, 52**,** 1015-1027.

Goncharov, N. 2011. Genus Triticum L. taxonomy: the present and the future. *Plant Systematics and Evolution,* 295**,** 1-11.

Gosti, F., Beaudoin, N., Serizet, C., Webb, A. A., Vartanian, N. & Giraudat, J. 1999. ABI1 protein phosphatase 2C is a negative regulator of abscisic acid signaling. *The Plant Cell*, 11**,** 1897-1910.

Graeber, K., Linkies, A., Steinbrecher, T., Mummenhoff, K., Tarkowská, D., Turečková, V., Ignatz, M., Sperber, K., Voegele, A., De Jong, H., Urbanová, T., Strnad, M. & Leubner-Metzger, G. 2014. DELAY OF GERMINATION 1 mediates a conserved coat-dormancy mechanism for the temperature- and gibberellin-dependent control of seed germination. *Proceedings of the National Academy of Sciences,* 111**,** E3571-E3580.

Graeber, K., Nakabayashi, K., Miatton, E., Leubner-Metzger, G. & Soppe, W. J. 2012. Molecular mechanisms of seed dormancy. *Plant Cell Environ,* 35**,** 1769-86.

Grdc. 2015. *Cereal Growth Stages* [Online]. Available: http://www.grdc.com.au/uploads/documents/GRDC%20Cereal%20Growth%20Stages%20Guide1.pdf.

Griffin, P. C., Robin, C. & Hoffmann, A. A. 2011. A next-generation sequencing method for overcoming the multiple gene copy problem in polyploid phylogenetics, applied to Poa grasses. *BMC Biol,* 9**,** 19.

Griffiths, J., Murase, K., Rieu, I., Zentella, R., Zhang, Z.-L., Powers, S. J., Gong, F., Phillips, A. L., Hedden, P., Sun, T.-P. & Thomas, S. G. 2006. Genetic Characterization and Functional Analysis of the GID1 Gibberellin Receptors in Arabidopsis. *The Plant Cell,* 18**,** 3399-3414.

Griffiths, S., Simmonds, J., Leverington, M., Wang, Y., Fish, L., Sayers, L., Alibert, L., Orford, S., Wingen, L., Herry, L., Faure, S., Laurie, D., Bilham, L. & Snape, J. 2009. Meta-QTL analysis of the genetic control of ear emergence in elite European winter wheat germplasm. *Theoretical and Applied Genetics,* 119**,** 383-395.

Griffiths, S., Simmonds, J., Leverington, M., Wang, Y., Fish, L., Sayers, L., Alibert, L., Orford, S., Wingen, L. & Snape, J. 2012. Meta-QTL analysis of the genetic control of crop height in elite European winter wheat germplasm. *Molecular Breeding,* 29**,** 159-171.

Groos, C., Gay, G., Perretant, M. R., Gervais, L., Bernard, M., Dedryver, F. & Charmet, D. 2002a. Study of the relationship between pre-harvest sprouting and grain color by quantitative trait loci analysis in a whitexred grain bread-wheat cross. *Theoretical and Applied Genetics,* 104**,** 39-47.

Groos, C., Gay, G., Perretant, M. R., Gervais, L., Bernard, M., Dedryver, F. & Charmet, G. 2002b. Study of the relationship between pre-harvest sprouting and grain color by quantitative trait loci analysis in a white×red grain bread-wheat cross. *Theoretical and Applied Genetics,* 104**,** 39-47.

Gu, X.-Y., Turnipseed, E. B. & Foley, M. E. 2008. The qSD12 Locus Controls Offspring Tissue-Imposed Seed Dormancy in Rice. *Genetics,* 179**,** 2263-2273.

Gu, X. Y., Liu, T., Feng, J., Suttle, J. C. & Gibbons, J. 2010. The qSD12 underlying gene promotes abscisic acid accumulation in early developing seeds to induce primary dormancy in rice. *Plant Mol Biol,* 73**,** 97-104.

Gualano, N. A. & Benech-Arnold, R. L. 2009. Predicting pre-harvest sprouting susceptibility in barley: Looking for "sensitivity windows" to temperature throughout grain filling in various commercial cultivars. *Field Crops Research,* 114**,** 35-44.

Gubler, F., Hughes, T., Waterhouse, P. & Jacobsen, J. 2008. Regulation of Dormancy in Barley by Blue Light and After-Ripening: Effects on Abscisic Acid and Gibberellin Metabolism. *Plant Physiology,* 147**,** 886-896.

Gubler, F., Millar, A. A. & Jacobsen, J. V. 2005. Dormancy release, ABA and pre-harvest sprouting. *Current Opinion in Plant Biology,* 8**,** 183-187.

Hagberg, S. 1960. A rapid method for determining α-amylase activity. *Cereal Chemist,* 37**,** 218.

Haider, N. 2013. The origin of the B-genome of bread wheat (Triticum aestivum L.). *Russian Journal of Genetics,* 49**,** 263-274.

Haiping, Z., Cheng, C., Jiming, F., Bo, Y., Hongqi, S. & Chuanxi I, M. 2009. Detecting and identifying polymorphism of Viviparous-B1 gene and its seed dormancy in mini-core collection of Chinese wheat varieties. *Journal of Agricultural Biotechnology/Nongye Shengwu Jishu Xuebao,* 17**,** 690-694.

Hamamura, Y., Nagahara, S. & Higashiyama, T. 2012. Double fertilization on the move. *Curr Opin Plant Biol,* 15**,** 70-7.

Hellsten, U., Wright, K. M., Jenkins, J., Shu, S., Yuan, Y., Wessler, S. R., Schmutz, J., Willis, J. H. & Rokhsar, D. S. 2013. Fine-scale variation in meiotic recombination in Mimulus inferred from population shotgun sequencing. *Proceedings of the National Academy of Sciences,* 110**,** 19478-19482.

Henderson, I. R. 2012. Control of meiotic recombination frequency in plant genomes. *Current Opinion in Plant Biology,* 15**,** 556-561.

Hernandez, P., Martis, M., Dorado, G., Pfeifer, M., Gálvez, S., Schaaf, S., Jouve, N., Šimková, H., Valárik, M., Doležel, J. & Mayer, K. F. X. 2012. Next-generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. *The Plant Journal,* 69**,** 377-386.

Higo, K., Ugawa, Y., Iwamoto, M. & Korenaga, T. 1999. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res,* 27**,** 297-300.

Himi, E., Maekawa, M., Miura, H. & Noda, K. 2011. Development of PCR markers for Tamyb10 related to R-1, red grain color gene in wheat. *Theor Appl Genet,* 122**,** 1561-76.

Hirano, K., Ueguchi-Tanaka, M. & Matsuoka, M. 2008. GID1-mediated gibberellin signaling in plants. *Trends Plant Sci,* 13**,** 192-9.

Hoang, H. H., Bailly, C., Corbineau, F. & Leymarie, J. 2013. Induction of secondary dormancy by hypoxia in barley grains and its hormonal regulation. *Journal of Experimental Botany,* 64**,** 2017-2025.

Holdsworth, M. J., Bentsink, L. & Soppe, W. J. J. 2008. Molecular networks regulating Arabidopsis seed maturation, after-ripening, dormancy and germination. *New Phytologist,* 179**,** 33-54.

Huang, B. E., George, A. W., Forrest, K. L., Kilian, A., Hayden, M. J., Morell, M. K. & Cavanagh, C. R. 2012. A multiparent advanced generation inter-cross population for genetic analysis in wheat. *Plant Biotechnology Journal,* 10**,** 826-839.

Ibgsc 2012. A physical, genetic and functional sequence assembly of the barley genome. *Nature,* 491**,** 711-716.

Imtiaz, M., Ogbonnaya, F. C., Oman, J. & Van Ginkel, M. 2008. Characterization of quantitative trait loci controlling genetic variation for preharvest sprouting in synthetic backcross-derived wheat lines. *Genetics,* 178**,** 1725-36.

Innan, H. & Kondrashov, F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet,* 11**,** 97-108.

International Barley Genome Sequencing Consortium, I. 2012. A physical, genetic and functional sequence assembly of the barley genome. *Nature,* 491**,** 711-716.

International, R. G. S. P. 2005. The map-based sequence of the rice genome. *Nature,* 436**,** 793-800.

International Rice Genome Sequencing Project 2005. The map-based sequence of the rice genome. *Nature,* 436**,** 793-800.

International Wheat Genome Sequencing Consortium (Iwgsc) 2014. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science,* 345.

Iyer, L. M., Burroughs, A. M. & Aravind, L. 2006. The ASCH superfamily: novel domains with a fold related to the PUA domain and a potential role in RNA metabolism. *Bioinformatics,* 22**,** 257-263.

Jaiswal, V., Mir, R. R., Mohan, A., Balyan, H. S. & Gupta, P. K. 2012. Association mapping for pre-harvest sprouting tolerance in common wheat (Triticum aestivum L.). *Euphytica,* 188**,** 89-102.

Jeffreys, A. J. & Neumann, R. 2005. Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot. *Hum Mol Genet,* 14**,** 2277-87.

Jiang, L., Zhang, W., Xia, Z., Jiang, G., Qian, Q., Li, A., Cheng, Z., Zhu, L., Mao, L. & Zhai, W. 2007. A paracentric inversion suppresses genetic recombination at the FON3 locus with breakpoints corresponding to sequence gaps on rice chromosome 11L. *Mol Genet Genomics,* 277**,** 263-72.

John, P., Reynolds, E. A., Prescott, A. G. & Bauchot, A. D. 1999. ACC Oxidase in the Biosynthesis of Ethylene. *In:* KANELLIS, A. K., CHANG, C., KLEE, H., BLEECKER, A. B., PECH, J. C. & GRIERSON, D. (eds.) *Biology and Biotechnology of the Plant Hormone Ethylene II.* Springer Netherlands.

Jordan, K. W., Wang, S., Lun, Y., Gardiner, L. J., Maclachlan, R., Hucl, P., Wiebe, K., Wong, D., Forrest, K. L., Sharpe, A. G., Sidebottom, C. H., Hall, N., Toomajian, C., Close, T., Dubcovsky, J., Akhunova, A., Talbert, L., Bansal, U. K., Bariana, H. S., Hayden, M. J., Pozniak, C., Jeddeloh, J. A., Hall, A. & Akhunov, E. 2015.

A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. *Genome Biol,* 16**,** 48.

Jupe, F., Witek, K., Verweij, W., Śliwka, J., Pritchard, L., Etherington, G. J., Maclean, D., Cock, P. J., Leggett, R. M., Bryan, G. J., Cardle, L., Hein, I. & Jones, J. D. G. 2013. Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *The Plant Journal,* 76**,** 530-544.

Kato, K., Nakamura, W., Tabiki, T., Miura, H. & Sawada, S. 2001. Detection of loci controlling seed dormancy on group 4 chromosomes of wheat and comparative mapping with rice and barley genomes. *Theoretical and Applied Genetics,* 102**,** 980-985.

King, R. W. & Von Wettstein-Knowles, P. 2000. Epicuticular waxes and regulation of ear wetting and pre-harvest sprouting in barley and wheat. *Euphytica,* 112**,** 157-166.

Kippes, N., Debernardi, J. M., Vasquez-Gross, H. A., Akpinar, B. A., Budak, H., Kato, K., Chao, S., Akhunov, E. & Dubcovsky, J. 2015. Identification of the VERNALIZATION 4 gene reveals the origin of spring growth habit in ancient wheats from South Asia. *Proceedings of the National Academy of Sciences*.

Kırmızı, S. & Bell, R. W. 2012. Responses of barley to hypoxia and salinity during seed germination, nutrient uptake, and early plant growth in solution culture. *Journal of Plant Nutrition and Soil Science,* 175**,** 630-640.

Knight, A. E. & Kendrick-Jones, J. 1993. A myosin-like protein from a higher plant. *J Mol Biol,* 231**,** 148-54.

Knox, R. E., Clarke, F. R., Clarke, J. M. & Fox, S. L. 2005. Genetic analysis of pre-harvest sprouting in a durum wheat cross. *Euphytica,* 143**,** 261-264.

Knox, R. E., Clarke, F. R., Clarke, J. M., Fox, S. L., Depauw, R. M. & Singh, A. K. 2012. Enhancing the identification of genetic loci and transgressive segregants for preharvest sprouting resistance in a durum wheat population. *Euphytica,* 186**,** 193-206.

Kobayashi, F., Wu, J., Kanamori, H., Tanaka, T., Katagiri, S., Karasawa, W., Kaneko, S., Watanabe, S., Sakaguchi, T., Hanawa, Y., Fujisawa, H., Kurita, K., Abe, C., Iehisa, J. C., Ohno, R., Safar, J., Simkova, H., Mukai, Y., Hamada, M., Saito, M., Ishikawa, G., Katayose, Y., Endo, T. R., Takumi, S., Nakamura, T., Sato, K., Ogihara, Y., Hayakawa, K., Dolezel, J., Nasuda, S., Matsumoto, T. & Handa, H. 2015. A high-resolution physical map integrating an anchored chromosome with the BAC physical maps of wheat chromosome 6B. *BMC Genomics,* 16**,** 595.

Koike, M., Takezawa, D., Arakawa, K. & Yoshida, S. 1997a. Accumulation of 19-kDa Plasma Membrane Polypeptide during Induction of Freezing Tolerance in Wheat Suspension-Cultured Cells by Abscisic Acid. *Plant and Cell Physiology,* 38**,** 707-716.

Koike, M., Takezawa, D., Arakawa, K. & Yoshida, S. 1997b. Accumulation of 19-kDa plasma membrane polypeptide during induction of freezing tolerance in wheat suspension-cultured cells by abscisic acid. *Plant Cell Physiol,* 38**,** 707-16.

Kondhare, K. R., Kettlewell, P. S., Farrell, A. D., Hedden, P. & Monaghan, J. M. 2012. Effects of exogenous abscisic acid and gibberellic acid on pre-maturity α-amylase formation in wheat grains. *Euphytica,* 188**,** 51-60.

Koornneef, M., Jorna, M. L., Brinkhorst-Van Der Swan, D. L. C. & Karssen, C. M. 1982. The isolation of abscisic acid (ABA) deficient mutants by selection of

induced revertants in non-germinating gibberellin sensitive lines of Arabidopsis thaliana (L.) heynh. *Theoretical and Applied Genetics,* 61**,** 385-393.

Koornneef, M., Reuling, G. & Karssen, C. M. 1984. The isolation and characterization of abscisic acid-insensitive mutants of Arabidopsis thaliana. *Physiologia Plantarum,* 61**,** 377-383.

Korbecka, G., Rymer, P. D., Harris, S. A. & Pannell, J. R. 2010. Solving the problem of ambiguous paralogy for marker loci: microsatellite markers with diploid inheritance in Allohexaploid Mercurialis annua (Euphorbiaceae). *J Hered,* 101**,** 504-11.

Koressaar, T. & Remm, M. 2007. Enhancements and modifications of primer design program Primer3. *Bioinformatics,* 23**,** 1289-91.

Kottearachchi, N. S., Uchino, N., Kato, K. & Miura, H. 2006. Increased grain dormancy in white-grained wheat by introgression of preharvest sprouting tolerance QTLs. *Euphytica,* 152**,** 421-428.

Krasileva, K., Buffalo, V., Bailey, P., Pearce, S., Ayling, S., Tabbita, F., Soria, M., Wang, S., Consortium, I., Akhunov, E., Uauy, C. & Dubcovsky, J. 2013. Separating homeologs by phasing in the tetraploid wheat transcriptome. *Genome Biology,* 14**,** R66.

Krattinger, S. G., Lagudah, E. S., Spielmeyer, W., Singh, R. P., Huerta-Espino, J., Mcfadden, H., Bossolini, E., Selter, L. L. & Keller, B. 2009. A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. *Science,* 323**,** 1360-3.

Kulwal, P., Ishikawa, G., Benscher, D., Feng, Z., Yu, L.-X., Jadhav, A., Mehetre, S. & Sorrells, M. 2012. Association mapping for pre-harvest sprouting resistance in white winter wheat. *Theoretical and Applied Genetics,* 125**,** 793-805.

Kulwal, P. L., Kumar, N., Gaur, A., Khurana, P., Khurana, J. P., Tyagi, A. K., Balyan, H. S. & Gupta, P. K. 2005. Mapping of a major QTL for pre-harvest sprouting tolerance on chromosome 3A in bread wheat. *Theoretical and Applied Genetics,* 111**,** 1052-1059.

Kulwal, P. L., Singh, R., Balyan, H. S. & Gupta, P. K. 2004. Genetic basis of pre-harvest sprouting tolerance using single-locus and two-locus QTL analyses in bread wheat. *Functional & Integrative Genomics,* 4**,** 94-101.

Kumar, S., Knox, R., Clarke, F., Pozniak, C., Depauw, R., Cuthbert, R. & Fox, S. 2015. Maximizing the identification of QTL for pre-harvest sprouting resistance using seed dormancy measures in a white-grained hexaploid wheat population. *Euphytica,* 205**,** 287-309.

Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D. L., Garcia-Hernandez, M., Karthikeyan, A. S., Lee, C. H., Nelson, W. D., Ploetz, L., Singh, S., Wensel, A. & Huala, E. 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research,* 40**,** D1202-D1210.

Lan, X. J., Wei, Y. M., Liu, D. C., Yan, Z. H. & Zheng, Y. L. 2005. Inheritance of seed dormancy in Tibetan semi-wild wheat accession Q1028. *J Appl Genet,* 46**,** 133-8.

Lazarus, C. M., Baulcombe, D. C. & Martienssen, R. A. 1985. alpha-amylase genes of wheat are two multigene families which are differentially expressed. *Plant Mol Biol,* 5**,** 13-24.

Lee, S., Cheng, H., King, K. E., Wang, W., He, Y., Hussain, A., Lo, J., Harberd, N. P. & Peng, J. 2002. Gibberellin regulates Arabidopsis seed germination via RGL2,

a GAI/RGA-like gene whose expression is up-regulated following imbibition. *Genes Dev,* 16**,** 646-58.

Lei, L., Zhu, X., Wang, S., Zhu, M., Carver, B. F. & Yan, L. 2013. TaMFT-A1 is Associated with Seed Germination Sensitive to Temperature in Winter Wheat. *PLoS ONE,* 8**,** e73330.

Lempe, J., Lachowiec, J., Sullivan, A. M. & Queitsch, C. 2013. Molecular mechanisms of robustness in plants. *Curr Opin Plant Biol,* 16**,** 62-9.

Leon-Kloosterziel, K. M., Van De Bunt, G. A., Zeevaart, J. A. & Koornneef, M. 1996. Arabidopsis mutants with a reduced seed dormancy. *Plant Physiol,* 110**,** 233-40.

Leubner-Metzger, G. 2002. Seed after-ripening and over-expression of class I β-1,3-glucanase confer maternal effects on tobacco testa rupture and dormancy release. *Planta,* 215**,** 959-968.

Leubner-Metzger, G. 2005. β-1,3-Glucanase gene expression in low-hydrated seeds as a mechanism for dormancy release during tobacco after-ripening. *The Plant Journal,* 41**,** 133-145.

Leymarie, J., Vitkauskaite, G., Hoang, H. H., Gendreau, E., Chazoule, V., Meimoun, P., Corbineau, F., El-Maarouf-Bouteau, H. & Bailly, C. 2012. Role of reactive oxygen species in the regulation of Arabidopsis seed dormancy. *Plant Cell Physiol,* 53**,** 96-106.

Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A. & Dewey, C. N. 2010. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics,* 26**,** 493-500.

Li, C., Ni, P., Francki, M., Hunter, A., Zhang, Y., Schibeci, D., Li, H., Tarr, A., Wang, J., Cakir, M., Yu, J., Bellgard, M., Lance, R. & Appels, R. 2004. Genes controlling seed dormancy and pre-harvest sprouting in a rice-wheat-barley comparison. *Functional & Integrative Genomics,* 4**,** 84-93.

Li, G., Yu, M., Fang, T., Cao, S., Carver, B. F. & Yan, L. 2013. Vernalization requirement duration in winter wheat is controlled by TaVRN-A1 at the protein level. *Plant J,* 76**,** 742-53.

Li Y-C, Z. C.-Y., Zhang N, Meng F-R, Ren J-P, Niu H-B, Wang X and Yin J 2012. Cloning of a plasma membrane protein gene TaPM19-1 and its response to abiotic atresses in wheat. *Scienticia Agricultural Sinica,* 4**,** 2502-2509.

Lin, M., Cai, S., Wang, S., Liu, S., Zhang, G. & Bai, G. 2015. Genotyping-by-sequencing (GBS) identified SNP tightly linked to QTL for pre-harvest sprouting resistance. *Theoretical and Applied Genetics,* 128**,** 1385-1395.

Lin, Q., Buckler Iv, E. S., Muse, S. V. & Walker, J. C. 1999. Molecular Evolution of Type 1 Serine/Threonine Protein Phosphatases. *Molecular Phylogenetics and Evolution,* 12**,** 57-66.

Linde-Laursen, I., Heslop-Harrison, J. S., Shepherd, K. W. & Taketa, S. 1997. The barley Genome and its Relationship with the Wheat Genomes. A Survey with an Internationally Agreed Recommendation for Barley Chromosome Nomenclature. *Hereditas,* 126**,** 1-16.

Ling, H.-Q., Zhao, S., Liu, D., Wang, J., Sun, H., Zhang, C., Fan, H., Li, D., Dong, L., Tao, Y., Gao, C., Wu, H., Li, Y., Cui, Y., Guo, X., Zheng, S., Wang, B., Yu, K., Liang, Q., Yang, W., Lou, X., Chen, J., Feng, M., Jian, J., Zhang, X., Luo, G., Jiang, Y., Liu, J., Wang, Z., Sha, Y., Zhang, B., Wu, H., Tang, D., Shen, Q., Xue, P., Zou, S., Wang, X., Liu, X., Wang, F., Yang, Y., An, X., Dong, Z., Zhang, K., Zhang, X., Luo, M.-C., Dvorak, J., Tong, Y., Wang, J., Yang, H., Li, Z., Wang, D., Zhang, A. & Wang, J. 2013. Draft genome of the wheat A-genome progenitor Triticum urartu. *Nature,* 496**,** 87-90.

Linkies, A. & Leubner-Metzger, G. 2012. Beyond gibberellins and abscisic acid: how ethylene and jasmonates control seed germination. *Plant Cell Reports,* 31**,** 253-270.

Linkies, A., Muller, K., Morris, K., Tureckova, V., Wenk, M., Cadman, C. S., Corbineau, F., Strnad, M., Lynn, J. R., Finch-Savage, W. E. & Leubner-Metzger, G. 2009. Ethylene interacts with abscisic acid to regulate endosperm rupture during germination: a comparative approach using Lepidium sativum and Arabidopsis thaliana. *Plant Cell,* 21**,** 3803-22.

Liu, A., Gao, F., Kanno, Y., Jordan, M. C., Kamiya, Y., Seo, M. & Ayele, B. T. 2013a. Regulation of wheat seed dormancy by after-ripening is mediated by specific transcriptional switches that induce changes in seed hormone metabolism and signaling. *PLoS One,* 8**,** e56570.

Liu, C. J., Atkinson, M. D., Chinoy, C. N., Devos, K. M. & Gale, M. D. 1992. Nonhomoeologous translocations between group 4, 5 and 7 chromosomes within wheat and rye. *Theoretical and Applied Genetics,* 83**,** 305-312.

Liu, S., Cai, S., Graybosch, R., Chen, C. & Bai, G. 2008. Quantitative trait loci for resistance to pre-harvest sprouting in US hard white winter wheat Rio Blanco. *Theoretical and Applied Genetics,* 117**,** 691-699.

Liu, S., Sehgal, S. K., Li, J., Lin, M., Trick, H. N., Yu, J., Gill, B. S. & Bai, G. 2013b. Cloning and characterization of a critical regulator for preharvest sprouting in wheat. *Genetics,* 195**,** 263-273.

Liu, X., Zhang, H., Zhao, Y., Feng, Z., Li, Q., Yang, H.-Q., Luan, S., Li, J. & He, Z.-H. 2013c. Auxin controls seed dormancy through stimulation of abscisic acid signaling by inducing ARF-mediated ABI3 activation in Arabidopsis. *Proceedings of the National Academy of Sciences,* 110**,** 15485-15490.

Liu, Y., Geyer, R., Van Zanten, M., Carles, A., Li, Y., Horold, A., Van Nocker, S. & Soppe, W. J. 2011. Identification of the Arabidopsis REDUCED DORMANCY 2 gene uncovers a role for the polymerase associated factor 1 complex in seed dormancy. *PLoS One,* 6**,** e22241.

Liu, Y., Koornneef, M. & Soppe, W. J. J. 2007. The Absence of Histone H2B Monoubiquitination in the Arabidopsis hub1 (rdo4) Mutant Reveals a Role for Chromatin Remodeling in Seed Dormancy. *The Plant Cell,* 19**,** 433-444.

Lohwasser, U., Rehman Arif, M. A. & Börner, A. 2013. Discovery of loci determining pre-harvest sprouting and dormancy in wheat and barley applying segregation and association mapping. *Biologia Plantarum,* 57**,** 663-674.

Lowe, T. M., Ailloud, F. & Allen, C. 2015. Hydroxycinnamic acid degradation, a broadly conserved trait, protects Ralstonia solanacearum from chemical plant defenses and contributes to root colonization and virulence. *Molecular plant-microbe interactions : MPMI,* 28**,** 286-297.

Lowry, D. B. & Willis, J. H. 2010. A Widespread Chromosomal Inversion Polymorphism Contributes to a Major Life-History Transition, Local Adaptation, and Reproductive Isolation. *PLoS Biol,* 8**,** e1000500.

Lucas, S. J., Akpinar, B. A., Kantar, M., Weinstein, Z., Aydinoglu, F., Safar, J., Simkova, H., Frenkel, Z., Korol, A., Magni, F., Cattonaro, F., Vautrin, S., Bellec, A., Berges, H., Dolezel, J. & Budak, H. 2013. Physical mapping integrated with syntenic analysis to characterize the gene space of the long arm of wheat chromosome 1A. *PLoS One,* 8**,** e59542.

Lukowitz, W., Gillmor, C. S. & Scheible, W.-R. 2000. Positional Cloning in Arabidopsis. Why It Feels Good to Have a Genome Initiative Working for You. *Plant Physiology,* 123**,** 795-806.

Luo, M. C., Ma, Y., You, F. M., Anderson, O. D., Kopecky, D., Simkova, H., Safar, J., Dolezel, J., Gill, B., Mcguire, P. E. & Dvorak, J. 2010. Feasibility of physical map construction from fingerprinted bacterial artificial chromosome libraries of polyploid plant species. *BMC Genomics,* 11**,** 122.

Lürssen, K., Naumann, K. & Schröder, R. 1979. 1-Aminocyclopropane-l-carboxylic Acid - An Intermediate of the Ethylene Biosynthesis in Higher Plants. *Zeitschrift für Pflanzenphysiologie,* 92**,** 285-294.

Ma, J., Stiller, J., Berkman, P. J., Wei, Y., Rogers, J., Feuillet, C., Dolezel, J., Mayer, K. F., Eversole, K., Zheng, Y.-L. & Liu, C. 2013. Sequence-Based Analysis of Translocations and Inversions in Bread Wheat (<italic>Triticum aestivum</italic> L.). *PLoS ONE,* 8**,** e79329.

Ma, J., Stiller, J., Wei, Y., Zheng, Y.-L., Devos, K. M., Doležel, J. & Liu, C. 2014. Extensive Pericentric Rearrangements in the Bread Wheat (Triticum aestivum L.) Genotype "Chinese Spring" Revealed from Chromosome Shotgun Sequence Data. *Genome Biology and Evolution,* 6**,** 3039-3048.

Ma, J., Wingen, L. U., Orford, S., Fenwick, P., Wang, J. & Griffiths, S. 2015. Using the UK reference population Avalon × Cadenza as a platform to compare breeding strategies in elite Western European bread wheat. *Molecular Breeding,* 35**,** 70.

Ma, Y., Szostkiewicz, I., Korte, A., Moes, D., Yang, Y., Christmann, A. & Grill, E. 2009. Regulators of PP2C phosphatase activity function as abscisic acid sensors. *Science,* 324**,** 1064-8.

Macgregor, D. R., Kendall, S. L., Florance, H., Fedi, F., Moore, K., Paszkiewicz, K., Smirnoff, N. & Penfield, S. 2015. Seed production temperature regulation of primary dormancy occurs through control of seed coat phenylpropanoid metabolism. *New Phytologist,* 205**,** 642-652.

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C. & Bryant, S. H. 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Res,* 43**,** D222-6.

Mares, D. & Mrva, K. 2008. Late-maturity α-amylase: Low falling number in wheat in the absence of preharvest sprouting. *Journal of Cereal Science,* 47**,** 6-17.

Mares, D. & Mrva, K. 2014. Wheat grain preharvest sprouting and late maturity alpha-amylase. *Planta,* 240**,** 1167-1178.

Mares, D., Mrva, K., Cheong, J., Williams, K., Watson, B., Storlie, E., Sutherland, M. & Zou, Y. 2005. A QTL located on chromosome 4A associated with dormancy in white- and red-grained wheats of diverse origin. *Theoretical and Applied Genetics,* 111**,** 1357-1364.

Mares, D., Rathjen, J., Mrva, K. & Cheong, J. 2009. Genetic and environmental control of dormancy in white-grained wheat (Triticum aestivum L.). *Euphytica,* 168**,** 311-318.

Martínez-Andújar, C., Ordiz, M. I., Huang, Z., Nonogaki, M., Beachy, R. N. & Nonogaki, H. 2011. Induction of 9-cis-epoxycarotenoid dioxygenase in Arabidopsis thaliana seeds enhances seed dormancy. *Proceedings of the National Academy of Sciences,* 108**,** 17225-17229.

Mascher, M., Jost, M., Kuon, J. E., Himmelbach, A., Assfalg, A., Beier, S., Scholz, U., Graner, A. & Stein, N. 2014. Mapping-by-sequencing accelerates forward genetics in barley. *Genome Biol,* 15**,** R78.

Matilla, A. J. & Matilla-Vázquez, M. A. 2008. Involvement of ethylene in seed physiology. *Plant Science,* 175**,** 87-97.

Matsuoka, Y. 2011. Evolution of Polyploid Triticum Wheats under Cultivation: The Role of Domestication, Natural Hybridization and Allopolyploid Speciation in their Diversification. *Plant and Cell Physiology,* 52**,** 750-764.

Mccarty, D. R., Carson, C. B., Stinard, P. S. & Robertson, D. S. 1989. Molecular Analysis of viviparous-1: An Abscisic Acid-Insensitive Mutant of Maize. *The Plant Cell,* 1**,** 523-532.

Mckibbin, R. S., Wilkinson, M. D., Bailey, P. C., Flintham, J. E., Andrew, L. M., Lazzeri, P. A., Gale, M. D., Lenton, J. R. & Holdsworth, M. J. 2002. Transcripts of Vp-1 homeologues are misspliced in modern wheat and ancestral species. *Proc Natl Acad Sci U S A,* 99**,** 10203-8.

Meimberg, H., Rice, K. J., Milan, N. F., Njoku, C. C. & Mckay, J. K. 2009. Multiple origins promote the ecological amplitude of allopolyploid Aegilops (Poaceae). *American Journal of Botany,* 96**,** 1262-1273.

Meimoun, P., Mordret, E., Langlade, N. B., Balzergue, S., Arribat, S., Bailly, C. & El-Maarouf-Bouteau, H. 2014. Is Gene Transcription Involved in Seed Dry After-Ripening? *PLoS ONE,* 9**,** e86442.

Merlot, S., Gosti, F., Guerrier, D., Vavasseur, A. & Giraudat, J. 2001. The ABI1 and ABI2 protein phosphatases 2C act in a negative feedback regulatory loop of the abscisic acid signalling pathway. *Plant J,* 25**,** 295-303.

Met Office. 2015. *Climate Summaries, UK actual and anomaly maps* [Online]. Available: http://www.metoffice.gov.uk/climate/uk/summaries/anomacts 21/09/2015].

Meyers, B. C., Scalabrin, S. & Morgante, M. 2004. Mapping and sequencing complex genomes: let's get physical! *Nat Rev Genet,* 5**,** 578-588.

Mezard, C. 2006. Meiotic recombination hotspots in plants. *Biochem Soc Trans,* 34**,** 531-4.

Miftahudin, Ross, K., Ma, X. F., Mahmoud, A. A., Layton, J., Milla, M. a. R., Chikmawati, T., Ramalingam, J., Feril, O., Pathan, M. S., Momirovic, G. S., Kim, S., Chema, K., Fang, P., Haule, L., Struxness, H., Birkes, J., Yaghoubian, C., Skinner, R., Mcallister, J., Nguyen, V., Qi, L. L., Echalier, B., Gill, B. S., Linkiewicz, A. M., Dubcovsky, J., Akhunov, E. D., Dvořák, J., Dilbirligi, M., Gill, K. S., Peng, J. H., Lapitan, N. L. V., Bermudez-Kandianis, C. E., Sorrells, M. E., Hossain, K. G., Kalavacharla, V., Kianian, S. F., Lazo, G. R., Chao, S., Anderson, O. D., Gonzalez-Hernandez, J., Conley, E. J., Anderson, J. A., Choi, D. W., Fenton, R. D., Close, T. J., Mcguire, P. E., Qualset, C. O., Nguyen, H. T. & Gustafson, J. P. 2004. Analysis of Expressed Sequence Tag Loci on Wheat Chromosome Group 4. *Genetics,* 168**,** 651-663.

Miko, I. 2008. Phenotype variability: penetrance and expressivity. *Nature Education,* 1**,** 137.

Millar, A. A., Jacobsen, J. V., Ross, J. J., Helliwell, C. A., Poole, A. T., Scofield, G., Reid, J. B. & Gubler, F. 2006. Seed dormancy and ABA metabolism in Arabidopsis and barley: the role of ABA 8'-hydroxylase. *Plant Journal,* 45**,** 942-954.

Miransari, M. & Smith, D. L. 2014. Plant hormones and seed germination. *Environmental and Experimental Botany,* 99**,** 110-121.

Mitchum, M. G., Yamaguchi, S., Hanada, A., Kuwahara, A., Yoshioka, Y., Kato, T., Tabata, S., Kamiya, Y. & Sun, T. P. 2006. Distinct and overlapping roles of two gibberellin 3-oxidases in Arabidopsis development. *Plant J,* 45**,** 804-18.

Miyoshi, K. & Sato, T. 1997. The Effects of Ethanol on the Germination of Seeds of Japonica and Indica Rice (Oryza sativa L.) under Anaerobic and Aerobic Conditions. *Annals of Botany,* 79**,** 391-395.

Mohan, A., Kulwal, P., Singh, R., Kumar, V., Mir, R., Kumar, J., Prasad, M., Balyan, H. S. & Gupta, P. K. 2009. Genome-wide QTL analysis for pre-harvest sprouting tolerance in bread wheat. *Euphytica,* 168**,** 319-329.

Mori, M., Uchino, N., Chono, M., Kato, K. & Miura, H. 2005. Mapping QTLs for grain dormancy on wheat chromosome 3A and the group 4 chromosomes, and their combined effect. *Theoretical and Applied Genetics,* 110**,** 1315-1323.

Mrva, K., Wallwork, M. & Mares, D. J. 2006. alpha-Amylase and programmed cell death in aleurone of ripening wheat grains. *J Exp Bot,* 57**,** 877-85.

Munkvold, J., Tanaka, J., Benscher, D. & Sorrells, M. 2009. Mapping quantitative trait loci for preharvest sprouting resistance in white wheat. *Theoretical and Applied Genetics,* 119**,** 1223-1235.

Muroi, A., Matsui, K., Shimoda, T., Kihara, H., Ozawa, R., Ishihara, A., Nishihara, M. & Arimura, G.-I. 2012. Acquired immunity of transgenic torenia plants overexpressing agmatine coumaroyltransferase to pathogens and herbivore pests. *Scientific Reports,* 2**,** 689.

Nakabayashi, K., Bartsch, M., Xiang, Y., Miatton, E., Pellengahr, S., Yano, R., Seo, M. & Soppe, W. J. J. 2012. The Time Required for Dormancy Release in Arabidopsis Is Determined by DELAY OF GERMINATION1 Protein Levels in Freshly Harvested Seeds. *The Plant Cell,* 24**,** 2826-2838.

Nakamura, S., Abe, F., Kawahigashi, H., Nakazono, K., Tagiri, A., Matsumoto, T., Utsugi, S., Ogawa, T., Handa, H., Ishida, H., Mori, M., Kawaura, K., Ogihara, Y. & Miura, H. 2011a. A Wheat Homolog of MOTHER OF FT AND TFL1 Acts in the Regulation of Germination. *The Plant Cell,* 23**,** 3215-3229.

Nakamura, S., Abe, F., Kawahigashi, H., Nakazono, K., Tagiri, A., Matsumoto, T., Utsugi, S., Ogawa, T., Handa, H., Ishida, H., Mori, M., Kawaura, K., Ogihara, Y. & Miura, H. 2011b. A wheat homolog of MOTHER OF FT AND TFL1 acts in the regulation of germination. *Plant Cell,* 23**,** 3215-29.

Nelson, W. & Soderlund, C. 2009. Integrating sequence with FPC fingerprint maps. *Nucleic Acids Research,* 37**,** e36.

Nevo, E., Korol, A.B., Beiles, A., Tzion, F. 2002. *Evolution of Wild Emmer and Wheat Improvement,* Heidelberg, Springer.

Nikolaeva, M. G. 2004. On criteria to use in studies of seed evolution. *Seed Science Research,* 14**,** 315-320.

Noll, J. S., Dyck, P. L. & Czarnecki, E. 1982. EXPRESSION OF RL 4137 TYPE OF DORMANCY IN F1 SEEDS OF RECIPROCAL CROSSES IN COMMON WHEAT. *Canadian Journal of Plant Science,* 62**,** 345-349.

Nomura, T., Ishizuka, A., Kishida, K., Islam, A. K., Endo, T. R., Iwamura, H. & Ishihara, A. 2007. Chromosome arm location of the genes for the biosynthesis of hordatines in barley. *Genes Genet Syst,* 82**,** 455-64.

Nyachiro, J. M., Clarke, F. R., Depauw, R. M., Knox, R. E. & Armstrong, K. C. 2002. Temperature effects on seed germination and expression of seed dormancy in wheat. *Euphytica,* 126**,** 123-127.

Ogbonnaya, F. C., Imtiaz, M. & Depauw, R. M. 2007. Haplotype diversity of preharvest sprouting QTLs in wheat. *Genome,* 50**,** 107-118.

Ogbonnaya, F. C., Imtiaz, M., Ye, G., Hearnden, P. R., Hernandez, E., Eastwood, R. F., Van Ginkel, M., Shorter, S. C. & Winchester, J. M. 2008. Genetic and QTL

analyses of seed dormancy and preharvest sprouting resistance in the wheat germplasm CN10955. *Theoretical and Applied Genetics,* 116**,** 891-902.

Oh, S.-J., Song, S. I., Kim, Y. S., Jang, H.-J., Kim, S. Y., Kim, M., Kim, Y.-K., Nahm, B. H. & Kim, J.-K. 2005. Arabidopsis CBF3/DREB1A and ABF3 in Transgenic Rice Increased Tolerance to Abiotic Stress without Stunting Growth. *Plant Physiology,* 138**,** 341-351.

Olesen, J. E., Trnka, M., Kersebaum, K. C., Skjelvåg, A. O., Seguin, B., Peltonen-Sainio, P., Rossi, F., Kozyra, J. & Micale, F. 2011. Impacts and adaptation of European crop production systems to climate change. *European Journal of Agronomy,* 34**,** 96-112.

Olsen, O.-A. 2004. Nuclear Endosperm Development in Cereals and Arabidopsis thaliana. *The Plant Cell,* 16**,** S214-S227.

Olsen, O. A. 2001. ENDOSPERM DEVELOPMENT: Cellularization and Cell Fate Specification. *Annu Rev Plant Physiol Plant Mol Biol,* 52**,** 233-267.

Oracz, K., Bouteau, H. E.-M., Farrant, J. M., Cooper, K., Belghazi, M., Job, C., Job, D., Corbineau, F. & Bailly, C. 2007. ROS production and protein oxidation as a novel mechanism for seed dormancy alleviation. *The Plant Journal,* 50**,** 452-465.

Osa, M., Kato, K., Mori, M., Shindo, C., Torada, A. & Miura, H. 2003. Mapping QTLs for seed dormancy and the Vp1 homologue on chromosome 3A in wheat. *Theor Appl Genet,* 106**,** 1491-6.

Ouaked, F., Rozhon, W., Lecourieux, D. & Hirt, H. 2003. A MAPK pathway mediates ethylene signaling in plants. *The EMBO Journal,* 22**,** 1282-1288.

Pallotta, M. A., Warner, P., Fox, R. L., Kuchel, H., Jefferies, S. J. & Langridge, P. Marker assisted wheat breeding in the southern region of Australia. *In:* ROMANO, M., POGNA, E. A. & GALTERIO, Z., eds. 10th International Wheat Genetics Symposium, 2003 Paestum. pp 789–791.

Paux, E., Sourdille, P., Salse, J., Saintenac, C., Choulet, F., Leroy, P., Korol, A., Michalak, M., Kianian, S., Spielmeyer, W., Lagudah, E., Somers, D., Kilian, A., Alaux, M., Vautrin, S., Bergès, H., Eversole, K., Appels, R., Safar, J., Simkova, H., Dolezel, J., Bernard, M. & Feuillet, C. 2008. A Physical Map of the 1-Gigabase Bread Wheat Chromosome 3B. *Science,* 322**,** 101-104.

Peeters, A. J., Blankestijn-De Vries, H., Hanhart, C. J., Leon-Kloosterziel, K. M., Zeevaart, J. A. & Koornneef, M. 2002. Characterization of mutants with reduced seed dormancy at two novel rdo loci and a further characterization of rdo1 and rdo2 in Arabidopsis. *Physiol Plant,* 115**,** 604-612.

Penfield, S., Josse, E. M., Kannangara, R., Gilday, A. D., Halliday, K. J. & Graham, I. A. 2005. Cold and light control seed germination through the bHLH transcription factor SPATULA. *Curr Biol,* 15**,** 1998-2006.

Peng, J. S., Dongfa; Nevo, Eviatar; 2011. Wild emmer wheat, Triticum dicoccoides, occupies a pivotal position in wheat domestication process. *Australian Journal of Crop Science,* Vol. 5**,** 1127-1143.

Periyannan, S., Moore, J., Ayliffe, M., Bansal, U., Wang, X., Huang, L., Deal, K., Luo, M., Kong, X., Bariana, H., Mago, R., Mcintosh, R., Dodds, P., Dvorak, J. & Lagudah, E. 2013. The Gene Sr33, an Ortholog of Barley Mla Genes, Encodes Resistance to Wheat Stem Rust Race Ug99. *Science,* 341**,** 786-788.

Perten, H. 1964. Application of the Falling Number Method for Evaluating Alpha-Amylase Activity. *Cereal Chemistry,* 41**,** 127-139.

Philippe, R., Paux, E., Bertin, I., Sourdille, P., Choulet, F., Laugier, C., Simkova, H., Safar, J., Bellec, A., Vautrin, S., Frenkel, Z., Cattonaro, F., Magni, F., Scalabrin,

S., Martis, M. M., Mayer, K. F., Korol, A., Berges, H., Dolezel, J. & Feuillet, C. 2013. A high density physical map of chromosome 1BL supports evolutionary studies, map-based cloning and sequencing in wheat. *Genome Biol,* 14**,** R64.

Poland, J. A. & Rife, T. W. 2012. Genotyping-by-Sequencing for Plant Breeding and Genetics. *The Plant Genome,* 5**,** 92-102.

Poursarebani, N., Nussbaumer, T., Simkova, H., Safar, J., Witsenboer, H., Van Oeveren, J., Dolezel, J., Mayer, K. F., Stein, N. & Schnurbusch, T. 2014. Whole-genome profiling and shotgun sequencing delivers an anchored, gene-decorated, physical map assembly of bread wheat chromosome 6A. *Plant J,* 79**,** 334-47.

Pretty, J. N., Ball, A. S., Lang, T. & Morison, J. I. L. 2005. Farm costs and food miles: An assessment of the full cost of the UK weekly food basket. *Food Policy,* 30**,** 1-19.

Raats, D., Frenkel, Z., Krugman, T., Dodek, I., Sela, H., Simkova, H., Magni, F., Cattonaro, F., Vautrin, S., Berges, H., Wicker, T., Keller, B., Leroy, P., Philippe, R., Paux, E., Dolezel, J., Feuillet, C., Korol, A. & Fahima, T. 2013. The physical map of wheat chromosome 1BS provides insights into its gene space organization and evolution. *Genome Biol,* 14**,** R138.

Radchuk, V. & Borisjuk, L. 2014. Physical, metabolic and developmental functions of the seed coat. *Frontiers in Plant Science,* 5**,** 510.

Rafalski, A. 2002. Applications of single nucleotide polymorphisms in crop genetics. *Curr Opin Plant Biol,* 5**,** 94-100.

Ral, J.-P., Whan, A., Larroque, O., Leyne, E., Pritchard, J., Dielen, A.-S., Howitt, C. A., Morell, M. K. & Newberry, M. 2015. Engineering high α-amylase levels in wheat grain lowers Falling Number but improves baking properties. *Plant Biotechnology Journal***,** n/a-n/a.

Rancurel, C., Khosravi, M., Dunker, A. K., Romero, P. R. & Karlin, D. 2009. Overlapping Genes Produce Proteins with Unusual Sequence Properties and Offer Insight into De Novo Protein Creation. *Journal of Virology,* 83**,** 10719-10736.

Ranford, J. C., Bryce, J. H. & Morris, P. C. 2002. PM19, a barley (Hordeum vulgare L.) gene encoding a putative plasma membrane protein, is expressed during embryo development and dormancy. *Journal of Experimental Botany,* 53**,** 147-148.

Rathjen, J. R., Strounina, E. V. & Mares, D. J. 2009. Water movement into dormant and non-dormant wheat (Triticum aestivum L.) grains. *Journal of Experimental Botany,* 60**,** 1619-1631.

Reddy, L. V., Metzger, R. J. & Ching, T. M. 1985. Effect of Temperature on Seed Dormancy of Wheat. *Crop Science,* 25**,** 455-458.

Rehman Arif, M. A., Neumann, K., Nagel, M., Kobiljski, B., Lohwasser, U. & Börner, A. 2012. An association mapping analysis of dormancy and pre-harvest sprouting in wheat. *Euphytica,* 188**,** 409-417.

Ren, X. B., Lan, X. J., Liu, D. C., Wang, J. L. & Zheng, Y. L. 2008. Mapping QTLs for pre-harvest sprouting tolerance on chromosome 2D in a synthetic hexaploid wheat x common wheat cross. *J Appl Genet,* 49**,** 333-41.

Rikiishi, K., Matsuura, T. & Maekawa, M. 2010. TaABF1, ABA response element binding factor 1, is related to seed dormancy and ABA sensitivity in wheat (Triticum aestivum L.) seeds. *Journal of Cereal Science,* 52**,** 236-238.

Rodríguez, M. V., Barrero, J. M., Corbineau, F., Gubler, F. & Benech-Arnold, R. L. 2015. Dormancy in cereals (not too much, not so little): about the mechanisms behind this trait. *Seed Science Research,* 25**,** 99-119.

Rolston, M. P. 1978. Water impermeable seed dormancy. *The Botanical Review,* 44**,** 365-396.

Ross, J. S. 2011. Next-Generation Pathology. *American Journal of Clinical Pathology,* 135**,** 663-665.

Rustenholz, C., Choulet, F., Laugier, C., Šafář, J., Šimková, H., Doležel, J., Magni, F., Scalabrin, S., Cattonaro, F., Vautrin, S., Bellec, A., Bergès, H., Feuillet, C. & Paux, E. 2011. A 3,000-Loci Transcription Map of Chromosome 3B Unravels the Structural and Functional Features of Gene Islands in Hexaploid Wheat. *Plant Physiology,* 157**,** 1596-1608.

Šafář, J., Bartoš, J., Janda, J., Bellec, A., Kubaláková, M., Valárik, M., Pateyron, S., Weiserová, J., Tušková, R., Číhalíková, J., Vrána, J., Šimková, H., Faivre-Rampant, P., Sourdille, P., Caboche, M., Bernard, M., Doležel, J. & Chalhoub, B. 2004. Dissecting large and complex genomes: flow sorting and BAC cloning of individual chromosomes from bread wheat. *The Plant Journal,* 39**,** 960-968.

Saintenac, C., Falque, M., Martin, O. C., Paux, E., Feuillet, C. & Sourdille, P. 2009. Detailed Recombination Studies Along Chromosome 3B Provide New Insights on Crossover Distribution in Wheat (Triticum aestivum L.). *Genetics,* 181**,** 393-403.

Saintenac, C., Jiang, D., Wang, S. & Akhunov, E. 2013a. Sequence-based mapping of the polyploid wheat genome. *G3 (Bethesda),* 3**,** 1105-14.

Saintenac, C., Zhang, W., Salcedo, A., Rouse, M. N., Trick, H. N., Akhunov, E. & Dubcovsky, J. 2013b. Identification of Wheat Gene Sr35 That Confers Resistance to Ug99 Stem Rust Race Group. *Science,* 341**,** 783-786.

Saxena, R. K., Edwards, D. & Varshney, R. K. 2014. Structural variations in plant genomes. *Briefings in Functional Genomics,* 13**,** 296-307.

Schmittgen, T. D. & Livak, K. J. 2008. Analyzing real-time PCR data by the comparative CT method. *Nat. Protocols,* 3**,** 1101-1108.

Schramm, E. C., Nelson, S. K. & Steber, C. M. 2012. Wheat ABA-insensitive mutants result in reduced grain dormancy. 188**,** 35-49.

Sears, E. R. & Sears, L. M. The telocentric chromosomes of common wheat. Proceedings of the 5th International Wheat Genetics Symposium, 1978. Indian Society for Genetics and Plant Breeding New Delhi, 389-407.

Shewry, P. R. 2009. Wheat. *Journal of Experimental Botany,* 60**,** 1537-1553.

Shewry, P. R., Mitchell, R. a. C., Tosi, P., Wan, Y., Underwood, C., Lovegrove, A., Freeman, J., Toole, G. A., Mills, E. N. C. & Ward, J. L. 2012. An integrated study of grain development of wheat (cv. Hereward). *Journal of Cereal Science,* 56**,** 21-30.

Shewry, P. R., Underwood, C., Wan, Y., Lovegrove, A., Bhandari, D., Toole, G., Mills, E. N. C., Denyer, K. & Mitchell, R. a. C. 2009. Storage product synthesis and accumulation in developing grains of wheat. *Journal of Cereal Science,* 50**,** 106-112.

Shimmen, T. 2007. The sliding theory of cytoplasmic streaming: fifty years of progress. *J Plant Res,* 120**,** 31-43.

Sidhu, D. & Gill, K. 2005. Distribution of genes and recombination in wheat and other eukaryotes. *Plant Cell, Tissue and Organ Culture,* 79**,** 257-270.

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. & Birol, I. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res,* 19**,** 1117-23.

Singh, A. K., Knox, R. E., Clarke, J. M., Clarke, F. R., Singh, A., Depauw, R. M. & Cuthbert, R. D. 2014. Genetics of pre-harvest sprouting resistance in a cross of Canadian adapted durum wheat genotypes. *Molecular Breeding,* 33**,** 919-929.

Singh, R., Matus-Cádiz, M., Båga, M., Hucl, P. & Chibbar, R. 2010. Identification of genomic regions associated with seed dormancy in white-grained wheat. *Euphytica,* 174**,** 391-408.

Slageren, M. W. V. 1994. *Wild wheats: a monograph of Aegilops L. and Amblyopyrum (Jaub. & Spach) Eig (Poaceae).* Aleppo, Syria, Wageningen Agricultural University Papers, Wageningen & ICARDA.

Smith, B. D. 1998. *The Emergence of Agriculture,* New York, Scientific American Library

Smith, D. B. & Flavell, R. B. 1975. Characterisation of the wheat genome by renaturation kinetics. *Chromosoma,* 50**,** 223-242.

Smith, S. M. & Maughan, P. J. 2015. SNP genotyping using KASPar assays. *Methods Mol Biol,* 1245**,** 243-56.

Soleimani, V. D., Baum, B. R. & Johnson, D. A. 2003. Efficient validation of single nucleotide polymorphisms in plants by allele-specific PCR, with an example from barley. *Plant Molecular Biology Reporter,* 21**,** 281-288.

Solovyev, V., Kosarev, P., Seledsov, I. & Vorobyev, D. 2006. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol,* 7 Suppl 1**,** S10 1-12.

Somers, D., Isaac, P. & Edwards, K. 2004. A high-density microsatellite consensus map for bread wheat (Triticum aestivum L.). *Theoretical and Applied Genetics,* 109**,** 1105-1114.

Sreenivasulu, N., Radchuk, V., Strickert, M., Miersch, O., Weschke, W. & Wobus, U. 2006. Gene expression patterns reveal tissue-specific signaling networks controlling programmed cell death and ABA- regulated maturation in developing barley seeds. *The Plant Journal,* 47**,** 310-327.

Srinivasachary, Gosman, N., Steed, A., Simmonds, J., Leverington-Waite, M., Wang, Y., Snape, J. & Nicholson, P. 2008. Susceptibility to Fusarium head blight is associated with the Rht-D1b semi-dwarfing allele in wheat. *Theor Appl Genet,* 116**,** 1145-53.

Steber, C. M. & Mccourt, P. 2001. A Role for Brassinosteroids in Germination in Arabidopsis. *Plant Physiology,* 125**,** 763-769.

Sugimoto, K., Takeuchi, Y., Ebana, K., Miyao, A., Hirochika, H., Hara, N., Ishiyama, K., Kobayashi, M., Ban, Y., Hattori, T. & Yano, M. 2010. Molecular cloning of Sdr4, a regulator involved in seed dormancy and domestication of rice. *Proceedings of the National Academy of Sciences,* 107**,** 5792-5797.

Sullivan, M. L., Carpenter, T. B. & Vierstra, R. D. 1994. Homologues of wheat ubiquitin-conjugating enzymes--TaUBC1 and TaUBC4 are encoded by small multigene families in Arabidopsis thaliana. *Plant Mol Biol,* 24**,** 651-61.

Suzuki, H., Nakayama, T., Yonekura-Sakakibara, K., Fukui, Y., Nakamura, N., Yamaguchi, M.-A., Tanaka, Y., Kusumi, T. & Nishino, T. 2002. cDNA Cloning, Heterologous Expressions, and Functional Characterization of Malonyl-Coenzyme A:Anthocyanidin 3-O-Glucoside-6"-O-Malonyltransferase from Dahlia Flowers. *Plant Physiology,* 130**,** 2142-2151.

Swan, M. 2010. Multigenic condition risk assessment in direct-to-consumer genomic services. *Genet Med,* 12**,** 279-288.

Tagi, T. a. G. I. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature,* 408**,** 796-815.

Takahashi, F., Yoshida, R., Ichimura, K., Mizoguchi, T., Seo, S., Yonezawa, M., Maruyama, K., Yamaguchi-Shinozaki, K. & Shinozaki, K. 2007. The Mitogen-Activated Protein Kinase Cascade MKK3–MPK6 Is an Important Part of the Jasmonate Signal Transduction Pathway in Arabidopsis. *The Plant Cell,* 19**,** 805-818.

Talbert, L. E., Smith, L. Y. & Blake, N. K. 1998. More than one origin of hexaploid wheat is indicated by sequence comparison of low-copy DNA. *Genome,* 41**,** 402-407.

Tanksley, S. D., Ganal, M. W. & Martin, G. B. 1995. Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. *Trends in Genetics,* 11**,** 63-68.

Teo, Y.-Y., Ong, R. T. H., Sim, X., Tai, E. S. & Chia, K.-S. 2010. Identifying candidate causal variants via trans-population fine-mapping. *Genetic Epidemiology,* 34**,** 653-664.

The International Brachypodium Initiative 2010. Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature,* 463**,** 763-768.

Torada, A., Ikeguchi, S. & Koike, M. 2005. Mapping and validation of PCR-based markers associated with a major QTL for seed dormancy in wheat. *Euphytica,* 143**,** 251-255.

Torada, A., Koike, M., Ikeguchi, S. & Tsutsui, I. 2008. Mapping of a major locus controlling seed dormancy using backcrossed progenies in wheat (Triticum aestivum L.). *Genome,* 51**,** 426-432.

Torii, K. U. 2004. Leucine-Rich Repeat Receptor Kinases in Plants: Structure, Function, and Signal Transduction Pathways. *International Review of Cytology.* Academic Press.

Trick, M., Adamski, N., Mugford, S., Jiang, C., Febrer, M. & Uauy, C. 2012. Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. *BMC Plant Biol,* 12**,** 14.

Tukey, J. W. 1949. Comparing Individual Means in the Analysis of Variance. *Biometrics,* 5**,** 99-114.

Tuteja, N. 2007. Abscisic Acid and Abiotic Stress Signaling. *Plant Signaling & Behavior,* 2**,** 135-138.

Tzarfati, R., Saranga, Y., Barak, V., Gopher, A., Korol, A. B. & Abbo, S. 2013. Threshing efficiency as an incentive for rapid domestication of emmer wheat. *Annals of Botany,* 112**,** 829-837.

Uauy, C., Paraiso, F., Colasuonno, P., Tran, R., Tsai, H., Berardi, S., Comai, L. & Dubcovsky, J. 2009. A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. *BMC Plant Biol,* 9**,** 115.

Ueno, K. & Miyoshi, K. 2005. Difference of optimum germination temperature of seeds of intact and dehusked japonica rice during seed development. *Euphytica,* 143**,** 271-275.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M. & Rozen, S. G. 2012. Primer3--new capabilities and interfaces. *Nucleic Acids Res,* 40**,** e115.

Vain, P. 2011. Brachypodium as a model system for grass research. *Journal of Cereal Science,* 54**,** 1-7.

Van Wijk, S. J. & Timmers, H. T. 2010. The family of ubiquitin-conjugating enzymes (E2s): deciding between life and death of proteins. *FASEB J,* 24**,** 981-93.

Veasey, E. A., Karasawa, M. G., Santos, P. P., Rosa, M. S., Mamani, E. & Oliveira, G. C. 2004. Variation in the loss of seed dormancy during after-ripening of wild and cultivated rice species. *Ann Bot,* 94**,** 875-82.

Walker-Simmons, M. 1987. ABA Levels and Sensitivity in Developing Wheat Embryos of Sprouting Resistant and Susceptible Cultivars. *Plant Physiology,* 84**,** 61-66.

Wan, Y., Poole, R. L., Huttly, A. K., Toscano-Underwood, C., Feeney, K., Welham, S., Gooding, M. J., Mills, C., Edwards, K. J., Shewry, P. R. & Mitchell, R. A. 2008. Transcriptome analysis of grain development in hexaploid wheat. *BMC Genomics,* 9**,** 121.

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., Maccaferri, M., Salvi, S., Milner, S. G., Cattivelli, L., Mastrangelo, A. M., Whan, A., Stephen, S., Barker, G., Wieseke, R., Plieske, J., International Wheat Genome Sequencing, C., Lillemo, M., Mather, D., Appels, R., Dolferus, R., Brown-Guedira, G., Korol, A., Akhunova, A. R., Feuillet, C., Salse, J., Morgante, M., Pozniak, C., Luo, M.-C., Dvorak, J., Morell, M., Dubcovsky, J., Ganal, M., Tuberosa, R., Lawley, C., Mikoulitch, I., Cavanagh, C., Edwards, K. J., Hayden, M. & Akhunov, E. 2014. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnology Journal,* 12**,** 787-796.

Wang, Y., Li, L., Ye, T., Zhao, S., Liu, Z., Feng, Y. Q. & Wu, Y. 2011. Cytokinin antagonizes ABA suppression to seed germination of Arabidopsis by downregulating ABI5 expression. *Plant J,* 68**,** 249-61.

Wang, Z., Cao, H., Sun, Y., Li, X., Chen, F., Carles, A., Li, Y., Ding, M., Zhang, C., Deng, X., Soppe, W. J. & Liu, Y. X. 2013. Arabidopsis paired amphipathic helix proteins SNL1 and SNL2 redundantly regulate primary seed dormancy via abscisic acid-ethylene antagonism mediated by histone deacetylation. *Plant Cell,* 25**,** 149-66.

Watanabe, N., Fujii, Y., Kato, N., Ban, T. & Martinek, P. 2006. Microsatellite mapping of the genes for brittle rachis on homoeologous group 3 chromosomes in tetraploid and hexaploid wheats. *J Appl Genet,* 47**,** 93-8.

Wicker, T., Mayer, K. F., Gundlach, H., Martis, M., Steuernagel, B., Scholz, U., Simkova, H., Kubalakova, M., Choulet, F., Taudien, S., Platzer, M., Feuillet, C., Fahima, T., Budak, H., Dolezel, J., Keller, B. & Stein, N. 2011. Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. *Plant Cell,* 23**,** 1706-18.

Willige, B. C., Ghosh, S., Nill, C., Zourelidou, M., Dohmann, E. M. N., Maier, A. & Schwechheimer, C. 2007. The DELLA Domain of GA INSENSITIVE Mediates the Interaction with the GA INSENSITIVE DWARF1A Gibberellin Receptor of Arabidopsis. *The Plant Cell,* 19**,** 1209-1220.

Winfield, M. O., Wilkinson, P. A., Allen, A. M., Barker, G. L. A., Coghill, J. A., Burridge, A., Hall, A., Brenchley, R. C., D'amore, R., Hall, N., Bevan, M. W., Richmond, T., Gerhardt, D. J., Jeddeloh, J. A. & Edwards, K. J. 2012. Targeted re-sequencing of the allohexaploid wheat exome. *Plant Biotechnology Journal,* 10**,** 733-742.

Worland, A. 1996. The influence of flowering time genes on environmental adaptability in European wheats. *Euphytica,* 89**,** 49-57.

Wu, J., Mizuno, H., Hayashi-Tsugane, M., Ito, Y., Chiden, Y., Fujisawa, M., Katagiri, S., Saji, S., Yoshiki, S., Karasawa, W., Yoshihara, R., Hayashi, A., Kobayashi, H., Ito, K., Hamada, M., Okamoto, M., Ikeno, M., Ichikawa, Y., Katayose, Y.,

Yano, M., Matsumoto, T. & Sasaki, T. 2003. Physical maps and recombination frequency of six rice chromosomes. *Plant J,* 36**,** 720-30.
Xiao-Bo, R., Xiu-Jin, L., Deng-Cai, L., Jia-Li, W. & You-Liang, Z. 2008. Mapping QTLs for pre-harvest sprouting tolerance on chromosome 2D in a synthetic hexaploid wheat×common wheat cross. *Journal of Applied Genetics,* 49**,** 333-341.
Xu, X., Hsia, A. P., Zhang, L., Nikolau, B. J. & Schnable, P. S. 1995. Meiotic recombination break points resolve at high rates at the 5' end of a maize coding sequence. *Plant Cell,* 7**,** 2151-61.
Yamauchi, Y., Ogawa, M., Kuwahara, A., Hanada, A., Kamiya, Y. & Yamaguchi, S. 2004. Activation of Gibberellin Biosynthesis and Response Pathways by Low Temperature during Imbibition of Arabidopsis thaliana Seeds. *The Plant Cell,* 16**,** 367-378.
Yao, H. & Schnable, P. S. 2005. Cis-effects on Meiotic Recombination Across Distinct a1-sh2 Intervals in a Common Zea Genetic Background. *Genetics,* 170**,** 1929-1944.
Yoo, S.-D., Cho, Y.-H., Tena, G., Xiong, Y. & Sheen, J. 2008. Dual control of nuclear EIN3 by bifurcate MAPK cascades in C(2)H(4) signalling. *Nature,* 451**,** 789-795.
Yoo, S.-D. & Sheen, J. 2008. MAPK signaling in plant hormone ethylene signal transduction. *Plant Signaling & Behavior,* 3**,** 848-849.
Yoon, G. M. & Kieber, J. J. 2013. 1-Aminocyclopropane-1-carboxylic acid as a signalling molecule in plants. *AoB Plants,* 5.
Zadoks, J. C., Chang T.T. And Konzak C. F. 1974. A decimal code for the growth stages of cereals. *Weed Research,* 14**,** 415-421.
Zhang, C., Gao, L., Sun, J., Jia, J. & Ren, Z. 2014. Haplotype variation of Green Revolution gene Rht-D1 during wheat domestication and improvement. *J Integr Plant Biol,* 56**,** 774-80.
Zhang, X.-Q., Li, C., Tay, A., Lance, R., Mares, D., Cheong, J., Cakir, M., Ma, J. & Appels, R. 2008. A new PCR-based marker on chromosome 4AL for resistance to pre-harvest sprouting in wheat (Triticum aestivum L.). *Molecular Breeding,* 22**,** 227-236.
Zohary, D. 1999. Monophyletic vs. polyphyletic origin of the crops on which agriculture was founded in the Near East. *Genetic Resources and Crop Evolution,* 46**,** 133-142.
Zohary D, H. M. 2000. *Domestication of plants in the old world,* Oxford, Oxford University Press.

8 Appendix

Appendix 1: Primers of the SNP markers designed from genes in the Phs physical map

Gene	Primer Name	Primer (probe) Sequence (5'-3')
Agmatine Coumarorl transferase	Agmatime_SNP1_FAM	GAAGGTGACCAAGTTCATGCTTTGGAAATTGCAAATTGAG
	Agmatine_SNP1_VIC	GAAGGTCGGAGTCAACGGATTTTGGAAATTGCAAATTGAC
	Agmatine_SNP1_COM	ACTTGCTTTTGGAGGCACAG
	ACC-SNP1_FAM	GAAGGTGACCAAGTTCATGCTAGCGGGAGTACAAATAACTG
ACC Oxidase-1	ACC-SNP1_VIC	GAAGGTCGGAGTCAACGGATTAGCGGGAGTACAAATAACTA
	ACC-SNP1_COM	ATGATCCCGATCTTGTACCT
	ACC-SNP2_FAM	GAAGGTGACCAAGTTCATGCTCTCCCGCTATTTAGATTTTTTT
	ACC-SNP2_VIC	GAAGGTCGGAGTCAACGGATTCTCCCGCTATTTAGATTTTTTA
	ACC-SNP2_COM	GCCATGGATTATTGGAGCTA
	ACC-SNP3_FAM	GAAGGTGACCAAGTTCATGCTCTCGTTGAAGCGCCGGAC
	ACC-SNP3_VIC	GAAGGTCGGAGTCAACGGATTCTCGTTGAAGCGCCGGAT
	ACC-SNP3_COM	CCAGGTGACGAACCACGA
Activating signal co-integrator	Brad677_SNP1_FAM	GAAGGTGACCAAGTTCATGCTCCAGAACAGTCAATCGCATCT
	Brad677_SNP1_VIC	GAAGGTCGGAGTCAACGGATTCCAGAACAGTCAATCGCATCC
	Brad677_SNP1_COM	ACAGGCGATGGATTCTGGA
	Brad677_SNP2_FAM	GAAGGTGACCAAGTTCATGCTGTTGTCCACAACAGTCCATT
	Brad677_SNP2_VIC	GAAGGTCGGAGTCAACGGATTGTTGTCCACAACAGTCCATC
	Brad677_SNP2_COM	CCATATACCGAGGATTCTGAGTTG
	Brad677_SNP3_FAM	GAAGGTGACCAAGTTCATGCTGTCCATCGTACTCGCAAAAC
	Brad677_SNP3_VIC	GAAGGTCGGAGTCAACGGATTGTCCATCGTACTCGCAAAAT
	Brad677_SNP3_COM	TTGCAAAAACAATGGACTGTTC
Protein Phosphatase 1-like	Brad690_SNP2_FAM	GAAGGTGACCAAGTTCATGCTGTCCTAGAGTTGAGACACTAATTTTAA
	Brad690_SNP2_VIC	GAAGGTCGGAGTCAACGGATTGTCCTAGAGTTGAGACACTAATTTTAG
	Brad690_SNP2_COM	AGACTTAACCAAGTCTCAGTCA
Hypothetical Protein	Brad710_SNP3_FAM	GAAGGTGACCAAGTTCATGCTGGAGCACAGGAACGTGATGTAT
	Brad710_SNP3_VIC	GAAGGTCGGAGTCAACGGATTGGAGCACAGGAACGTGATGTTG
	Brad710_SNP3_COM	TAAACGGCACCCTCTAGCAG
ERF_B	ERF_B_SNP1_FAM	GAAGGTGACCAAGTTCATGCTATTGATTTTTTTCTTTAACATTCC
	ERF_B_SNP1_VIC	GAAGGTCGGAGTCAACGGATTATTGATTTTTTTCTTTAACATTCA
	ERF_B_SNP1_COM	AGTTATTTGTATGATGTGATAGGG
	ERF_B_SNP6_FAM	GAAGGTGACCAAGTTCATGCTGAGGGTTCGAGAGTTTGAGG
	ERF_B_SNP6_VIC	GAAGGTCGGAGTCAACGGATTGAGGGTTCGAGAGTTTGAGA
	ERF_B_SNP6_COM	CTGGCCTCCTTCTTTTCTTAACTA
ERF_C	ERF_C _ SNP1_FAM	GAAGGTGACCAAGTTCATGCTTTAGGTGAAGCATCCAAAACC
	ERF_C_SNP1_VIC	GAAGGTCGGAGTCAACGGATTTTAGGTGAAGCATCCAAAAAA
	ERF_C_SNP1_COM	GCAGGTTTGGTTCGGAAGT
	ERF_C_SNP3_FAM	GAAGGTGACCAAGTTCATGCTGACCATGTCCGGCGTTCT
	ERF_C_SNP3_VIC	GAAGGTCGGAGTCAACGGATTGACCATGTCCGGCGTTCG
	ERF_C_SNP3_COM	CCTTCGACTTCCTGCCTGT
	ERF_C_SNP4_FAM	GAAGGTGACCAAGTTCATGCTTGTGAGGCTGTTTGGTTTGA
	ERF_C_SNP4_VIC	GAAGGTCGGAGTCAACGGATTTGTGAGGCTGTTTGGTTTGG
	ERF_C_SNP4_COM	AGCCGAAGCCTGGTCAAC
	ERF_C_SNP5_FAM	GAAGGTGACCAAGTTCATGCTTTGCATTTTGTGTAGAGCACC
	ERF_C_SNP5_VIC	GAAGGTCGGAGTCAACGGATTTTGCATTTTGTGTAGAGCACG

	ERF_C_SNP5_COM	GACATTCGTGGTGCTATGGA
LRR Kinase 1	LRR-1_SNP1_FAM	GAAGGTGACCAAGTTCATGCTTTACACTATGTAGTTGGTTTTCTG
	LRR-1_SNP1_VIC	GAAGGTCGGAGTCAACGGATTTTACACTATGTAGTTGGTTTTCTA
	LRR-1_SNP1_COM	CACCTACGATTTCAAATTTAACCA
	LRR-4_SNP1_FAM	GAAGGTGACCAAGTTCATGCTTTCCAGTGAATCGTATCCTC
	LRR-4_SNP1_VIC	GAAGGTCGGAGTCAACGGATTTTCCAGTGAATCGTATCCTG
	LRR-4_SNP1_COM	GACTTTGTCGCCACCGTATT
PM19-A1	*PM19-A1*_SNP1_FAM	GAAGGTGACCAAGTTCATGCTTTGATGTCGTAGTATGTTTCT
	*PM19-A1*_SNP1_VIC	GAAGGTCGGAGTCAACGGATTTTGATGTCGTAGTATGTTTCC
	*PM19-A1*_SNP1_COM	TGTGATGTCGGTAAGCTC
	*PM19-A1*_SNP12_FAM	GAAGGTGACCAAGTTCATGCTGAAGATTCAAAAGGTAGATTAG
	*PM19-A1*_SNP12_VIC	GAAGGTCGGAGTCAACGGATTGAAGATTCAAAAGGTAGATTAC
	*PM19-A1*_SNP12_COM	GCTGGCGTATTCCACCAC
	*PM19-A1*_DEL1_FAM	GAAGGTGACCAAGTTCATGCTGGCACGCGTCTCACCACC
	*PM19-A1*_DEL1_VIC	GAAGGTCGGAGTCAACGGATTGGCACGCGTCTCACCACG
	*PM19-A1*_DEL1_COM	CTGGGATTGGATTTGGCG
PM19-A2	*PM19-A2*_DEL1_FAM	GAAGGTGACCAAGTTCATGCTTCGAGCAGTGCAACCTAAGAT
	*PM19-A2*_DEL1_VIC	GAAGGTCGGAGTCAACGGATTTCGAGCAGTGCAACCTAAGAG
	*PM19-A2*_DEL1_COM	ATGCGGCCGCTTAGAAAT
	*PM19-A2*_SNP1_FAM	GAAGGTGACCAAGTTCATGCTCCAGCCCCTCGTGTGATG
	*PM19-A2*_SNP1_VIC	GAAGGTCGGAGTCAACGGATTCCAGCCCCTCGTGTGATC
	*PM19-A2*_SNP1_COM	CAACCACACACACACATCCA
RPM1	Sc4326385_SNP1_FAM	GAAGGTGACCAAGTTCATGCTCCAACTCCAAAACCTTGAGAC
	Sc4326385_SNP1_VIC	GAAGGTCGGAGTCAACGGATTCCAACTCCAAAACCTTGAGAT
	Sc4326385_SNP1_COM	AAACTAACGGCCATTAAAGGAG
	Sc4326385_Del_FAM	GAAGGTGACCAAGTTCATGCTTTGACTGACTTGGCTGCAGT
	Sc4326385_Del_VIC	GAAGGTCGGAGTCAACGGATTTTGACTGACTTGGCTGGGTC
	Sc4326385_Del_COM	AGGAAGAGGGCGAGCTAGGT
STW17	STW17_FAM	TTGCAGTCATGGTGTGATAAT
	STW17_VIC	TTGCAGTCATGGTGTGATAAC
	STW17_Com	GCTAGCCACACACCGTCT
STW18	STW18_FAM	GAAGGTGACCAAGTTCATGCTAGGGGAGGTGATCGTGGAT
	STW18-VIC	GAAGGTCGGAGTCAACGGATTAGGGGAGGTGATCGTGGAG
	STW18_COM1	ATCGCGTTGTAGCCGAAT

Note that the first 21 bp in the FAM and VIC primers represent the FAM and VIC dye probe sequences, respectively.

Appendix 2: Primers of the SNP markers designed from non-genic BAC sequence in the Phs physical map

BAC	**Primer Name**	**Primer (probe) Sequence (5'-3')**
TaaCsp4AL031F05	031_F05_T7_SNP8_FAM	GAAGGTGACCAAGTTCATGCTGGTGTCTGCAACAATCATAGC
	031_F05_T7_SNP8_VIC	GAAGGTCGGAGTCAACGGATTGGTGTCTGCAACAATCATAGT
	031_F05_T7_SNP8_COM	CATAATTGTTGCCGATGCTG

TaaCsp4AL032F12	032_Ctg1_SNP1_FAM	GAAGGTGACCAAGTTCATGCTAAAGAATCCCCACCGAAG
	032_Ctg1_SNP1_VIC	GAAGGTCGGAGTCAACGGATTAAAGAATCCCCACCGAAT
	032_Ctg1_SNP1_COM	TTCCTGCTCTCTCTTCATCTTCTAC
TaaCsp4AL052N17	052_N17_SNP2_FAM	GAAGGTGACCAAGTTCATGCTGTGGCCTGGCATGACTTT
	052_N17_SNP2_FAM	GAAGGTCGGAGTCAACGGATTGTGGCCTGGCATGACTTC
	052_N17_SNP2_COM	AAACTTCACATTCCCCCAAA
TaaCsp4AL170N01	170_N01_T7_DEL_FAM	GAAGGTGACCAAGTTCATGCTCAATGGATGTTTATAAGCCGC
	170_N01_T7_DEL_VIC	GAAGGTCGGAGTCAACGGATTCAATGGATGTTTATAAGCCGA
	170_N01_T7_DEL_COM	CAAAGTTGCCTAGCTACGCG
TaaCsp4AL151A19	151_A19_T7_FAM	GAAGGTGACCAAGTTCATGCTGCCAGATTGCCTGCGTAT
	151_A19_T7_VIC	GAAGGTCGGAGTCAACGGATTGCCAGATTGCCTGCGTAC
	151_A19_T7_COM	CACCGCAGCTCACATCATAG

Appendix 3: Primers used for amplification and qPCR analysis the PM19 genes

Purpose	**Primer Name**	**Primer (probe) Sequence (5'-3')**
Amplification of wheat PM19-A2	*PM19-A2*_FULL_F	GCATTATGTGTCAAAATTTATAG
	*PM19-A2*_FULL_R	GAAACAAGTGCTGCCATTCA
Amplification of wheat PM19-A1	*PM19-A1*_FULL_F	CATATCGGTATCCTTCGATGCG
	*PM19-A1*_FULL_R	gCTGGCGTATTCCACCAC
qPCR of copies of wheat PM19	*PM19-A2*-QPCR_F3	TTTTTGTAGGGACGACgcg
	PM19-A2-QPCR_R3	AGCATGGACTGATCGATCTCt
	PM19-A1-QPCR_F3	CTGGAGGCGTTCGTGATTg
	PM19-A1-QPCR_R3	GCATACAAGATCGATCATCAgc
	PM19_D2-QPCR_F2	GTGCAAGGAGATTCATATTGGt
	PM19_D2-QPCR_R2	AGCAGCTGCGTGAACATc
	PM19_B1-QPCR_F2	AACCAGTTCGGTAACCATGC
	PM19_B1-QPCR_R2	CCACGCCGAGATAGACGac
	PM19_B2-QPCR_F1	CGCAGCTGCTGTACGTGATa
	PM19_B2-QPCR_R1	TCACTGTCACCGCCGAtt
Reference genes used for the qPCR analysis of the PM19 copies	TaEF-1A_F	TGGTGTCATCAAGCCTGGTATGGT
	TaEF-1A_R	ACTCATGGTGCATCTCAACGGACT

Appendix 4: PM19-A1 sequence for contrasting Phs parent

>*PM19-A1*_ROBIGUS

GTCTTTCATTGTCAAAACTTCTGTAGTGATTACATCTGCAGCGAAAGGAATAAATTTGCATTGTGCTGCCACAAAGTT
TTCTCTAACTGGTTTGCTGAATTCGTGGATCGGGAGAGCGAAGAAGAGCTTGACACCGAAGCAAAGCGTCCCCACGTG
AGGTGGTTGATTCGATAGACAGAGCGTCCATTTCTATTTTAGTGGGCGTTTGCCAAGCGGAGGCGAAGTGTGGAAGCA
TCCGCACGCTGCTCGAAACAGCTACCGTGTAAAGCCCCCATGCGGCACGCGTCGCACCACGCGTGTCCCTCCTGCACA
CCTATATATTTATCCCACTACACTCCACTTCACCACCTCATCCAGAAATCATCACATTCCATTTGATTCTTGCTTCCG
CCAAATCCAATCCCAGCACGTAGCTAGATCAAGATCAACCAAAGCGAATCCAGTGAGTGTCAGAGCTCCATTTCCATC
TCCACCTCGACCAAGAAGAAACGGGAGCAGCAATGGCGGGCGTTGGGCGTAAACATGGTGGCGCCGCTGCTGGTGCTG
AACCTCATCATGTACATCATCGTCATAGGGTTCGCGAGCTGGAACCTCAACCACTTCATCAACGGCCTCACCAACCG
CCCCGGCGTCGGCGGCAACGGCGCCACCTTCTACTTCCTCGTCTTCGCCATCCTCGCCGGGGTCGTGGGCGCCGCATC
CAAGCTCGCCGGCGTGCACCACGTCCGCACCTGGCGCGGCGACAGCCTCGCCACCTCCGCGTCCTCCTCGCTGGTGGC
CTGGGCGATCACCGCCTTGGCGTTCGGGCTGGCGTGCAAGGAGATCCACATCGGCGGGTACCGAGGGTGGCGCCTCCG
GGTGCTGGAGGCGTTCGTGATTGTTCTCATGTTCACGCAGCTGCTGTACGTGCTGGCGCTGCACTCGGGCCTCTTCGG
GAACCAGTTCGGTGGTAACCATGGTGGTGGGTATCCGGCGGAGCACGCGTACGGCGCAGGCGTCGGCGACCCGCACAA
CAAGGGCATGGGCATGGGCACCGGCGGGGTCGCCAGGGTCTGA

>*PM19-A1*_ALCHEMY

GTCTTTCATTGTCAAAACTTCTGTAGTGATTACATCTGCAGCGAAAGGAATAAATTTGCATTGTGCTGTCACAAAGTT
TTCTCTAACTGGTTTGCTGAATTCGTGGATCGGGAGAGCGAAGAAGAGCTTGACACCGAAGCAAAGCGTCCCCACGTG
AGGTGGTTGATTCGATAGACAGAGCGTCCATTTCTATTTTAGTGGGCGTTTGCCAAGCGGAGGCGAAGTGTGGAAGCA
TCCGCACGCTGCTCGAAACAGCTACCGTGTAAAGCCCCCATGCGGCACGCGTCGCACCACCCGTAAGCGGCGCATCAC
GCGTGTCCCTCCTGCACACCTATATATTTATCCCACTACACTCCACTTCACCACCTCATCCAGAAATCATCACATTCC
ATTTGATTCTTGCTTCCGCCAAATCCAATCCCAGCACGTAGCTAGATCAAGATCAACCAAAGCGAATCCAGTGAGTGT
CAGAGCTCCATTTCCATCTCCACCTCGACCAAGAAGAAACGGGAGCAGCAATGGCGGGCGTGGGTCGCAACATGGTG
GCGCCGCTGCTGGTGCTGAACCTCATCATGTACCTCATCGTCATCGGCTTCGCGAGCTGGAACCTCAACCACTTCAT
CAACGGCCTCACCAACCGCCCCGGCGTCGGCGGCAACGGCGCCACCTTCTACTTCCTCGTCTTCGCCATCCTCGCCGG
GGTCGTGGGCGCCGCATCCAAGCTCGCCGGCGTGCACCACGTCCGCACCTGGCGCGGCGACAGCCTCGCCACCTCCGC
GTCCTCCTCGCTGGTGGCCTGGGCGATCACCGCCTTGGCGTTCGGGCTGGCGTGCAAGGAGATCCACATCGGCGGGTA
CCGAGGGTGGCGCCTCCGGGTGCTGGAGGCGTTCGTGATTGTTCTCATGTTCACGCAGCTGCTGTACGTGCTGGCGCT
GCACTCGGGCCTCTTCGGGAACCAGTTCGGTGGTAACCATGGTGGTGGGTATCCGGCGGAGCACGCGTACGGCGCAGG
CGTCGGCGACCCGCACAACAAGGGCATGGGCATGGGCACCAGGGTCTGA

>*PM19-A1*_CLAIRE

GTCTTTCATTGTCAAAACTTCTGTAGTGATTACATCTGCAGCGAAAGGAATAAATTTGCATTGTGCTGCCACAAAGTT
TTCTCTAACTGGTTTGCTGAATTCGTGGATCGGGAGAGCGAAGAAGAGCTTGACACCGAAGCAAAGCGTCCCCACGTG
AGGTGGTTGATTCGATAGACAGAGCGTCCATTTCTATTTTAGTGGGCGTTTGCCAAGCGGAGGCGAAGTGTGGAAGCA
TCCGCACGCTGCTCGAAACAGCTACCGTGTAAAGCCCCCATGCGGCACGCGTCGCACCACGCGTGTCCCTCCTGCACA
CCTATATATTTATCCCACTACACTCCACTTCACCACCTCATCCAGAAATCATCACATTCCATTTGATTCTTGCTTCCG
CCAAATCCAATCCCAGCACGTAGCTAGATCAAGATCAACCAAAGCGAATCCAGTGAGTGTCAGAGCTCCATTTCCATC
TCCACCTCGACCAAGAAGAAACGGGAGCAGCAATGGCGGGCGTTGGGCGTAAACATGGTGGCGCCGCTGCTGGTGCTG
AACCTCATCATGTACATCATCGTCATAGGGTTCGCGAGCTGGAACCTCAACCACTTCATCAACGGCCTCACCAACCG
CCCCGGCGTCGGCGGCAACGGCGCCACCTTCTACTTCCTCGTCTTCGCCATCCTCGCCGGGGTCGTGGGCGCCGCATC
CAAGCTCGCCGGCGTGCACCACGTCCGCACCTGGCGCGGCGACAGCCTCGCCACCTCCGCGTCCTCCTCGCTGGTGGC
CTGGGCGATCACCGCCTTGGCGTTCGGGCTGGCGTGCAAGGAGATCCACATCGGCGGGTACCGAGGGTGGCGCCTCCG
GGTGCTGGAGGCGTTCGTGATTGTTCTCATGTTCACGCAGCTGCTGTACGTGCTGGCGCTGCACTCGGGCCTCTTCGG
GAACCAGTTCGGTGGTAACCATGGTGGTGGGTATCCGGCGGAGCACGCGTACGGCGCAGGCGTCGGCGACCCGCACAA
CAAGGGCATGGGCATGGGCACCGGCGGGGTCGCCAGGGTCTGA

>*PM19-A1*_OPTION

GTCTTTCATTGTCAAAACTTCTGTAGTGATTACATCTGCAGCGAAAGGAATAAATTTGCATTGTGCTGTCACAAAGTT
TTCTCTAACTGGTTTGCTGAATTCGTGGATCGGGAGAGCGAAGAAGAGCTTGACACCGAAGCAAAGCGTCCCCACGTG
AGGTGGTTGATTCGATAGACAGAGCGTCCATTTCTATTTTAGTGGGCGTTTGCCAAGCGGAGGCGAAGTGTGGAAGCA
TCCGCACGCTGCTCGAAACAGCTACCGTGTAAAGCCCCCATGCGGCACGCGTCGCACCACCCGTAAGCGGCGCATCAC
GCGTGTCCCTCCTGCACACCTATATATTTATCCCACTACACTCCACTTCACCACCTCATCCAGAAATCATCACATTCC
ATTTGATTCTTGCTTCCGCCAAATCCAATCCCAGCACGTAGCTAGATCAAGATCAACCAAAGCGAATCCAGTGAGTGT
CAGAGCTCCATTTCCATCTCCACCTCGACCAAGAAGAAACGGGAGCAGCAATGGCGGGCGTGGGTCGCAACATGGTG
GCGCCGCTGCTGGTGCTGAACCTCATCATGTACCTCATCGTCATCGGCTTCGCGAGCTGGAACCTCAACCACTTCAT
CAACGGCCTCACCAACCGCCCCGGCGTCGGCGGCAACGGCGCCACCTTCTACTTCCTCGTCTTCGCCATCCTCGCCGG

GGTCGTGGGCGCCGCATCCAAGCTCGCCGGCGTGCACCACGTCCGCACCTGGCGCGGCGACAGCCTCGCCACCTCCGC
GTCCTCCTCGCTGGTGGCCTGGGCGATCACCGCCTTGGCGTTCGGGCTGGCGTGCAAGGAGATCCACATCGGCGGGTA
CCGAGGGTGGCGCCTCCGGGTGCTGGAGGCGTTCGTGATTGTTCTCATGTTCACGCAGCTGCTGTACGTGCTGGCGCT
GCACTCGGGCCTCTTCGGGAACCAGTTCGGTGGTAACCATGGTGGTGGGTATCCGGCGGAGCACGCGTACGGCGCAGG
CGTCGGCGACCCGCACAACAAGGGCATGGGCATGGGCACCAGGGTCTGA

Appendix 5: PM19-A2 sequence for contrasting Phs parent

>*PM19-A2*_ROBIGUS
GGATCCCCTAATTTGCTATTATTGCACTAAACACACCATAGTTGCGCGAAAAGTTGGAAAGGCCCTCGTGTTTTCTTG
GAGCTAAATGGATGATGAGTGTAAGCAGAAAAAAAATGGATGTAAAAGCAGAGCAAACACGTGAGACACGTACAGTAC
AGCGGGAAGCGGCGAGGAAGGGAAGAGGAAGAGGATAAGGTTGCATGGGAGGGTGGATGGACCGATCGAGCAGTGCAA
CCTAAGATCCAAGGAAGCGTACGTTTAGTTTACGTTACGTGAGCTGATTGATTATTCCAGTCCTAGTCTAGCAGCAGC
AGCAGCAGCTATAAACCGTGCCACGTACGGTCTCGAAGCCGCTGGAAGCATCCGCCAACGGCTAGAGAGCTACGACGT
CGACTGGAAGAAGCAGAAGCAGAAGCAGCAGCAGCAGCAGCTTCAGCTCGGTGTCAAGCGCTCACCACGCGTCGCGCC
ACCTCCATTTCTAAGCGGCCGCATCACACGTGTCCCGTCTTTCCCTACCTCCTGTGCATCTATATATCCCACTACGCT
CCACCCCGGCACCTCATCCACAAAGCAATCACCAATCTCCCATCCTTGTCGATTCGATCTGTTCCACAAAGCCACTTC
AGTGCTCGATCCAAGCCAACAACAACAACAACAACAAGAAGAAGAAAGAAGCAAGGCAGCAGCAATGGCGGGCGTGGG
TCGCAACATGGTGGCGCCGCTGCTGGTGCTGAACCTCATCATGTACCTCATCGTCATCGGCTTCGCGAGCTGGAACCT
CAACCACTTCATCAACGGCCTCACCAACCGCCCCGGCGTCGGCGGCAACGGCGCCACCTTCTACTTCCTCGTCTTCGC
CATCCTCGCCGGGGTCGTGGGCGCCGCATCCAAGCTCGCCGGCGTGCACCACCGTCCGCACCTGGCGCGGCGACAGCC
TCGCCACCTCCGCGTCCTCCTCGCTGGTGGCCTGGGCGATCACCGCCTTGGCGTTCGGGCTGGCGTGCAAGGAGATCC
ACATCGGCGGGTACCGAGGGTGGCGCCTCCGGGTGCTGGAGGCGTTCGTCATCATCCTGGCCTTCACGCAGCTGCTCT
ACGTGCTGGCGCTGCACTCTGGGCTGTTTGGCAACCAGTTTGGTAACCATGCTGGTGGATACGGCGCGGAGCACGGCG
GGTATGGCGCCGGCGATCCGCACAATAAGGGCATGGGCACCGGCGGCGTCGCCAGGGTCTGA

>*PM19-A2*_ALC1.F_ALCHEMY
GGATCCCCTAATTTGCTATTATTGCACTAAACACACCATAGTTGCGCGAAAAGTTGGAAAGGCCCTCGTGTTTTCTTG
GAGCTAAATGGATGATGAGTGTAAGCAGAAAAAAAATGGATGTAAAAGCAGAGCAAACACGTGAGACACGTACAGTAC
AGCGGGAAGCGGCGAGGAAGGGAAGAGGAAGAGGATAAGGTTGCATGGGAGGGTGGATGGACCGATCGAGCAGTGCAA
CCTAAGAGCCTCAGCTCGGTGTCAAGCGCTCACCACGCGTCGCGCCACCTCCATTTCTAAGCGGCCGCATCACACGTG
TCCCGTCTTTCCCTACCTCCTGTGCATCTATATATCCCACTACGCTCCACCCCGGCACCTCATCCACAAAGCAATCAC
CAATCTCCCATCCTTGTCGATTCGATCTGTTCCACAAAGCCACTTCAGTGCTCGATCCAAGCCAACAACAACAACAAC
AAGAAGAAGAAAGAAGCAAGGCAGCAGCAATGGCGGGCGTGGGTCGCAACATGGTGGCGCCGCTGCTGGTGCTGA
ACCTCATCATGTACCTCATCGTCATCGGCTTCGCGAGCTGGAACCTCAACCACTTCATCAACGGCCTCAATAACCGCC
CCGGCGTCGGCGGCAACGGCGCCACCTTCTACTTCCTCGTCTTCGCCATCCTCGCCGGGGTCGTGGGCGCCGCATCCA
AGCTCGCCGGCGTGCACCATTGTCCGCACCTGGCATGGGGACAGCCTCGCCACCTCCGCGTCCTCCTCGCTGGTTTGCC
TGGGCGATTTACCGCCTTGGCGTTCGGGCTGGCGTGCAAGGAGATCCACATCGGCGGGTACCGAGGGTGGCGCCTCCGG
GTGCTGGAGGCGTTCGTCATCATCCTGGTTCTTCACGCAGCTGCTCTACGTGCTGGCGCTGCACTCTGGGCTGTTTGGT
AACCATGCTGGTGGATACGGCGCGGAGCACGGCGGGTATGGCGCCGGCGATCCGCACAATAAGGGCATGGGCACCGGC
GGCGTCGCCAGGGTCTGA

>*PM19-A2*_CLAIRE
GGATCCCCTAATTTGCTATTATTGCACTAAACACACCATAGTTGCGCGAAAAGTTGGAAAGGCCCTCGTGTTTTCTTG
GAGCTAAATGGATGATGAGTGTAAGCAGAAAAAAAATGGATGTAAAAGCAGAGCAAACACGTGAGACACGTACAGTAC
AGCGGGAAGCGGCGAGGAAGGGAAGAGGAAGAGGATAAGGTTGCATGGGAGGGTGGATGGACCGATCGAGCAGTGCAA
CCTAAGATCCAAGGAAGCGTACGTTTAGTTTACGTTACGTGAGCTGATTGATTATTCCAGTCCTAGTCTAGCAGCAGC
AGCAGCAGCTATAAACCGTGCCACGTACGGTCTCGAAGCCGCTGGAAGCATCCGCCAACGGCTAGAGAGCTACGACGT
CGACTGGAAGAAGCAGAAGCAGAAGCAGCAGCAGCAGCAGCTTCAGCTCGGTGTCAAGCGCTCACCACGCGTCGCGCC
ACCTCCATTTCTAAGCGGCCGCATCACACGTGTCCCGTCTTTCCCTACCTCCTGTGCATCTATATATCCCACTACGCT
CCACCCCGGCACCTCATCCACAAAGCAATCACCAATCTCCCATCCTTGTCGATTCGATCTGTTCCACAAAGCCACTTC
AGTGCTCGATCCAAGCCAACAACAACAACAACAACAAGAAGAAGAAAGAAGCAAGGCAGCAGCAATGGCGGGCGTGGG
TCGCAACATGGTGGCGCCGCTGCTGGTGCTGAACCTCATCATGTACCTCATCGTCATCGGCTTCGCGAGCTGGAACCT

CAACCACTTCATCAACGGCCTCACCAACCGCCCCGGCGTCGGCGGCAACGGCGCCACCTTCTACTTCCTCGTCTTCGC
CATCCTCGCCGGGGTCGTGGGCGCCGCATCCAAGCTCGCCGGCGTGCACCACGTCCGCACCTGGCGCGGCGACAGCC
TCGCCACCTCCGCGTCCTCCTCGCTGGTGGCCTGGGCGATCACCGCCTTGGCGTTCGGGCTGGCGTGCAAGGAGATC
CACATCGGCGGGTACCGAGGGTGGCGCCTCCGGGTGCTGGAGGCGTTCGTCATCATCCTGGCCTTCACGCAGCTGCTC
TAC

>*PM19-A2*_OPTION
GGATCCCCTAATTTGCTATTATTGCACTAAACACACCATAGTTGCGCGAAAAGTTGGAAAGGCCCTCGTGTTTTCTTG
GAGCTAAATGGATGATGAGTGTAAGCAGAAAAAAAATGGATGTAAAAGCAGAGCAAACACGTGAGACACGTACAGTAC
AGCGGGAAGCGGCGAGGAAGGGAAGAGGAAGAGGATAAGGTTGCATGGGAGGGTGGATGGACCGATCGAGCAGTGCAA
CCTAAGAGCCTCAGCTCGGTGTCAAGCGCTCACCACGCGTCGCGCCACCTCCATTTCTAAGCGGCCGCATCACACGTG
TCCCGTCTTTCCCTACCTCCTGTGCATCTATATATCCCACTACGCTCCACCCCGGCACCTCATCCACAAAGCAATCAC
CAATCTCCCATCCTTGTCGATTCGATCTGTTCCACAAAGCCACTTCAGTGCTCGATCCAAGCCAACAACAACAACAAC
AAGAAGAAGAAGAAAGAAGCAAGGCAGCAGCAATGGCGGGCGTGGGTCGCAACATGGTGGCGCCGCTGCTGGTGCTGA
ACCTCATCATGTACCTCATCGTCATCGGCTTCGCGAGCTGGAACCTCAACCACTTCATCAACGGCCTCACCAACCGCC
CCGGCGTCGGCGGCAACGGCGCCACCTTCTACTTCCTCGTCTTCGCCATCCTCGCCGGGGTCGTGGGCGCCGCATCCA
AGCTCGCCGGCGTGCACCATGTCCGCACCTGGCATGGGGACAGCCTCGCCACCTCCGCGTCCTCCTCGCTGGTTGCC
TGGGCGATTACCGCCTTGGCGTTCGGGCTGGCGTGCAAGGAGATCCACATCGGCGGGTACCGAGGGTGGCGCCTCCGG
GTGCTGGAGGCGTTCGTCATCATCCTGGTCTTCACGCAGCTGCT