GENETIC STUDY OF PRE-HARVEST SPROUTING IN SPRING WHEAT

By

RAMANPREET RAMANPREET

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Department of Plant Science

University of Manitoba

Winnipeg

ACKNOWLEDGEMENT

To begin with I would like to acknowledge The Almighty God who has been The Prime Mover in all my efforts. I owe my deepest sincere gratitude to him for he is the one who has given me the strength, the spirit, the will and the courage to work in the right direction irrespective of the hurdles that came in the path of life.

With a deep sense of respect, I owe my sincere thanks to my revered supervisor, Prof. Belay Ayele. As my supervisor, his guidance helped me to focus on achieving my goal. His observations and assistance helped me to establish the overall direction of the research and to move forward with investigation in depth. Without his able guidance, painstaking efforts and encouragement, the work wouldn't have been what it is. I am truly grateful to him for extending his total co-operation and understanding when things were not too smooth. I am thankful for all those uncounted hours he has spent going through reading my thesis. I am thankful to my committee member Dr. Curt McCartney, and Dr. Marta Izydorczyk for providing the necessary facilities during the course of my M.Sc. work. I am so grateful to Gurkamal for her help and support. I owe my deep sense of gratitude to my lab members Dr. Anh Pham, Dr. Murali, Dr. Manisha, Dr. Aihua, Dr. Uzzal, Ginny, Riya for their concern and cooperation.

I take this opportunity to express my deepest gratitude to the better half of my life, my husband Amandeep Singh. He has always pumped in enthusiasm and positivity in me and made me believe 'I can'. Special thanks to my lovely daughter Sargun Kaur and cute son Fateh Singh for their love and support. I also thank my mother-in-law, father-in-law, Pawan di, Baljeet Singh jiju and my nephew. Last but not least, I would like to pay high regards to my Mummy, Papa, brothers Jagpreet, Harpreet and sisters-in-law Bubleen, Kaminder and my nephews.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Ramanpreet, Ramanpreet. M.Sc. The University of Manitoba, November 2023. Genetic study of pre-harvest sprouting in spring wheat. M.Sc. Supervisor: Dr. Belay T. Ayele

Pre-harvest sprouting (PHS), which is defined as the germination of seeds prior to harvest under moist conditions, occurs in wheat and is closely related to seed dormancy. Low levels of seed dormancy result in pre-harvest sprouting, which adversely impacts wheat grain yield and end-use quality; for example, sprouted seeds have poor quality that is unsuitable for baking. On the other hand, high levels of seed dormancy cause a delay in germination and seedling establishment. Therefore, prevention of PHS or minimizing its negative effects in wheat requires the development of wheat cultivars with moderate level of dormancy in the seeds. However, the development of such cultivars requires the need to identify genomic regions and genes that are responsible for PHS resistance. Therefore, this research was focused on genome wide association study of a collection of spring wheat lines, which show a wide range of variation in seed dormancy, over six trial environments. Analysis of the association between the phenotypic and genotypic data of wheat lines studied led to the detection of the genomic regions and candidate genes that might have the potential to introduce PHS resistance in spring wheat.

FOREWORD

This thesis is written in manuscript style. A general introduction about the research project and a literature review precedes the manuscript. An abstract, introduction, materials and methods, results and discussion form a complete manuscript. The manuscript is followed by a general discussion and conclusions, a list of references and appendices.

1.0 GENERAL INTRODUCTION

Wheat is one of the most important food crops in terms of planted acreage and production volume. It provides 20 percent of calories consumed by the global population (http://faostat.fao.org/). Because of its use as a staple food for humans and feed for animals, there is an increased demand for wheat. However, its production is negatively affected by several biotic and abiotic stress factors. Pre-harvest sprouting (PHS), which is defined as the germination of seeds on the mother plant before harvest, is one of the factors that lead to significant yield loss and poor grain quality in wheat. PHS usually occurs due to wet and moist summer conditions that occur before harvest. PHS causes activation of hydrolytic enzymes such as alpha-amylase which causes breakdown of starch into simple sugars that are required to support the germination process. As a result, sprouted grains have poor quality and are sold at a significant discount since the sprouting induces poor end-use quality (Castor and Frederiken, 1977). Farmers and food processors can suffer significant financial losses as a result of poor grain quality and lower yields caused by PHS (Thomason *et al*., 2019).

The degree of resistance and susceptibility to PHS depends upon the level of dormancy in the wheat seeds. Seed dormancy is an adaptive trait that blocks the germination of seeds under optimal environmental conditions (Gao and Ayele, 2014). High and low levels of seed dormancy are undesirable traits since they cause non-uniform germination and PHS, respectively. Therefore, one of the major objectives of wheat breeding programs around the world has been to develop wheat cultivars with moderate levels of seed dormancy. Seed dormancy or PHS are regulated by signaling compounds and a variety of environmental factors. The involvement of plant hormones, particularly that of abscisic acid (ABA) and gibberellin (GA), in regulating PHS is crucial (Nonogaki *et al*., 2018). In wheat grains, these two phytohormones act antagonistically to regulate

seed dormancy and germination. Gibberellin promotes germination, for example by inducing the synthesis of cell wall hydrolytic enzymes that induce cell elongation and seed germination whereas ABA inhibits germination/promotes dormancy via repressing germination-related physiological processes such as inhibition of the synthesis of hydrolytic enzymes in seeds (Steinbach *et al*., 1997; Li *et al*., 2018).

PHS is a complex and quantitative trait controlled by multiple genes. Genetic mapping has been the main approach to analyze the genetic architecture of such complex traits in field crops such as wheat. Genetic mapping mainly involves two methods, linkage analysis and association mapping. Linkage analysis is often characterized by poor QTL resolution and identification of only a small number of alleles. Moreover, the development of mapping populations is a more labor-intensive and time-consuming process. Association mapping has been an alternative method to find associations between markers and traits and this approach is better than linkage analysis in terms of high mapping resolution and avoids the time required to develop a mapping population. The main goal of this study was therefore to identify genomic regions and candidate genes associated with PHS through genome-wide association mapping of diverse spring wheat collection over several environmental conditions. The specific objectives of the thesis project are to:

1. characterize the dormancy/PHS characteristics of the diverse wheat genotypes.
2. identify marker-trait association for PHS by combining the genotypic and phenotypic data.
3. identify and validate genomic regions and candidate genes controlling PHS resistance.

2.0 LITERATURE REVIEW

2.1 Wheat: the common food crop

Wheat is one of the important cereal crops belonging to the Poaceae family. It is grown extensively in a variety of environments around the world and known to be the second essential cultivated cereal crop in the developing world, after rice, in terms of food security; approximately 80 million farmers dependent on it for their living (Patricia *et al*., 2019). Wheat is one of the first food crops to be domesticated and its domestication first took place in the Middle East around 10,000 years ago, when transition from hunting and gathering food to established agriculture occurred (Dubcovsky and Dvorak, 2007; Shewry, 2009). Wheat has been used as the main food crop in Europe, West Asia and North Africa for about 8000 years (Giraldo *et al.*, 2019), and this is most likely due to its agronomical versatility, ease of storing the grains and converting the flour to a variety of cuisines (Curtis, 2019). In the year 2021, wheat was cultivated on 220 million hectares around the world with a production volume of approximately 777 million tonnes, making it as a second cereal crop produced after maize (Food and Agricultural Organization, 2022). With an expected global population of above 9 billion people by 2050, the demand for wheat is estimated to rise by 60% as compared to the demand reported in 2010 (Food and Agricultural Organization, 2022). As such, an increase of a global annual yield from 1% per year (2001-2010) to 1.6% per year (2011-2050) is required to satisfy the increasing demand of the crop (United Nations, 2019). About 95% of the cultivated wheat is hexaploid while the remaining 5% is durum wheat and some other less important types (Shewry, 2009). Wheat comes in a variety of natural forms however white and red wheat are the most popular types of wheat.

2.2 Wheat production in Canada

Canada is one of the largest producer of wheat, and, for example, in the year of 2020, it ranked as the six largest producer with over 24.8 million acres harvested and 21.7 million tonnes produced (Statistics Canada, 2021). It is the third-largest exporter of wheat after Russia and USA and is also known for its high-quality wheat (Food and Agricultural Organization, 2021). Wheat is mainly cultivated in Canada's Prairie provinces (Alberta, Saskatchewan and Manitoba) with a less amount in British Columbia and eastern Canada (McCallum and DePauw, 2008; Randhawa *et al*., 2012). Hexaploid wheat is the most common type of wheat produced in Canada, and in the year 2020 it constituted 77 % of total wheat production followed by durum wheat (*Triticum turgidum* L.), which contributed the remaining 23% of the total produce (Statistics Canada, 2020). In Canada, to date, 34 research institutes and corporations have registered 363 cultivars of spring wheat (Canadian Grain Commission, 2021). In Western Canada, these cultivars of wheat are currently divided into eight classes on the basis of grain baking and milling quality, grain protein content and end product use (Canadian Grain Commission, 2021). These classes include Canada Northern Hard Red (CNHR), Canada Prairie Spring Red (CPSR), Canada Prairie Spring White (CPSW), Canada Western Extra Strong (CWES), Canada Western Hard White Spring (CWHWS), Canada Western Red Spring (CWRS), Canada Western Soft White Spring (CWSWS), and Canada Western Special Purpose Spring (CWSP).

2.3 Evolution of common wheat

Common wheat (*Triticum aestivum*) is one of the important cultivated and ancient cereal crops. It has a complex allohexaploid genome (2n = 6*x* = 42, AABBDD), which is formed by natural hybridization and chromosome doubling (Figure 2.1; El Baidouri *et al*., 2017; Marcussen *et al*.,

2014). The first phase of chromosome doubling resulted in the production of *T. turgidium* ($2n = 4x = 28$, AABB), which is a tetraploid wheat with genome composition AABB; the A genome is donated by *T. urartu* (Dvorak *et al*., 1998; Friebe and Gill, 1994) while donor of the B genome is extinct but assumed to be an ancestor of *Aegilops speltoides*. In the second phase, hybridization between *T. turgidum* ($2n = 4x = 28$, AABB) and *Ae. tauschii* ($2n = 2x = 14$, DD) resulted in the formation of *T. aestivum* (Heun *et al*., 1997; Huang *et al*., 2002). Archeological evidence has been used to estimate the timeline of evolution of common wheat and its non-existence in wild populations. Although available evidence suggested that the sub-genomes of hexaploid wheat are related to the diploid species of wheat, the phylogeny and divergence history of the three genomes remain unknown (Eversole *et al*., 2014).

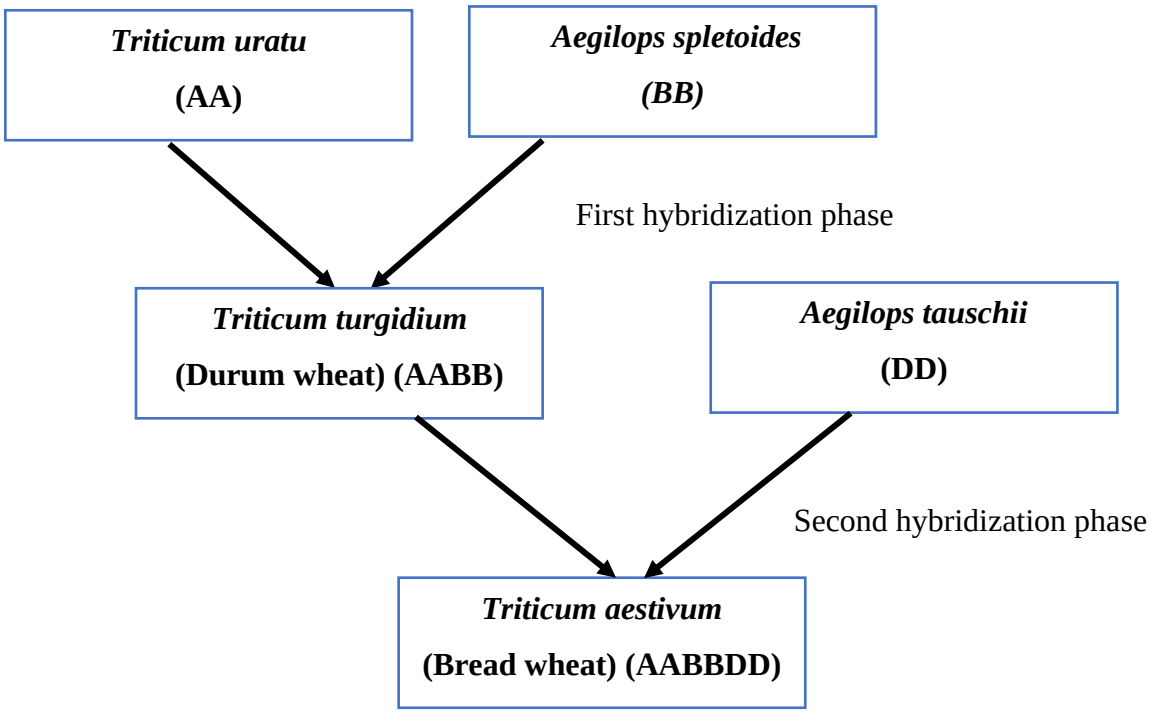


Figure 2.1 The two phases of hybridization between the three species of wheat during evolution of hexaploid wheat.

2.4 Wheat genome

The genome of common wheat (*Triticum aestivum* L.), which is composed of three sub-genomes, is very complex and highly repetitive. There are seven pairs of homologous chromosomes in each of the three sub-genomes, which makes a total of 42 chromosomes. Homoeologous chromosomes are duplicated chromosomes that come from distinct parental species and are brought together in the same genome by allopolyploidization (Glover *et al*., 2016). Wheat is a self-pollinating crop, therefore, genotypes are largely homozygous and homologs at most loci have identical alleles (Dixon *et al*., 2018). The complex genome of the wheat makes molecular and genetic studies very difficult; however, the latest advancements in molecular approaches and the advent of next-generation sequencing technologies have made the genetic and molecular studies of wheat possible. The International Wheat Genome Sequencing Consortium (IWGSC) recently released a completely annotated reference genome for the hexaploid wheat variety Chinese Spring, which contains 107,891 projected high-confidence protein-coding genes in a 14.5 Gb assembly (IWGSC, 2018). The hexaploid genome is believed to have 17 billion nucleotides, or five times the size of the human genome, with over 85 percent of the genome being repetitive (IWGSC, 2018). The genome sizes of tetraploid and diploid wheat are approximately two-third and one-third of that of hexaploid wheat, respectively. Overall, availability of the wheat genome reference sequence allows wheat scientists to develop novel techniques for wheat improvement and wheat breeders to employ genomics-assisted breeding strategies.

2.5 Uses of wheat

Wheat is the main source of food for about 40% of the global population and plays an important role in imparting nutrition to mankind through providing 20% of the dietary calories and protein

(Patricia *et al*., 2019). It also contains minerals, vitamins, fats, micronutrients and fibers (Shewry and Hey, 2015; Lafiandra *et al*., 2014). Hexaploid wheat is commonly used for making a wide range of leavened and flatbread, and other forms of baked goods. The main cause of bread loaves to be raised is the presence of gluten in the wheat that helps to trap the carbon dioxide bubbles during fermentation (Hoseney and Rogers, 1990). On the other hand, *T. turgidium*, which is commonly known as durum wheat is used to make pasta (Patricia *et al*., 2019). Wheat is also used as an animal feed, especially when rain has hampered harvests and large amounts of the grains are unfit for human consumption. Einkorn wheat (*T. monococcum*) is currently a relic crop used for animal feed in the Mediterranean regions (Peng *et al*., 2011). Moreover, low-quality grains are frequently used to create glues, paper additives, and a variety of other products, including malt and beer (Curtis, 2019).

Wheat grain also contains therapeutic characteristics and can be used to treat a variety of ailments (Kumar *et al*., 2011). Wheat germ helps to prevent heart disease, while the whole wheat grain with bran helps to avoid constipation, ischaemia, diverticulum, appendicitis, obesity and diabetes (Kumar *et al*., 2011). However, the gluten molecule of wheat is linked to celiac disease, a persistent bowel swelling that causes nutrient malabsorption, and this disease has been reported to affect about 1% of population in Western Europe (Tye-Din *et al*., 2018).

2.6 Challenges of wheat production

Wheat production is influenced by various biotic and abiotic stress factors. The major biotic factors that affect wheat production include pathogens and insect pests. The most common diseases that affect wheat production in Canada and worldwide include fusarium head blight (FHB), rusts (stem, leaf, and stripe), powdery mildew, leaf spot complex and wheat blast (Singh *et al*., 2016). The

abiotic stress factors that affect wheat production include drought, salinity, heat and cold (Dhakal *et al*., 2021) Heat stress is known to increases pollen sterility, reduces carbon dioxide assimilation and promotes photorespiration in wheat (Hlaváčová *et al*., 2017), whereas low temperatures cause sterility in up to 98 percent of wheat plants (Bandyopadhyay, 2011). Drought affects more than half of the world's wheat-growing regions and causes yield losses of up to 27.5% (Zhang *et al*., 2018).

Pre-harvest sprouting (PHS), which refers to the germination of grains while being on the mother plant due to the occurrence of wet and moist conditions before harvest (Gao and Ayele, 2014) is among the abiotic stress related problems that causes significant reductions in wheat yield and quality, and therefore major financial loss to the wheat producers (Varughese *et al*., 1987). This is due to the fact that PHS increases alpha-amylase activity and induces hydrolysis of starch in the endosperm, leading to loss of grain weight and end-use quality (Castor and Frederiken, 1977). It also reduces the storage life of grains (Bason *et al*., 1991). Globally, PHS is estimated to cause an annual loss of about ~$1 billion in wheat crops (Black *et al*., 2006).

2.7 Pre-harvest sprouting and its effect on wheat production in Canada

Pre-harvest sprouting is also a major concern to Canadian wheat producers. Reductions in wheat yield and quality due to PHS occurred in four of ten years between 1978 and 1988, costing an estimated loss of $100 million in each of those four years (Derera, 1990). PHS also caused serious damage to Canadian wheat production and significant economic loss to the producers in 2000, 2002 and 2010 (Singh *et al*., 2014). In the Prairie provinces of Alberta, Saskatchewan and Manitoba, which are the major wheat-producing regions of Canada, PHS occurs frequently and it is one of the major concerns of wheat producers. PHS damage is caused by increased activity of

α-amylase in the grains, and the level of damage is usually assessed via falling number and stirring number tests. The effects of alpha-amylase on the viscosity of a starch paste are measured by both of these methods (Hocking *et al*., 1993). Gluten content of the flour made from PHS-affected cereal grains is normally low, resulting in poor flour baking quality, sticky crumbs and collapsed loaves (Imtiaz *et al*., 2008). Loaves with sticky crumbs have a slicing issue, which is an undesirable feature that affects the market value of the crop (Moot and Every, 1990).

Resistance to PHS is dependent on a number of morphological traits, physiological factors and environmental conditions. The morphological traits include erectness of the spike, structure of the awn and spike, and openness of the floret (King and Richards, 1984) while the physiological factors are related to seed dormancy, phytohormones such as abscisic acid (ABA), gibberellin, auxin, brassinosteroids and ethylene, seed coat colour and structure of the grain, activity of alpha-amylase and germination-inhibiting compounds (Emebiri *et al*., 2010; Gao and Ayele, 2014; Mares and Mrva, 2014; Gatford *et al*., 2002). The environmental conditions that cause major influence on PHS include temperature and relative humidity (Gao and Ayele, 2014). Among all these factors, seed dormancy is considered to be the most important trait that determines PHS resistance.

2.8 Seed dormancy

Seed dormancy is an adaptive trait that prevents the germination of viable seeds under optimal environmental conditions (Gao and Ayele, 2014). Both genetic and environmental factors influence dormancy, and the depth of dormancy or the speed at which dormancy is released in a given genotype can differ under different environmental conditions (Koornneef *et al*., 2002 and Gao *et al*., 2013). Based on the timing of its onset, dormancy is divided into primary dormancy and secondary dormancy. Primary seed dormancy is innate and achieved during seed maturation

phase, whereas secondary dormancy is generated in non-dormant seeds as a result of unfavorable environmental conditions (Finkelstein *et al*., 2008).

Whereas, on the basis of the cause of primary dormancy, it is divided into seed coat dormancy and embryo dormancy (Bewley and Black, 1994). The seed coat and other enclosing tissues, which includes the endosperm, pericarp and extra-floral organs, cause seed coat dormancy mainly via limiting water absorption and gas exchange or secreting germination inhibitory chemicals to the embryo (Bewley 1997).

Embryo dormancy is due to the presence of an underdeveloped or rudimentary embryo, and seed germination is inhibited due to a lack ripening of the embryo or presence of germination inhibitors with little effect originated from the seed coat or surrounding tissues. (Derera *et al*., 1977). Phytohormones such as ABA and GA are considered to be responsible for the presence or absence of embryo dormancy. ABA promotes and maintains seed dormancy whereas GA induces seed germination (Finkelstein *et al*., 2008). A detail of the function of these two hormones in inducing, maintaining and release of seed dormancy is discussed in detail in a later section.

The degree of resistance to PHS is related to the level of seed dormancy in cereal seeds. High level of seed dormancy delays germination and results in poor seedling establishment and low level of dormancy is also undesirable in cereal crops because it makes the seeds susceptible to PHS (Derera, 1989). It is, therefore, preferable to have a moderate level of dormancy to prevent seeds from PHS. Wheat breeding programs have had a goal of selection against seed dormancy in order to achieve rapid and uniform germination (Simpson, 1990). However, this has a drawback in that it increases the occurrence of PHS. Therefore, it is highly desirable to develop cultivars that stay dormant before harvest but then break the dormancy soon after harvesting to ensuring rapid and uniform germination (Ullrich *et al*., 2009).

There are several established approaches to determine the level of seed dormancy, which include germination index (GI), sprouting index (SI) and falling number (FN). Germination index (GI) is a direct determination of seed dormancy, and this can be evaluated through seed germination test. Sprouting index (SI) can be estimated by wetting the whole spikes (Paterson *et al*., 1989). Falling number test involves evaluation of the action of the alpha-amylase which degrades the starch in germinating grains (Hagberg, 1961). Low FN values reflects grain-starch degradation caused by increased alpha-amylase activity and are an indicator of low PHS resistance.

2.9 Seed germination

The process of germination starts with water uptake by the seed and ends with radicle protrusion through the seed coat (Bewley and Black, 1994). The uptake of water during seed germination is divided into three phases. The quiescent dry seed uptakes water rapidly during the first phase, resulting in the enlargement of the seed. Moreover, during this phase, the cell membrane of the imbibed seeds undergoes many structural changes that result in rapid leakage of solutes and other small metabolites to the surrounding medium (Bewley, 1997). The first phase of water uptake also includes repair of cell membrane, DNA and mitochondrial damage, and the seed begins respiratory activity and protein synthesis utilizing transcripts already present in the dry seed (Bewley, 1997) The second phase of water uptake, is also known as the plateau phase i.e., slow and constant rate of water uptake by the seeds. During this phase, the seed's metabolic activity resumes and it prepares for further growth. Protein synthesized from newly synthesized mRNA and mitochondrial synthesis occurs during this phase. Seeds go through a phase change from embryo axis growth to germination. Then the seed enters into the third phase that is associated with the growth phase, which includes mobilization of stored food reserves, elongation of the radicle, cell division and

DNA synthesis. The seedling's water intake increases during phase three, which is used for divisions of the cells and the enlargement of the developing radicle (Nonogaki *et al*., 2010). In wheat, various physiological and biochemical processes that are controlled by hormonal and environmental factors affect the induction and maintenance of seed dormancy and germination, and these alterations are primarily controlled by two growth regulators, abscisic acid (ABA) and gibberellin (GA) (Tuan *et al*., 2018). Alterations in environmental conditions such as temperature, light, and oxygen also affect seed dormancy and germination in grain crops (Benech-Arnold *et al*., 2006).

2.10 Hormonal control of seed dormancy and germination

The involvement of plant hormones (ABA and GA) in seed dormancy and germination regulation is well documented in various crops (Kucera *et al.,* 2005; Finkelstein *et al*., 2008; Linkies and Leubner-Metzger, 2012). Genes that encode starch hydrolytic enzymes such as α -amylases, which are required for seed germination, are stimulated by GA but repressed by ABA in the aleurone layer of the grain (Lovegrove and Hooley, 2000; Brady and McCourt, 2003).

Although the molecular mechanisms and functions of ABA and GA in seed dormancy have been the focus of research in seed biology, a number of reports investigated the involvement of other plant hormones such as jasmonate, ethylene and brassinosteroid in regulating seed dormancy and germination through affecting the interaction between GA and ABA (Shu *et al*., 2016).

2.10.1 Role of abscisic acid

Abscisic acid (ABA) is a sesquiterpenoid plant hormone that regulates seed dormancy and plant response to abiotic stress conditions such as drought, cold and salinity (Zeevaart and Creelmann,

1988; Seo and Koshiba, 2002; Seki *et al*., 2007). The balance between its biosynthesis and catabolism controls ABA levels in the seeds (Nambara and Marion-Poll, 2005; Nambara *et al*., 2010). As a result, dormancy and germination of seeds are significantly influenced by the expression of genes that act as key regulators of ABA biosynthesis and catabolism (Tuan *et al*., 2018). Several enzymes are responsible for the biosynthesis of ABA; however, NINE-CIS-EPOXY CAROTENOID DIOXYGENASE (NCED) enzyme is considered to be the main regulator of ABA synthesis (Gao and Ayele, 2014). The *NCED6* and *NCED9* genes are reported to be highly expressed in developing Arabidopsis seeds, and a mutational study of these two genes revealed that they play a role in ABA production and seed dormancy (Lefebvre et al., 2006). For example, in the absence of plant hormone GA, ABA-deficient mutants of maize show vivipary (McCarty and Donald, 1995) and ABA-deficient mutants of Arabidopsis show germination (Koornneef *et al*., 1982). For the seed to germinate, the high concentration of ABA is lowered after imbibition (Gubler *et al*., 2005). The ABA concentration of after-ripened non-dormant seeds of Arabidopsis has been shown to drop rapidly during imbibition and this leads to the germination of seeds. It has been suggested that a higher level of dormancy is connected with the maintenance of an elevated embryonic ABA level for an extended period of time during the late-maturation phase of wheat seed development (Walker-Simmons and Sesing, 1990). On the other hand, developing seeds of a non-dormant mutant line of wheat show a substantially lower amount of embryonic ABA than that observed in the dormant line from which the mutant was generated (Kawakami *et al*., 1997).

The most common mechanism of ABA catabolism is the 8'-hydroxylation of ABA, which produces 8'-hydroxy ABA that immediately isomerizes to phaseic acid (Nambara and Poll, 2005). The enzyme ABA 8'-hydroxylase, which is a cytochrome P450 monooxygenase, catalyzes this

process, and this enzyme is encoded by *CYP707A* genes. Previous molecular and genetics studies have shown that *CYP707A2* gene is highly expressed in non-dormant than dormant seeds of Arabidopsis (Millar *et al*., 2006). Likewise, a *CYP707* homologue of barley (*HvABA8'OH-1*) was found to be more active in non-dormant than dormant embryos (Millar *et al*., 2006). After-ripening (dry storage of seeds for some time at room temperature) and stratification (cold treatment) induce ABA catabolism and result in a reduced amount of ABA (Gubler *et al*., 2005).

2.10.2 Role of gibberellins

Gibberellins (GAs) are tetracyclic diterpenoid compounds. To date, 136 GAs have been identified in higher plants and fungi; however only few of them such as GA_1, GA_3, GA_4, and GA_7 are physiologically active (Hedden and Phillips, 2000; Sun, 2008). GAs are involved in seed germination, stem elongation, blooming and fruit development in plants. The bioactive level of GA in plant tissues is maintained by a balance between its production, which is controlled by GA 20-oxidase and GA 3-oxidase enzymes, and its inactivation, which is controlled by GA 2-oxidase enzyme (Yamaguchi, 2008). When the embryo releases GA to the aleurone layer, the cells of the aleurone produce hydrolytic enzymes, mainly α-amylases, which hydrolysis the endosperm reserves and regulate the process of germination. It also causes the initiation of programmed cell death in aleurone cells when combined with internally generated reactive oxygen species (Bethke *et al*., 2001). GA released from the embryo activates a variety of reactions, including gene induction and repression, as well as secretory response upregulation. An increase in cytoplasmic Ca^{2+} concentration occurs first after GA treatment. The GA signal is then transduced in the aleurone layer by hetero-trimeric G protein (Lovegrove and Hooley, 2000). Alterations in cytosolic Ca^{2+}, cGMP, calmodulin (CaM) and the transactivating protein GAMyb are used in the signal

transduction pathway that leads to hydrolase production. Within 2-5 minutes of GA treatment, intact wheat aleurone cells have been reported to show a response (Bush, 1996). The majority of reports that link GA to the regulation of seed germination and dormancy in cereals are based on a comparison of dormant and non-dormant seeds. For instance, it has been demonstrated that dormancy loss in wheat and barley seeds owing to after-ripening is associated with increased *TaGA20ox* and *TaGA3ox* gene expression and higher amount of bioactive GA_1 during imbibition (Gubler *et al*., 2008; Liu *et al*., 2013; Kashiwakura *et al.*, 2016). Gene expression of a GA 20-oxidase is likewise suppressed by ABA, indicating that ABA may hinder the process of germination by inhibiting GA production (Pérez-Flores *et al*., 2003). Gibberellins appear to be engaged in germination stimulation and maintenance rather than seed dormancy (Bewley, 1997). Seeds germinate or do not germinate depending on the balance between these two hormones as shown in Figure 2.2, which varies by species and may be affected by concentrations and responsiveness. Furthermore, environmental factors have a significant impact on ABA and GA metabolism and signaling (Footitt *et al*., 2020).

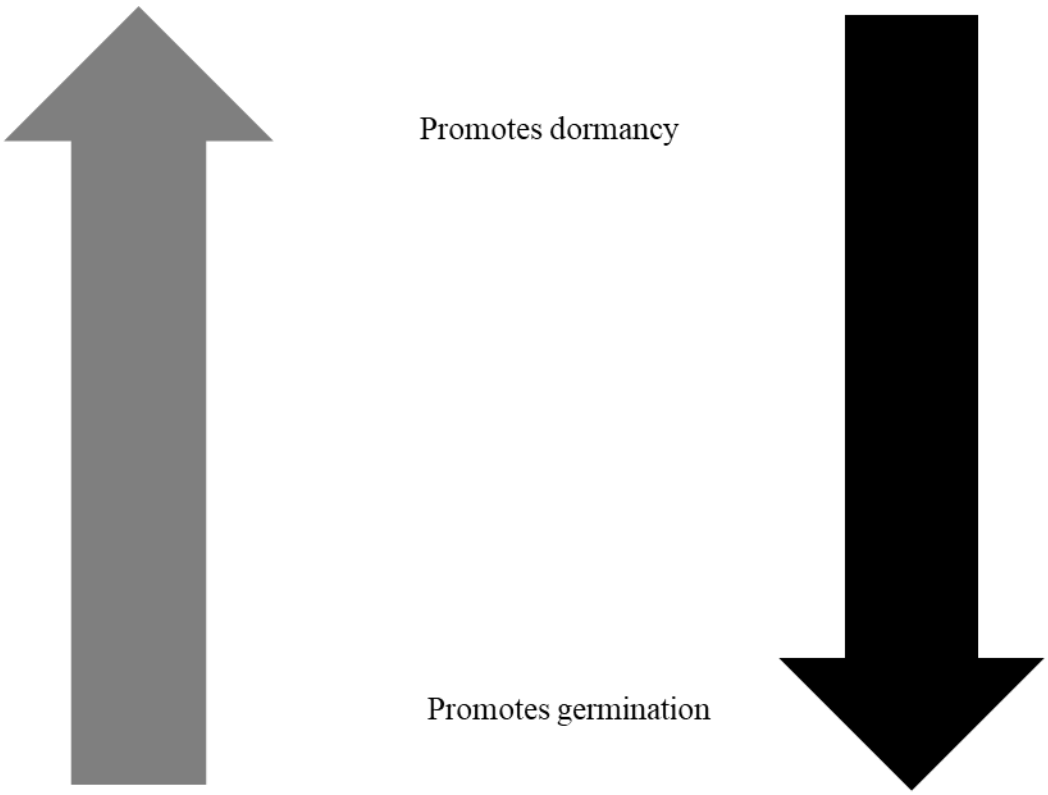


Figure 2.2 Shows the antagonistic role of ABA and GA in seed dormancy and seed germination.

2.11 Role of environmental factors in Seed dormancy

For successful seed germination, different plants require different factors. Under ideal environmental conditions (light, moisture, and temperature), dormant seeds do not germinate (Bewley and Black, 1994). After-ripening is the process of releasing dormancy by storing seeds at room temperature for an extended period of time (Finkelstein *et al*., 2008). Dormancy can also be interrupted by treatment with low temperature, which is known as cold stratification, or by seed coat scarification (Paterson *et al*., 1989; Finkelstein *et al*., 2008). Water, oxygen, optimal temperature and, on occasion, light or darkness are the most crucial external conditions.

Temperature is an important environmental component that influences seed dormancy during seed development and imbibition while humidity levels during grain development and harvest were found to be highly linked to PHS (Thomason *et al*., 2009).

In case of wheat, the study by Walker-Simmons and Sesing in 1990 reported that at 15°C, seeds have a high level of dormancy than those developed under 25°C. This was attributed to the presence of high level of ABA in seeds developed at 15°C. Warm to hot seed ripening conditions (20-30°C) increased sensitivity to PHS while cold seed ripening (12-18°C) promoted seed dormancy and increased PHS resistance (Gerjets *et al*., 2010). The effect of temperature could be linked to O_2 availability in the embryo because O_2 solubility reduces with increasing temperature (Hoang *et al*., 2013). It has also been reported that the level of dormancy in wheat increases when seed development takes place under drought conditions (Biddulph *et al*., 2005). In a usually non-dormant genotype of wheat, Biddulph *et al*. (2007) discovered that sustained high temperatures combined with moisture stress caused dormancy. Under certain conditions, after-ripened seeds become dormant again, and environmental factors such as cold temperatures and high humidity can induce this secondary dormancy (Belderok and Habekotte, 1980). Another environmental component that influences seed dormancy and germination during imbibition is light. Blue light, for example, increases ABA concentration and thus dormancy in barley seeds (Gubler *et al*., 2008). Therefore, to enhance genetic inheritance and reduce environmental influences, the effects of temperature, rainfall, and humidity on PHS must be considered while breeding PHS-resistant cultivars.

2.12 Molecular genetic markers

Molecular genetic markers are used to construct genetic maps to examine genetic diversity within and between species and investigate the correlation between genotype and phenotype on the basis of variations in the genome (Batley and Edwards, 2007). In the 19th century, Gregor Mendel's genetic study used phenotypic-based genetic markers. Subsequently, genetic markers based on phenotype enhanced the development of linkage theory, and the isozyme markers were the most popular early marker technology but these markers are limited in number and affected by environmental factors. As a result, they are hardly used nowadays (Koebner and Summers, 2003; Winter and Kahl, 1995). To circumvent the drawbacks of phenotypic-based genetic markers, DNA-based markers have eventually been introduced (Agarwal e*t al*., 2008). Since its introduction, substantial progress has been made in the development and application of molecular markers mainly due to advancements in DNA technologies. Among the ever-evolving DNA-based genetic markers are restriction fragment length polymorphism (RFLP) (Tanksley *et al*., 1989), random amplified polymorphic DNA (RAPD) (Williams *et al.*, 1990), amplified fragment length polymorphism (AFLP) (Vos *et al*., 1995), diversity arrays technology (DArT) (Jaccoud *et al*., 2001), simple sequence repeat (SSR) or microsatellites (Zietkiewicz *et al*., 1994) and single nucleotide polymorphism (SNP) (Marth *et al*., 1999).

RFLP markers are able to discover specific loci in each of the three sub genomes of hexaploid wheat at the same time (Chao *et al*., 1989). However, they are time-consuming and labor-intensive, and therefore they have been less useful in wheat (Gupta *et al*., 1999). Few maps utilizing AFLP markers have been generated in wheat, for instance a wheat molecular map using AFLP and SSR markers was recently constructed in winter wheat (Korzun., 2002). However, AFLP markers were not widely utilised in molecular breeding because of the time-consuming and

labor-intensive detection procedure (Mammadov *et al*., 2012). Subsequently, the development of DArT and RAPD has made the generation of high-density linkage maps possible (Wenzl *et al*., 2004). Owing to the complexity of the wheat genome, low levels of polymorphisms and lack of reproducibility of results, RAPD technology and DArT have not been widely used in genetic studies of wheat (Gupta *et al*., 1999). Before the discovery of SNP markers, SSR or microsatellites have been the most popular molecular markers used in genetic analysis of wheat, mainly because they are abundant, easy to detect, dispersed throughout the genome and show higher levels of polymorphism than RFLP and RAPD markers (Gupta *et al*., 1999). Among all the markers, SNPs are the common source of polymorphisms in plant and animal genomes as they represent variations in the genome at single nucleotide base level (Xu and Crouch, 2008). Moreover, SNPs are highly reliable and cost-effective. The development of high-density SNP arrays in wheat such as the 9K (Cavanagh *et al*., 2013), 90K (Wang *et al*., 2014) and 820K (Winfield *et al*., 2016) gene-related SNP arrays have opened up a horizon of new opportunities for dissecting complex genetic characteristics and therefore advancing marker-assisted breeding. This is because these SNP arrays represent a significant advancement in wheat genotyping, allowing for the collection of comprehensive genotypic data (Akhunov *et al*., 2009). Association mapping to detect associations between phenotypic traits and genetic polymorphisms, population structure analysis and genomic selection are all examples of genome-wide studies wherein SNP markers are widely used (Chen *et al*., 2014, Wang *et al*., 2016). The application of high-density SNP arrays along with the completion of the sequencing and assembly of the whole genome of bread wheat variety Chinese Spring by the International Wheat Genome Sequencing Consortium (IWGSC) is playing critical role in providing a high-resolution link between wheat traits and genetic variations.

2.13 Quantitative trait loci analysis and genetic mapping

The majority traits contributing to yield and quality are complex and regulated by a large number of loci or genes. Quantitative trait loci (QTL) are areas of the genome that include genes associated with polygenic trait (Collard *et al.*, 2005). The two most common approaches of identifying QTLs for complex traits are linkage mapping and association mapping. Both mapping approaches use the fact that recombination divides the genome into small parts that can be linked to phenotypes (Myles *et al*., 2009). Linkage mapping which is also known as traditional mapping uses bi-parental population and relies on the recombination of events occurring during the development of a mapping population, which is often highly expensive and time-consuming. This mapping has low resolution because of limited recombination events and low allele frequency since developing the population involves crossing between two parental lines (Xu *et al*., 2017). Association mapping which is also known as linkage disequilibrium, on the other hand, uses a collection of germplasm having genetic diversity and involves all the historical and evolutionary recombination events that have arisen in the population, resulting in high resolution and greater allele frequency (Zhu *et al*., 2008).

2.13.1 Linkage mapping

The principle behind linkage mapping is the co-segregation of markers with traits due to their tight linkage. The first and most crucial step of linkage mapping is the development of a mapping populations such as F2, back cross (BC), double haploid (DH), recombinant inbred line (RIL) and near-isogenic line (NIL), which are derived from a crossing between two parental genotypes with a contrasting trait of interest. The population will then be phenotyped to examine if the trait of interest segregates under different environmental conditions. After that the population is

genotyped in order to identify DNA markers that exhibit polymorphism between the parents and segregants (Xu *et al*., 2017). The next steps involve linking the phenotypic data of the traits with the genotypic data. The markers that are highly correlated with a phenotypic data are thought to be closely linked to the QTL region controlling the trait, and the linkage analysis is conducted using software packages such as JoinMap, MAPMAKER/EXP, GMENDEL, LINKAGE and Map Manager QTX (Manly *et al*., 2001). The odds ratios are used to assign markers to linkage groups, and a logarithm of odds (LOD) is the most convenient method to describe this ratio (Risch, 1992).

Some of the widely used statistical methods used to identify QTL are single marker analysis (SMA), simple interval mapping (SIM), multiple interval mapping (MIM), composite interval mapping (CIM), Bayesian interval mapping (BIM), expression QTL (eQTL) or protein QTL (pQTL) (Collard *et al.*, 2005). In crops such as maize (Li *et al*., 2010) and barley (Ayoub *et al*., 2002), significant success has been achieved in discovering important QTL and genes for the weight and size of the kernel through linkage mapping. The QTL for agronomically relevant traits such as flowering time in Arabidopsis (Werner *et al*., 2005) and frost tolerance in wheat (Vagujfalvi *et al*., 2003) have also been identified with linkage mapping. Several QTLs for seed dormancy and PHS have been discovered in the major cereal crops including barley (Ullrich *et al*., 2008), rice (Gu *et al*., 2004), sorghum (Lijavetzky *et al*., 2000) and wheat (Kumar *et al*., 2015). The details of PHS associated QTLs identified in wheat through linkage mapping are discussed in the later section of this chapter. In linkage mapping, the number of recombination within the population is quite low and this leads to low mapping resolution, 10-20 cM (Zhu *et al*., 2008). The use of association mapping is an alternative to linkage mapping in overcoming such limitations of linkage mapping (Ross-Ibarra *et al*., 2007).

2.13.2 Association mapping

Association mapping (AM) is an approach that uses linkage disequilibrium (LD) to determine the association between phenotypic variation and genetic polymorphisms in germplasms with a range of genetic diversity (Flint-Garcia *et al*., 2003). The main principle behind association mapping is that LD between loci that are genetically related to one another tends to be preserved over many generations (Neumann *et al*., 2011). Given AM uses populations that have gone through large numbers of recombination events since domestication, historical recombination events between SNPs and QTLs at the population level are used to identify QTLs (Nordborg & Tavaré, 2002). As a result, markers that are strongly linked with the gene of interest can be identified (Flint-Garcia, Thornsberry and Buckler, 2003).

Association mapping is divided into two types, genome-wide association mapping (GWAS) and candidate gene association mapping. GWAS examines the entire genome for random genetic variation using dense genome-wide markers (Rafalski, 2010). On the other hand, candidate gene association examines the relationship between a candidate gene's DNA variation and the trait of interest. In the case of GWAS, greater number of molecular markers are required to obtain high resolution of the association analysis. For example, depending on the genetic diversity and size of the genome used, 5,000 to 750,000 markers could be required to completely cover the entire genome of maize (Flint- Garcia *et al*., 2003). However, in crops such as barley, about 1,000 equally dispersed markers could be enough to achieve excellent genome coverage (Rostoks *et al*., 2006). Despite all its advantages, AM has a few limitations. To begin, knowledge about the genome is required for building SNP arrays and determining where SNPs are located in the genome. Secondly, uncommon variations/rare alleles that are not represented on the SNP array could cause

the phenotype. Furthermore, identification of false positives or negatives as well as the issue of missing heritability have been cited as key challenges in GWAS (Gupta *et al*., 2019).

High-throughput DNA sequencing is being used to circumvent some of these constraints related to AM. Several QTLs for important agronomic traits such as resistance to disease and end-use quality have been identified in wheat using GWAS. For example, a GWAS that involved a heterogeneous group of 94 wheat lines, identified marker-trait relationships for five major agronomical traits including kernel hardness, thousand-kernel weight, protein content, test weight and plant height. Another GWAS that involved the use of DArT markers and studied 20 agronomic traits in a collection of winter wheat identified substantial marker-trait associations for plant height, yield and resistance to disease (Neuman *et al*., 2011). Moreover, SSR and sequence-tagged sites (STS) markers have been used to undertake association mapping for identifying genes resistant to stem rust in the US winter wheat germplasm.

2.13.3 Linkage disequilibrium and association mapping

Linkage disequilibrium is defined as the non-random correlation of alleles at two or more distinct loci (Flint- Garcia *et al*, 2003). The LD between molecular markers and functional loci is used in both linkage and association mapping approaches. The distinction between the two methodologies is that the LD used in linkage mapping is generated by the mating design, whereas the LD used in association mapping is already present in the germplasm collection. The resolution and strength of association mapping are dependent on the degree of LD over the entire genome (Zhang *et al*., 2009). Many factors that influence LD include frequency of the alleles, rate of recombination, size and structure of the population, mutation, mating pattern and admixture (Flint-Garcia *et al*., 2003). As a result, association mapping requires knowledge of population structure and LD through the

entire genome. In plants, reproductive behavior appears to be one of the most critical variables that influence the degree of LD, and LD is more common in species with a high level of inbreeding (Zhang *et al*., 2009). If LD decays over a narrow distance, the resolution will be high, but more markers will be required (Rafalski, 2002). If LD decay is spread out across a large area, mapping resolution will be less, but only fewer markers will be needed. The computation of LD at the start of the association study is critical since a long-range LD (between pairs of loci that are distant from one another or even on separate chromosomes) enhances the likelihood of false association (Stich *et al*., 2006). The disequilibrium coefficient (D') or squared allele frequency correlation (r^2) are the two most commonly used methods to compute LD which exhibits pairwise measurements between loci (Flint-Garcia *et al*., 2003). A scattering plot of r^2 values with genetic/physical distances of all SNP pairs in a genome can be used to visualize LD decay. The value of r^2 provides the information on whether a set of significant SNPs is inherited together and represents the same QTL, or they stand for individual QTLs (Alqudah *et al*., 2020).

2.13.4 Factors influencing genome-wide association study

Various factors including population structure and relatedness, sample size, environmental effects on the phenotype and genotypic error have significant influence on GWAS. These factors are discussed in detail in the following sections.

2.13.4.1 Population structure and familial relatedness

Genome-wide association study uses a diverse set of genotypes from a single species and the structure of a population is defined as how closely the individuals are related to one another (Alqudah *et al*., 2020). It is mainly based on growth habits, geographic origin and some other

factors and this causes genetic variation and has a strong influence on the association study. Correction for population structure in a germplasm collection is crucial for minimizing statistical error in AM, mainly when genetic interactions among genotypic lines are varied (Zhu *et al*., 2008). The presence of a population structure can result in false-positive associations between markers and trait of interest (Hirschhorn and Daly, 2005). False associations can also develop when alleles are present in the original population at an extremely low frequency (Breseghello and Sorrells, 2006). Population structure causes LD between loci that aren't connected physically, resulting in more chances of false associations between genotypic and phenotypic traits, which must be addressed (Crossa *et al*., 2007). As a result, in association analyses, distinguishing LD owing to physical connection from the LD created by population structure is crucial. The STRUCTURE program is a statistical approach used to define population structure and estimate the number of subpopulations within the population, which is referred to as the Q matrix (Pritchard *et al*., 2000). Based on their genetic characteristics, the individuals are then classified into distinct groups. The STRUCTURE and STRUCTURE HARVESTER are used to find the proportion of an individual's genome that originate from multiple estimated populations using a Markov Chain Monte Carlo (MCMC) Bayesian approach.

Since Structure program is time-consuming and computationally intensive, there is another approach as well through which we can estimate the structure of the population. This approach uses EIGENSTRAT method that uses principal component analysis (PCA) to estimate the population structure (Price *et al*., 2006). This method is effective, much faster and less computationally intensive than STRUCTURE. A mixed model method in Trait Analysis by Association, Evolution and Linkage (TASSEL) software is used to correct spurious associations

between the genotypes. It produces a pairwise genetic relatedness matrix (kinship matrix; K) by calculating the levels of relatedness between all pairs of individuals (Yu *et al*., 2006).

Combining the Q matrix obtained from population structure along with the kinship matrix generated from TASSEL as covariates into a unified linear mixed-model approach reduces false-positive associations (Yu *et al*., 2006). By reducing the confounding effects of population structure and kinship matrix, this approach proved to be successful in enhancing the power of marker-trait associations.

2.13.4.2 Sample size

The size of the population is important for achieving the intended results of an association between genotype and phenotype. Individuals in the population could be chosen based on their expected phenotypic and genotypic variance as well as their genetic background, which could include geographical origin or growth habits (Alqudah *et al*., 2020). The power of having significant associations with a greater effect, an acceptable frequency within the population, and overcoming rare variants is improved by increasing the population size. Therefore, a small number of individuals is a limitation that reduces the statistical power of GWAS. GWAS is usually conducted on a large number, ranging from 100 to 500 genotypes (Kumar *et al*., 2012)**.** However, even if the numbers are restricted (less than 100), new cultivars with desirable traits can be developed by combining genotypically diverse and phenotypically superior lines. Several GWAS analyses with population sizes ranging from 60 to 150 have been reported for different traits in various crop species such as flowering time in maize (Remington *et al*., 2001), kernel size and milling quality in wheat (Breseghello and Sorrells, 2006b), and yield and its components in rice (Agrama *et al*., 2007).

2.13.4.3 Environmental effects on the phenotype

Broad-sense heritability is a useful indicator of how much genetic variance contributes to the traits and how much the genotype influences the phenotype (Alqudah *et al*., 2020). Furthermore, GWAS should only evaluate features with moderate to high heritability since low heritability is a restricted factor to identify associations between the phenotype and genotype in GWAS. Heritability is calculated as H = VG/(VG + VE), where VG and VE indicate genetic and environmental variance, respectively (Smith *et al*., 1998). Genotypes that are repeatedly grown over different sites or years may exhibit strong genotype-environment interaction and ultimately lower heritability. Thus, several methods are used for adjusting the phenotypic data such as the best linear unbiased predictor (BLUP) and best linear unbiased estimator (BLUE) methods, which provide good estimates of the phenotypic data when genotype-environment interaction is taken into account. The best linear unbiased prediction (BLUP) is also used to reduce the environmental effect from the analysis (Piepho *et al*., 2008).

2.13.4.4 Genotypic error

Error-free genotyping of the mapping panel is an important criterion for association mapping. An error of 3% or even less can have a significant impact on the precision of LD estimates (Akey *et al*., 2002). This is especially important for SNPs because even within a single experiment the rate of error varies dramatically between various SNP sites (Ingvarsson and Street 2011). However, the genotyping error rate is continuously getting better with the advancement of next-generation sequencing technologies.

2.13.5 Different stages of association mapping

The first step in conducting a GWAS experiment is to choose the population, which includes taking into account of the population's size. Once selected, the population will be subjected to three critical stages (Figure 2.3). Stage I is represented by phenotypic analysis of the association panel for a specific trait or combination of traits and calculating the heritability of the panel. Given that precise phenotyping is required for examining the genotypic-phenotypic relationships, phenotyping should be performed in replicated field trails across multiple locations and years. Stage II of AM is genotyping of the association panel subjected to phenotypic analysis using polymorphic DNA molecular markers (Ibrahim *et al*., 2020). The most popular method for genotyping is genotyping by sequencing since it produces a lot of SNPs that cover the crop genome at a low cost. Population structure, kinship matrix and LD should be assessed before doing GWAS in order to determine the optimum GWAS model. The general linear model (GLM) and the mixed linear model (MLM) are two statistical models that are frequently used for GWAS. The population structure is not taken into consideration by the GLM while the MLM incorporates population structure and kinship matrix into its model. Stage III of AM is analysis of the association between the phenotypic and genotypic data using an appropriate software like TASSEL in order to identify markers linked with a particular trait of interest. The false discovery rate (FDR), Bonferroni correction (BC) or 0.1 percentile method can be used to detect the significant markers that cross a particular threshold (Chan *et al*., 2010; Pasam *et al*., 2012).

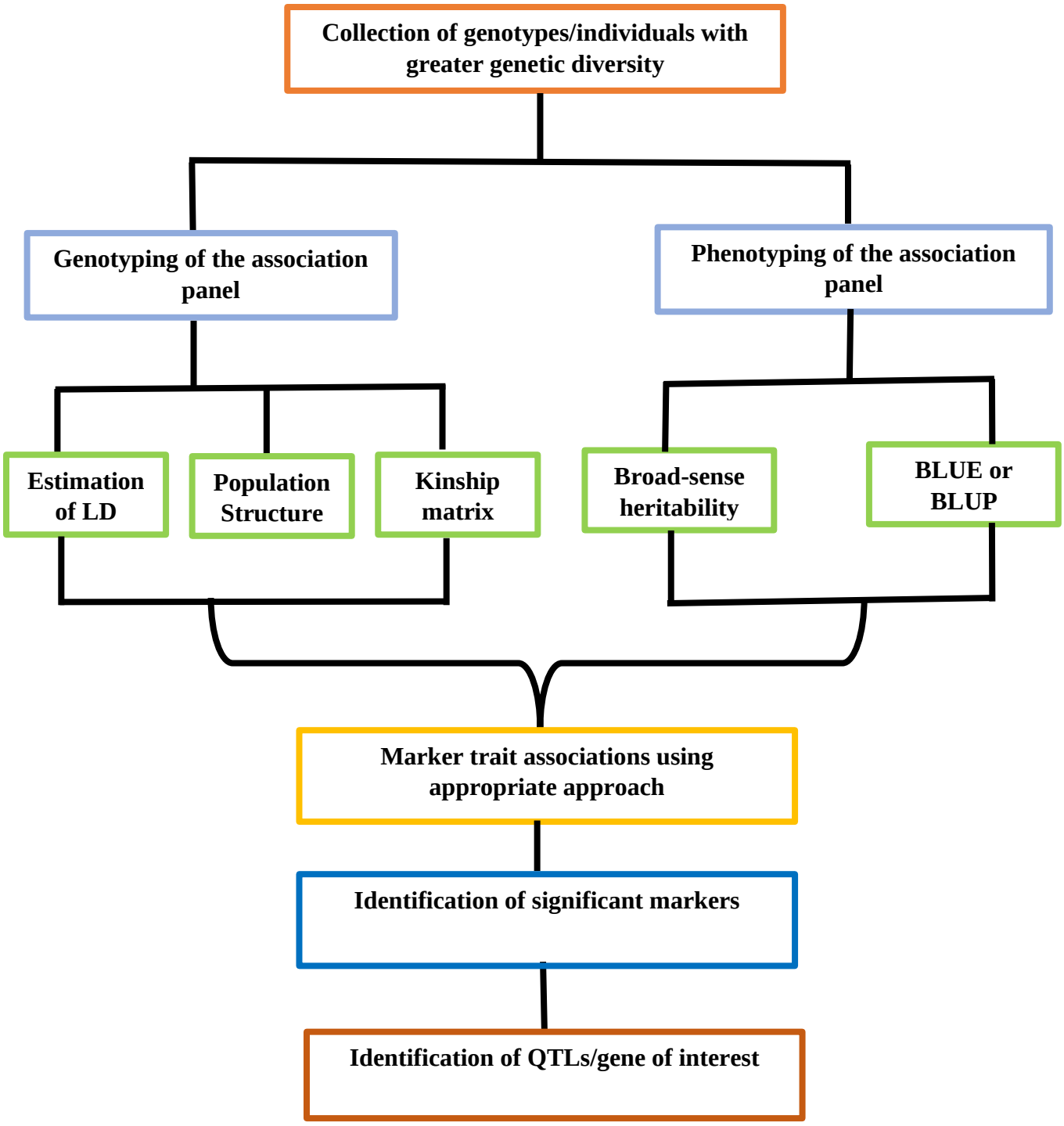


Figure 2.3 Schematic representation of the different stages of association mapping.

2.13.6 Merits and challenges of association mapping

Association mapping has a few advantages over bi-parental mapping. Its key advantage is that it has high resolution and less time-consuming as compared to linkage mapping as it eliminates the need for the construction of mapping population (Zhu *et al*., 2008). Furthermore, genetic diversity in complex traits may be measured and screened across a considerably broader and more representative gene pool (Neumann *et al*., 2011; Zhao *et al*., 2011), and more traits can be studied in the population using the same genotypic data (Zhao *et al*., 2011).

Although AM has a number of advantages, it has a few limitations too. Firstly, due to the possibility of false associations between the marker and trait of interest in the germplasm, the procedures required to perform statistical analysis of the data to correct population structure are more complicated (Neumann *et al*., 2011). Secondly, multiple testing such as bonferroni correction (Holm, 1979) and FDR (Benjamini and Hochberg, 1995) may result in a genome-wide significance threshold that is too stringent. Moreover, the ability to find significant marker-trait relationships is further influenced by the types of phenotypic data, size of the sample and heritability in association mapping (Barabaschi *et al*., 2016). As a result, multi-parent cross populations such as Nested Association Mapping (NAM), Multi-parent Advanced Generation Inter-Crosses (MAGIC), and Random-Open-parent Association Mapping (ROAM) populations are being used for mapping to overcome false associations between the trait of interest and genotypes and thereby enhance the chances of identifying rare alleles (Yu and Buckler, 2006; Dell'Acqua *et al*., 2015; Xiao *et al*., 2016).

2.13.7 Family based association mapping

Several studies have combined linkage and association mapping, resulting in a new approach known as "nested association mapping," which reduces false correlations between the genotype and phenotype by taking population structure into account (Alqudah *et al.,* 2020). The NAM population was first used in maize (Yu and Buckler, 2006), with 5,000 RILs derived by breeding a homozygous line with 25 lines (McMullen *et al*., 2009). Following the success achieved in maize, NAM populations were developed for genetic mapping/studies of different quantitative traits in other crop species such as rice, wheat, barley, soybean and sorghum (Maurer *et al*., 2015; Schmutzer *et al*., 2015; Bajgain *et al*., 2016; Bouchet *et al*., 2017; Fragoso *et al*., 2017). Intercrossing multiple genetically different parents such as MAGIC can improve allelic variation within a mapping population (Sannemann *et al*., 2015). MAGIC populations, in contrast to other multiparent populations, include intermating several inbred founders for multiple generations prior to the formation of inbred lines, which greatly enhances the precision of QTL detection. The first MAGIC population was developed in Arabidopsis and this population contained 527 lines obtained from a diverse panel of 19 founders (Kover *et al*. 2009). In wheat, MAGIC populations have been used to identify QTL for hectoliter weight (weight per unit volume) and plant height (Huang *et al*., 2012). Another study in wheat involved the development of a MAGIC population using four different wheat types; Yitpi, Chara, Baxter, and Westonia, and analysis of this population led to the identification of seed dormancy gene *PM19* on Chromosome 4 A (Barreo *et al*., 2015). The ROAM population, which combines multiple existing RIL populations, has been used in crop species such as maize to study 10 above-ground plant architecture traits (Pan *et al*., 2016; Xiao *et al*., 2016). The use of such multiparent population will undoubtedly facilitate the

identification of novel causal genes for dormancy QTLs in wheat and therefore enhance the development of PHS-resistant wheat cultivars.

2.13.8 Genetics of pre-harvest sprouting resistance in wheat

Pre-harvest sprouting is a complex trait affected by the genotype and environment. Therefore, molecular markers are used as an effective tool to identify the genetic cause of PHS resistance since markers are not affected by the changing environment. Developing PHS resistant cultivars of wheat is crucial to minimize the negative effects of PHS on wheat yield and quality (Humphreys and Noll, 2002). As a result, PHS resistance has become a key goal in wheat breeding programs around the world. However, the development of such cultivars requires identifying genomic regions or genes responsible for resistance to PHS.

The first report on QTL for seed dormancy in barley and wheat was documented in the early 1990s (Anderson *et al.*,1993, Ullrich *et al.*,1993). Since then QTLs for seed dormancy have been identified on all chromosomes of barley and wheat (Flintham *et al.*, 2002; Kulwal *et al.*, 2010; Mares and Mrva, 2014). Over 110 quantitative trait loci linked to resistance to PHS have been reported in wheat (Zhou *et al*., 2017). Using bi-parental mapping populations, several QTLs for PHS have been discovered on all 21 wheat chromosomes (Mares and Mrva, 2014). Major QTLs for PHS have been found on chromosomes 3A (Osa *et al*., 2003; Kulwal *et al*., 2005; Mori *et al*., 2005; Liu *et al*., 2008; Tyagi and Gupta, 2012); 2BS (Munkvold *et al*., 2009; Somyong *et al*., 2014); and 4AL (Torada *et al*., 2005; Chen *et al*., 2008; Ogbonnay *et al*., 2008; Tyagi and Gupta, 2012; Lin *et al*., 2015).

Using a Yangxiaomai/Zhongyou RIL population, a previous study identified a QTL associated with germination index at the at the TaSdr-A1 locus on chromosome 2 (Zhang *et al*.,

2014). Subsequently, the *TaSdr* genes, which are orthologs to the rice *Sdr4* gene, were cloned, and these genes have been reported as important regulators of seed dormancy and PHS in wheat (Zhang *et al*., 2014; Zhang *et al*., 2017). QTLs for PHS resistance and dormancy have been found on both the short and long arms of chromosome 3A. A significant QTL from Zenkoujikomugi, a Japanese wheat variety, explained 23–38 % of the phenotypic diversity on the short arm of 3A using a red by red grained wheat cross-derived RIL population (Osa *et al*., 2003). A study that used RILs developed from a cross between SPR8198 (PHS resistant) and HD2329 (PHS susceptible) parents identified a significant QTL associated with PHS, which is designated as QPhs.ccsu-3A.1 and explained up to 78 percent of the phenotypic variation across six trial environments (Kulwal *et al*., 2005). The long arm of group 3 chromosomes contains the red grain color gene designated as *R-1*, and this gene encodes a *Myb* transcription factor (*TaMyb10*). The *Tamyb10* gene is known to control grain colour and dormancy in wheat through regulating the accumulation of anthocyanin and ABA level respectively (Groos *et al*., 2002). Group 3 chromosomes also contain the viviparous gene, *TaVp1*, which has been implicated to have a significant role in controlling seed dormancy (Nakamura and Toyama, 2001; Groos *et al*., 2002). Furthermore, Mother of *FT* and *TFL1* (*MFT*) has been identified as a causal gene for QPhs.ocs-3A.1 (Nakamura *et al.*, 2011).

Using numerous parental combinations of wheat, a number of genetic studies have identified a significant QTL for dormancy, designated as *Phs1*, on 4A chromosome (Flintham *et al*., 2002; Mares *et al*., 2005; Torada *et al*., 2005) and mitogen-activated protein kinase kinase 3 (*MKK3*) was recently identified as the causal gene for *Phs1* (Torada *et al*. 2016). Two genes encoding the ABA-inducible Plasma Membrane 19 proteins, *PM19-A1* and *PM19-A2*, have been identified through transcriptomic study as other candidate causal genes for *Phs-A1 by* (Barrero *et al*., 2015). Using comparative genomics, *MKK3* was detected close to *PM19* (Torada *et al*., 2016).

Genetic analysis of DH population produced by crossing AC Domain and Haruyutaka hybrid identified three QTLs on group 4 chromosomes of wheat (Kato *et al*., 2001). One of the three QTLs, QPhs.ocs-4B.2, which is located in the telomere region of 4 BL, was homoeologous to a PHS QTL on 4D. Under field conditions, Kumar *et al*. (2015) identified a significant QTL, QPhs.spa-4B, which accounts for 35–60% of the phenotypic variance. In addition, Qphs.pseru-4B.1, a minor QTL on 4B for both PHS resistance and seed dormancy, was identified by Lin *et al*. (2015) using composite interval mapping. This QTL accounts for 6.3–8.7% of the phenotypic variance.

The mapping resolution from genome-wide association mapping studies in wheat has been hindered as compared to that of maize, Arabidopsis and rice due to the absence of a high-quality reference genome sequence and the availability of a small number of mapped SNPs (Allen *et al*., 2013). As a result, SSR markers or DArT markers have been utilised in the majority of genome-wide association mapping studies of wheat (Breseg-hello and Sorrells, 2006; Tommasini *et al*., 2007; Reif *et al*., 2011). Nonetheless, several genome-wide association studies (GWAS) have identified possible PHS resistance loci. For example, Jaiswal *et al* (2012) used diverse wheat accessions and SSR markers to identify six QTLs (QPhs.ccsu-2A, QPhs.ccsu-2B.2, QGi.crc-3B/Qphs.ccsu-3B.2, QGi.crc-3D, QPhs) linked with PHS through the use of DArT markers. Kulwal *et al* (2012) performed association mapping of white winter wheat accessions and discovered a previously documented significant QTL for PHS on chromosome 4AL, along with a novel QTL on 7BS chromosome. Association mapping of Chinese wheat founder parents led to the identification of *vp1D* and *AIP2-1* genes that are linked with PHS resistance (Yu *et al*., 2017). It has also been shown previously that PHS resistance is associated with the germination-related gene *vp1D* (Bailey *et al*., 1999; Kulwal *et al*., 2005; Imtiaz *et al*., 2008). The wheat *AIP2s* may

play significant roles in seed germination and PHS by negatively regulating the ABA signaling pathway (Gao *et al*., 2014). Overall, through analysis of germplasm lines that exhibit diversity in seed dormancy and PHS resistance, GWAS has become an effective approach for the identification of marker-trait relationships and the subsequent localisation of novel loci.

3.0 GENETIC STUDY OF PRE-HARVEST SPROUTING IN SPRING WHEAT

Contribution Statement

This thesis project part of a collaborative project between the labs of my supervisor, Dr. Belay Ayele, at the University of Manitoba and Dr. Dean Spaner lab at the University of Alberta. The plant materials and the genotypic data used for the study are provided by Dr. Spaner's group. Dr. Spaner's group and Dr. Curt McCartney's group at Agriculture and Agri-Food Canada were also involved in managing the field trails conducted in Alberta and Morden (Manitoba), respectively. Ramanpreet Ramanpreet managed the greenhouse trial and the field trial at The Point (University of Manitoba, Fort Gary Campus), performed all the required experiments, analyzed the data and wrote the first draft of the manuscript.

Abstract

Wheat (*Triticum aestivum* L.) is an important cereal crop that provides ~20% of the daily global calorie and protein intake. However, wheat production is negatively affected by pre-harvest sprouting (PHS), which refers to the germination of seeds on the parent plant due to wet and moist conditions before harvest. It causes significant yield and quality loss. The incidence of PHS depends upon the level of dormancy, which is defined as an adaptive trait that blocks the germination of seeds under optimal environmental conditions, in which a low level of seed dormancy leads to susceptibility to PHS. To identify genomic regions and the associated candidate genes that are important for PHS resistance, this study performed genome-wide association analyses of 194 diverse wheat lines grown in six different environments. The mapping panel was genotyped using a 90k Illumina iSelect SNP array and the seeds of each wheat line were characterized for their germination index and the data obtained showed the presence of significant variation in dormancy levels among the wheat lines. After filtering the genotypic data, a total of 18,672 markers were used in mixed linear model for analysis of marker-trait association, and a total of 19 significant markers were identified with a *P*-value threshold of 2.938×10^{-4}. The significant markers were located on chromosomes 2B, 3B, 4A and 5B representing five QTLs associated with PHS. The phenotypic variation explained by these markers ranges from 7.6% to 25.8%. The significant SNPs associated with these QTLs are linked to 11 genes, which are considered as candidate genes controlling PHS. In conclusion, the findings of the study have the potential to be applied for marker assisted breeding of PHS resistance in spring wheat.

3.1 Introduction

Wheat is the most extensively grown cereal crop under various agro-climatic conditions. Breads, cookies, pasta, biscuits, cakes, noodles, and morning cereals are just a few of the staple foods made from wheat (Olaerts *et al*., 2018). However, due to climate change, key wheat-growing regions worldwide have recently faced intense and unpredictable weather conditions that affect its productivity. Pre-harvest sprouting (PHS), a phenomenon in which seeds germinate while intact on the spike or on the mother plant before harvest due to rainfall and high humidity is one of the main factors that limit wheat production (Wang *et al*., 2019). The market value of grains is substantially decreased by sprouting, and the degree of sprouting determines whether the grain is suitable for human food (Cunha *et al*., 2004).

Pre-harvest sprouting is associated with as a lack of sufficient dormancy level during the maturity of seeds. Thus, one of the key elements influencing PHS is seed dormancy, a trait that blocks germination under favourable conditions (Kumar *et al*., 2015). The presence of adequate level of dormancy reduces PHS damage in wheat and other cereal crops (Rodriguez *et al*., 2015). Among the different factors regulating seed dormancy and germination in wheat and other species are plant hormones, mainly abscisic acid and gibberellin, which are known to promote seed germination and dormancy, respectively (Kucera *et al*., 2005; Finkelstein *et al*., 2008).

Quantitative traits (QTs) have been divided into simple QTs and complex QTs, with the latter being governed by a large number of genes significantly impacted by epistatic and environmental interactions. Pre-harvest sprouting is an example of an important grain quality trait in wheat that is complex and quantitative in nature (Roy *et al*., 1999; Tuberosa and Salvi, 2004; Kulwal *et al*., 2005, 2012; Mori *et al*., 2005; Kumar *et al*., 2006). Linkage analysis and association mapping (AM) are the two main approaches used to analyze the genetic architecture of such

complicated traits. The discovery of segregating QTL using the linkage method has been proven to be successful, but the QTL resolution is typically low and only a small number of alleles are sampled. In addition, developing the mapping population takes more time, effort and money.

Where large segregating populations are not available, association mapping is an alternate approach to uncover marker-trait relationships and it has been widely applied in a number of plant species (Remington *et al*., 2001; Breseghello and Sorrels, 2006a; Kump *et al*., 2011; Kulwal *et al*., 2012). Higher mapping resolution and the existence of a considerably greater range of variance in the traits of interest in the unrelated accessions that comprise the populations are just two of the advantages association mapping has over interval mapping (Buckler and Thornsberry, 2002; Yu and Buckler, 2006). Although structured populations are always used in association mapping studies as compared to traditional linkage-based QTL mapping studies, it gives a higher proportion of spurious associations. However, the impact of structured relationships within the population may be reduced using various statistical approaches that have become available (Pritchard *et al*., 2000; Price *et al*., 2006; Zhou and Stephens, 2012). A candidate gene-based approach (Remington *et al*., 2001; Stich *et al*., 2008) and whole genome scans (Rafalski, 2002; Kulwal *et al*., 2012; Mir *et al*., 2012) are the two methods widely used in association mapping studies to identify marker-trait associations.

Nearly all bread wheat chromosomes have been reported to contain genes and QTL underlying PHS resistance or seed dormancy with QTLs on 3A, 3B, and 4AL being detected consistently in populations exhibiting different genetic backgrounds under diverse environmental conditions (Anderson *et al*., 1993; Mares and Mrva, 2001; Mares *et al*., 2005; Lohwaser *et al*., 2005; Kumar *et al*., 2015). Previous studies have consistently detected the QTL on group 4 chromosome (prominently 4A) thus, it is considered as a major genetic component for PHS

resistance and dormancy (Kato *et al*., 2001; Mori *et al*., 2005; Ogbonnaya *et al*., 2008; Singh *et al*., 2010; Lohwasser *et al*., 2005; Barrero *et al*., 2015; Torada *et al*., 2016).

There are several genes that have been identified and linked to PHS tolerance in wheat. Among these genes, a map-based approach identified a *mitogen-activated protein kinase 3* (*MKK3*) as a causal gene for phs1 QTL that is associated with seed dormancy and located on chromosome 4A (Torada *et al*., 2016). A previous study has also identified *TaPHS1* as a gene underlying a QTL associated with PHS resistance, which is located on the short arm of chromosome 3A (Liu *et al*. 2013). This gene has been reported to be a homolog of the *mother of flowering time* (*MFT*) gene. In line with this result, *MFT* has been identified as a gene underlying a seed dormancy QTL on chromosome 3A (Nakamura *et al*., 2011). In addition, *TaSdr-B1*, which is located on chromosome 2B has been shown to be crucial in controlling seed dormancy (Zhang *et al*., 2017; Zhang *et al*., 2014). The *viviparous 1* (*Vp1*) homolog of wheat (*TaVP1*) is also extensively studied with respect to its role in seed dormancy and pre-harvest sprouting resistance (Yang *et al*., 2007; Chang *et al*., 2010; Nakamura and Toyama, 2001; Xia *et al*., 2008; McKibbin *et al*., 2002). Recent developments in genome sequencing technologies at the chromosomal or whole genome level (Clavijo *et al*., 2017; IWGSC, 2018) is opening up new possibilities for the identification of candidate genes underlying QTLs controlling PHS resistance.

The purpose of this study was to identify loci and candidate genes involved in the control of seed dormancy and PHS using genome-wide association mapping of a diverse collection of spring wheat grown over different environments.

3.2 Material and Methods

3.2.1 Plant material and trial environment

In this study, 194 diverse spring wheat genotypes were used; ~80% of the genotypes are registered as cultivars in western Canada while ~20% of the genotypes represent pre-registration lines. These genotypes were grown in different environments in replicated field trials in different years such as Morden, MB (2016 and 2017), Edmonton, AB (2021) under two different growth conditions, and at the Point, University of Manitoba Field station, Winnipeg, MB (2021), and in a greenhouse at the Crop Technology Centre, University of Manitoba, Winnipeg, MB (2021).

For the greenhouse experiment, fifteen seeds from each genotype were imbibed with 7 ml distilled water between filter papers in a Petri-plate. The plates were then kept under darkness at room temperature for three days. On the fourth day, seedlings were transplanted into one-gallon plastic pots (two seeds/pot) filled with LA sunshine mix (Sungro Horticulture, Bellevue, WA, USA) and slow-release fertilizer (ACER®nt 13-12-12 consisting of 13% N, 12% P_2O_5, 12% K_2O) was added to it. The study was carried out in a randomized complete block design with three replications for each genotype. The transplanted pots were then grown in the greenhouse at 22/18°C (day/night temperatures) and 16/8 h photoperiod at the Crop Technology Centre, University of Manitoba. The plants were well watered regularly and after 30 days of planting, they were supplemented with N-P-K (20:20:20) fertilizers.

The five field trials were also carried out in a randomized complete block design with three replications. For both greenhouse-grown as well as field plants, the spikes were harvested at maturity when the peduncle of most of the spikes became yellow and the seeds were also tough sufficient to dent with a fingernail. The seeds were then collected and kept at room temperature for two weeks before being stored at -20°C until needed.

3.2.2 Assessment of dormancy

Mature seeds from all six trials were characterized for their germination index (GI) by performing germination tests (Gao *et al*., 2012). Fifty seeds from each genotype were imbibed between filter papers using Petri-plates at room temperature and in the darkness. Germination was examined daily for 15 days, and a seed was considered germinated when the coleorhiza arose beyond the seed coat. The GI was calculated using the following formula:

$$GI= [(15\times g1) + (14\times g2) + (13\times g3) \ldots.1\times g15)] / (15\times n)$$

Where g is the number of seeds germinated on the first, second, third… and 15^{th} day of imbibition and n denotes the total number of seeds imbibed.

3.2.2.1 Phenotypic data analysis

The phenotypic data analysis including analysis of variance (ANOVA), basic statistics (maximum, minimum and mean GI values) and frequency distribution was performed using SPSS (V.26) statistics software (https://www.ibm.com/analytics/spss-statistics-software). The GLM was used to investigate changes in GI between genotypes and environments as well as the presence of genotype-environment interactions using a univariate model and type III sum of squares to explain the missing data. This was performed based on the following equation (Imtiaz *et al*., 2008; Shao *et al*., 2018)

$$GLMyijkl = \mu + Gi + Ej + Bk(j) + GEij + eijkl$$

Where Gi, Ej and Bk(j) represent the genotype, environment and blocking effect, respectively. GEij represents the genotype-environment interaction and eijkl is the random error.

At the 5% level of significance, the mean values were separated using the F-protected least significant difference (LSD) value. The multi-environment experiment analysis using R version 6.0 in META_R (http://data.cimmyt.org/dvn/dv/cimmytswdvn) was used to determine broad-sense heritability (H^2), and the correlation between the environments (Alvarado *et al*., 2020). Heritability was calculated using the following formula (Imtiaz *et al*., 2008; Shao *et al*., 2018):

$$H^2 = Vg / (Vg + Vge / e + Ve / re)$$

Where Vg, Vge, Ve represents the genotype, genotype-environment and error variance, respectively; r and e denote the number of replications and number of environments.

Moreover, to account for environmental variances and more exact estimates of phenotypic values, best linear unbiased predictors (BLUEs) estimates were generated. The BLUEs were calculated by considering genotypes as a fixed effect while environment and replication as random effects. The resulting BLUEs were used in the association analysis as the mean values for the dormancy trait.

3.2.3 Genotyping

Genotyping of the wheat lines was performed using the 90 K Illumina I select genotyping array (Wang *et al*., 2014). A total of 20,205 SNP markers were derived and the SNP markers were filtered to remove those with more than 20% missing values and a minor allele frequency (MAF) of less than 5%. After removing markers with heterozygosity, a total of 18,672 markers remained

for final analyses. The genotypic information was available for 194 wheat genotypes. Therefore, to identify marker-trait associations, the 194 wheat genotypes were scanned with the 18,672 SNP markers using TASSEL v5.2.60 3.2.4 (http://www. maizegenetics.net/).

3.2.4 Estimation of linkage disequilibrium

Linkage disequilibrium (LD), which identifies significant differences between the SNPs, was computed using TASSEL by calculating the squared allele frequency correlation (r^2) for all pairwise comparisons of SNP distances. TASSEL was used to import the genotype data, and a full matrix LD analysis was performed with the default parameters. The LD of the whole genome, as well as the subgenomes (A, B, and D) was computed. LD decay scatter plots were produced using r^2 and the genetic map distance between markers according to the Hill and Weir methods (Hill and Weir, 1988). The 95^{th} percentile in the distribution of r^2 of the selected loci was used to calculate the threshold value of r^2 and this threshold value is used as a cutoff point to distinguish between the LD that is mainly due to the genetic linkage and the LD that could be due to other factors. A non-linear regression line was fitted to r^2 data to visualize the link between LD decay and genetic distance by using R v.4.1.1 software (R: The R Project for Statistical Computing; https://www.r-project.org/).

3.2.5 Population Structure and Kinship

In order to assess the structure of the population and estimate the real number of subpopulations (K), a clustering method using the admixture model of STRUCTURE software 2.3.4 was employed (Pritchard *et al*., 2000). By using 245 unlinked SNPs, the population structure of the 194 wheat genotypes was studied. For each hypothetical number of subpopulations (K) from one

to seven, 25 independent runs were performed, with a burn-in time of 100,000 iterations followed by 100,000 Markov Chain Monte Carlo iterations to get an accurate parameter estimate. Using the ΔK method and the STRUCTURE HARVESTER programme (http://taylor0.biology.ucla.edu./struct_harvest/), the number of subpopulations most likely to exist was calculated (Evanno *et al*., 2005); ΔK was plotted against the number of sub-groups K.

Furthermore, a principal component analysis (PCA) and a neighbor-joining phylogenetic tree analysis were performed in TASSEL to analyze and visualize the population structure. TASSEL v5.2.39 was also used to calculate an identity-by-state (IBS)-based relative kinship matrix (K-matrix) between genotypes as a measure of relatedness (Bradbury et al. 2007). To avoid false positives in the linear model, the principal components (PCs), Q, and K matrix were used as covariates.

3.2.6 Association analysis for dormancy/ PHS

The general linear model (GLM) and the mixed linear model (MLM) in TASSEL were used for association analysis (Bradbury *et al*., 2007). These models were used to measure correlations between genetic markers and the trait of interest. Mean BLUE values of GI of the phenotypic data were used for the association with the genotypic data. Two GLM models with principal components (GLM-PCA) and Q matrix (GLM-Q) as covariates, and two MLM models with K matrix as an additional covariate (MLM-PCA+K) and (MLM-Q+K) were used. Quantile-quantile (QQ) and Manhattan plots of different models and the significant markers from the best models were analyzed. In Manhattan plots, *P* values are transformed into -log *P* values and one can easily identify the regions of the significant markers that cross a particular threshold. Bonferroni correction (α=1%) and 0.1% distribution of the *P* values were used to remove spurious associations

and declare the significant markers. The threshold with Bonferroni correction was too stringent, so the 0.1% *P*-value distribution was chosen (Chan *et al*., 2010). The significant markers were selected on the basis of the *P*-value threshold of 2.99×10^{-4} with a corresponding -log *P* value of 3.52.

3.2.7 Identification of candidate gene

The significant SNPs were screened for their chromosome locations (genetic positions) using the wheat consensus genetic map (Wang *et al*., 2014). Physical positions were determined using the Chinese Spring reference genome sequence (https://urgi.versailles.inra.fr/blast_iwgsc/blast.php, IWGSC 2018). The nucleotide sequence for the significant markers (Wang *et al*., 2014) was blast searched against the IWGSC RefSeq v2.0 database. The sequences of these markers were searched against the Chinese Spring IWGSC RefSeq v2.1 genome using the basic local alignment search tool (BLAST) (https://blast.ncbi.nlm.nih.gov/Blast.cgi) in GrainGenes (GrainGenes | A Database for Triticeae and Avena (usda.gov) and Ensembl Plants databases (http://uswest.ensembl.org/index.htmL) to identify candidate genes linked with the significant markers.

3.3 Results

3.3.1 Evaluation of phenotypic variation and heritability trait

Germination index (GI) of the association mapping panel was evaluated over six environments. Our results revealed significant variations in GI among the genotypes in each of the trial

environments (Table 3.1). The GI value varied from 0.07 to 0.98 with a mean GI of 0.54 (MOR16), 0.14 to 0.99 with a mean GI of 0.71 (MOR17), 0.01 to 0.98 with a mean GI of 0.64 (GH21), 0.22 to 0.76 with a mean GI of 0.48 (Point21), 0.08 to 0.89 with a mean GI of 0.55 (EDM CONV21) and 0.10 to 0.95 with a mean GI of 0.57 (EDM ORG21). MOR17 showed the highest mean GI value of 0.71 followed by GH21 with a mean GI of 0.64. The lowest germination index was observed in Point21 with a mean GI of 0.48. The minimum and maximum GI for the dataset pooled from all the environments were 0.26 and 0.83 while the mean GI value was 0.57.

Broad sense heritability estimates were also calculated in all the trial environments as well as for the dataset pooled from all environments (Table 3.1). Our data showed the highest heritability of 0.97 in MOR16, followed by MOR17 and EDM ORG21 with heritability of 0.90. The lowest heritability of 0.60 was observed at Point21. The overall heritability across all the environments was 0.81. These results indicate that the GI variation observed among the genotypes is mostly due to genetic not environment effects.

Table 3.1 Germination index, skewness, kurtosis and heritability of the mapping population in the different trial environments.

Environment[a]	Minimum GI	Maximum GI	Mean GI	Skewness	Kurtosis	Heritability
MOR16	0.07	0.98	0.54	-0.07	-0.93	0.97
MOR17	0.14	0.99	0.71	-0.84	-0.33	0.90
GH21	0.01	0.98	0.64	-0.70	-0.45	0.88
Point21	0.22	0.76	0.48	-0.01	-1.00	0.60
EDM CONV21	0.08	0.89	0.55	-0.32	-0.08	0.88
EDM ORG21	0.10	0.95	0.57	-0.29	0.07	0.90
Pooled	0.26	0.83	0.57	-0.43	-0.32	0.81

[a]MOR16, Morden 2016; MOR17, Morden 2017; GH21, Winnipeg Greenhouse 2021; Point21, Point 2021 (University of Manitoba Point Field Station); EDM CONV21, Edmonton Conventional 2021; EDM ORG21, Edmonton Organic 2021

Frequency distribution was also calculated for each environment as well as for the dataset pooled from all the trial environments (Figure 3.1 and 3.2). Our analysis showed that the kurtosis and skewness values are within the acceptable range of normal distribution in all trial environments except in MOR17. Thus, the frequency distribution is normal, indicating varied levels of dormancy in all the environments except MOR17 in which the frequency distribution is skewed towards low level of dormancy and PHS (Figure 3.1B).

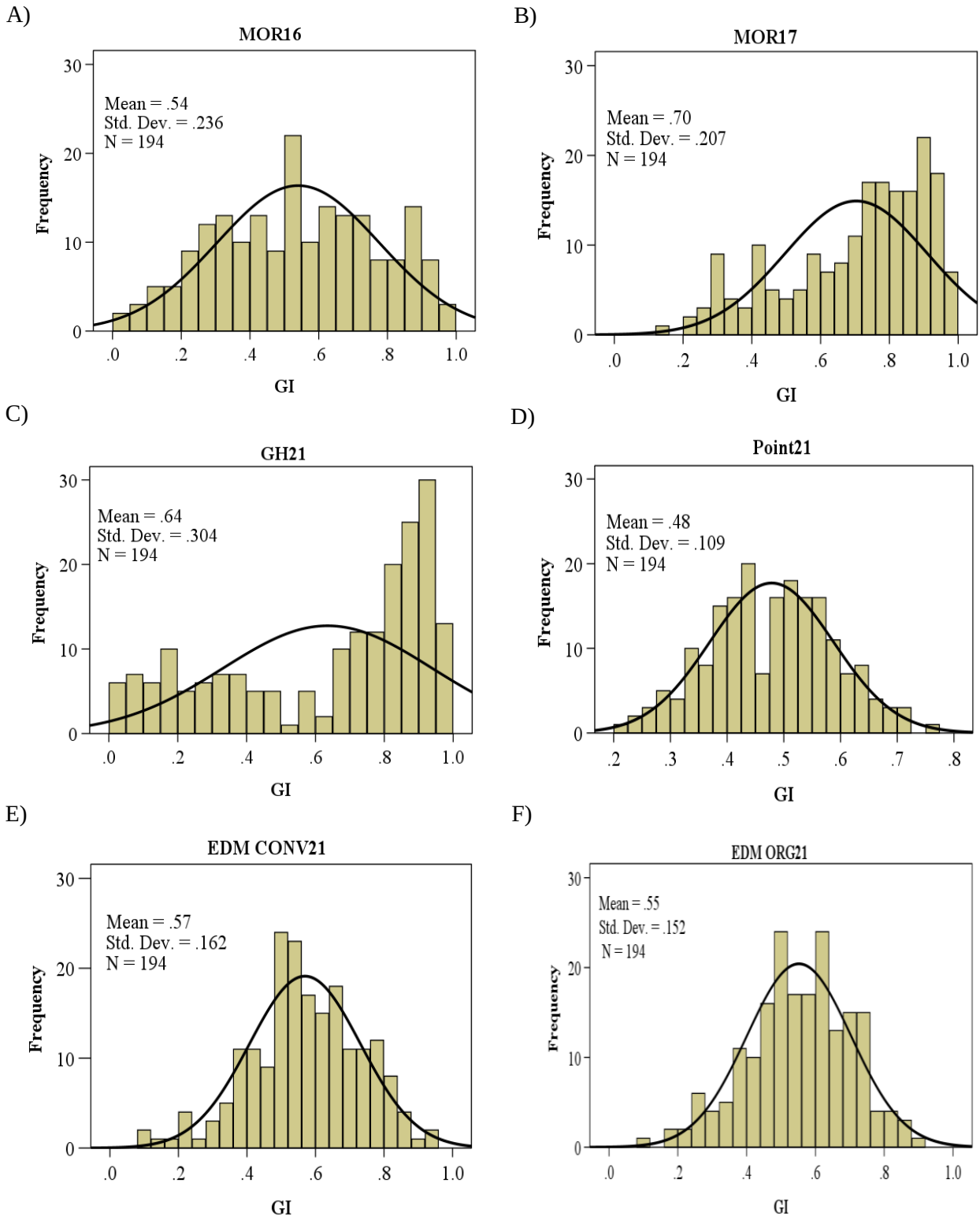


Figure 3.1 Frequency distribution of germination index in each environment (a) MOR16, (b) MOR17, (c) GH21, (d) Point21, (e) Edm Conv21, (f) Edm Org21. The black curve represents a normal distribution. Indicated on each graph are, the overall mean, standard deviation and number of genotypes examined in each trail environment.

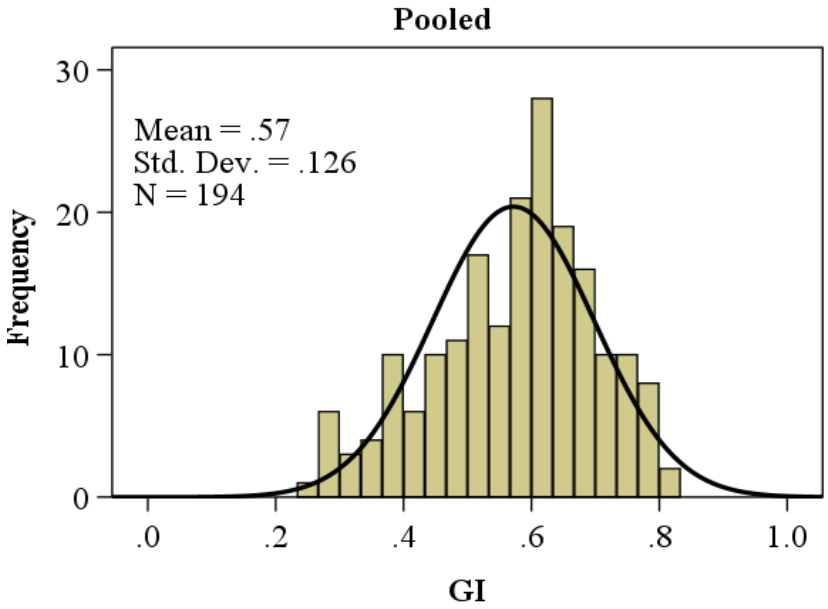


Figure 3.2 Frequency distribution of germination index pooled from all trial environments. The black curve represents a normal distribution. Indicated on the graph are the mean and standard deviation of the dataset pooled from all trial environments, and number of genotypes examined in each trial environment.

3.3.2 Phenotypic correlations among different environments

Correlation among the different trial environments for germination index was analyzed for the association mapping panel. Our analysis showed positive and significant correlation of GI for all trial environments (Table 3.2). The correlation between EDM ORG21 and EDM CON21 was the highest (r = 0.58).

Table 3.2 Correlation among the different trial environments for germination index in the association mapping panel.

Environment[a]	MOR16	MOR17	GH21	Point21	EDM CONV21
MOR17	0.55***				
GH21	0.25***	0.48***			
Point21	0.05***	0.11***	0.03***		
EDM CONV21	0.36***	0.42***	0.10***	0.07***	
EDM ORG21	0.57***	0.53***	0.19***	0.07***	0.58***

[a]Environment names are as shown in table 1
***Significance at P < 0.001

3.3.3 Analysis of variance

Analysis of variance (ANOVA) was performed to determine statistically significant differences in GI among the genotypes of the association mapping panel, the trial environments, as well as interaction between genotype and environment. Our ANOVA analysis revealed statistically significant difference for genotypes, environments, and genotype and environment interactions (Table 3.3).

Table 3.3 Analysis of variance of germination index in the association mapping panel.

Source	DF	Type III SS	Mean Square	Expected Mean Square	F	P>F
Environment (E)	5	21.15	4.23		268.89	<0.001
Genotype (G)	193	54.05	0.28	Ve+rVge+reVg	17.79	<0.001
G*E	944	84.17	0.09	Ve+rVge	5.66	<0.001
Error	2286	35.98	0.02	Ve		
Total	3429	1310.75				

*** significant at P < 0.001, ns not significant
DF is the degrees of freedom; SS is the sum of squares; where Vg is the genotypic variance; Vge is the variance of genotype by the environment; Ve is the error variance; r is the number of replications; e is the number of environments; P is the probability

3.3.4 Analysis of population structure

Analysis of the population structure was conducted using three different methods including the Bayesian cluster approach, principal component analysis, and phylogenetic tree approach. Using the genotypic data of the association mapping panel, the Bayesian cluster analysis method in STRUCTURE was used to determine the number of genetic clusters (K-value) by plotting sub-groups 'K' against delta K using STRUCTURE HARVESTER. Our result showed a sharp peak at K = 2 (Figure 3.3A), indicating that the association panel is divided into two sub-populations. The first subpopulation, designated as group Q1, contains 76 genotypes while the second subpopulation, designated as group Q2, contains 118 genotypes. Structure Plot v2.0 (Earl, 2012) was used to create the structure bar plot for the optimum number of clusters and this analysis also indicated that the association mapping panel is divided into two subpopulations (Figure 3.3B).

As an alternative approach, principal component analysis (PCA) was also performed to determine the number of subpopulations. A scatter plot was produced between the Q matrix generated in STRUCTURE and the individuals were assigned to their respective clusters, and the first two principal components, PC1 and PC2 explained 13.16% and 12.23% of the genetic variation (Figure 3.3C). This approach also revealed that the association mapping panel is composed of two subpopulations, confirming the results of the Bayesian cluster analysis of population structure.

The number of subpopulations of the mapping panel was also determined using phylogenetic tree analysis, which is based on membership to different expected groups obtained from STRUCTURE analysis. Our phylogenetic tree also revealed the presence of two distinct groups (Figure 3.3D) as shown by the Bayesian cluster analysis method and PCA.

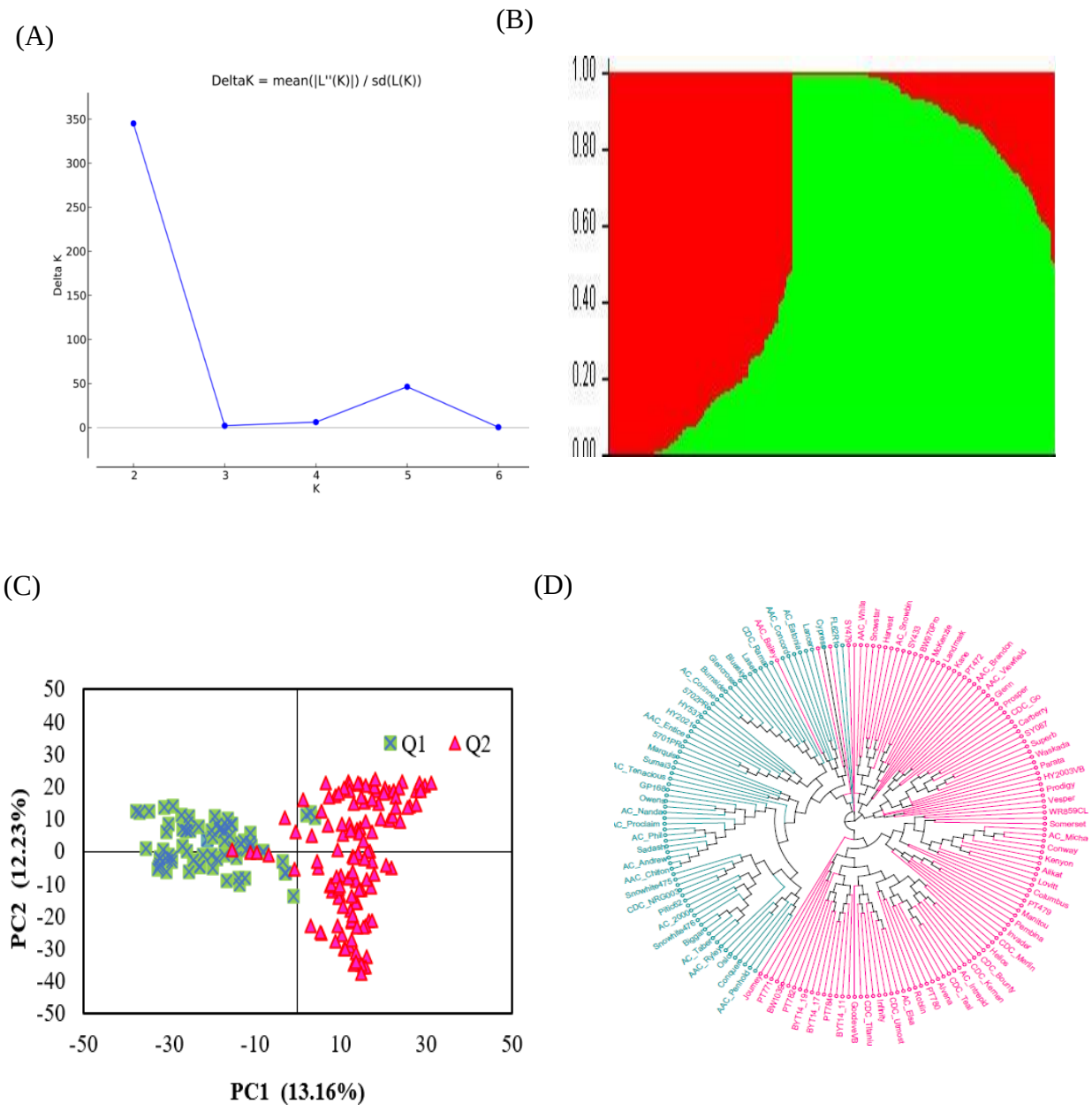


Figure 3.3 Determination of number of subpopulations. A) Delta K method, the x-axis and y-axis represent subgroups K and delta K respectively, the ΔK peak value at K = 2 showed the presence of two subpopulations; B) Structure bar plot showing two subpopulations represented in two colours Q1 - green color and Q2 - red color; C) Scatter plot, x axis and y axis represent the principal component 1 (PC1) versus principal component 2 (PC2), respectively, where each dot represents an individual genotype in the association mapping panel in two subpopulations and the percentage of genetic variance shown in parenthesis is explained by the major components. D) Neighbour-based phylogenetic tree demonstrating the grouping of the 194 wheat lines into two groups, each indicated by a distinct colour.

3.3.5 Estimation of linkage disequilibrium

A total of 18,672 SNPs was used to calculate the extent of the linkage disequilibrium (LD) in the association mapping panel. The squared allele-frequency correlations (r^2) were used to calculate pair-wise LD. The LD decay was calculated for the whole genome as well as for each sub-genome A, B and D by plotting the r^2 against the corresponding genetic distance between pairwise SNPs (cM). The whole genome LD decayed at a distance of 4.51 cM with a critical threshold of r^2 = 0.29. Therefore, a distance of ± 4.51 cM was used as a confidence interval for the map location of QTLs. The extent of LD decay for the sub-genomes A, B and D were 3.21 cM, 2.18 cM and 10.78 cM, respectively (Figure 3.4 A, B and C). The D genome had the highest level of LD, while the B genome had the lowest.

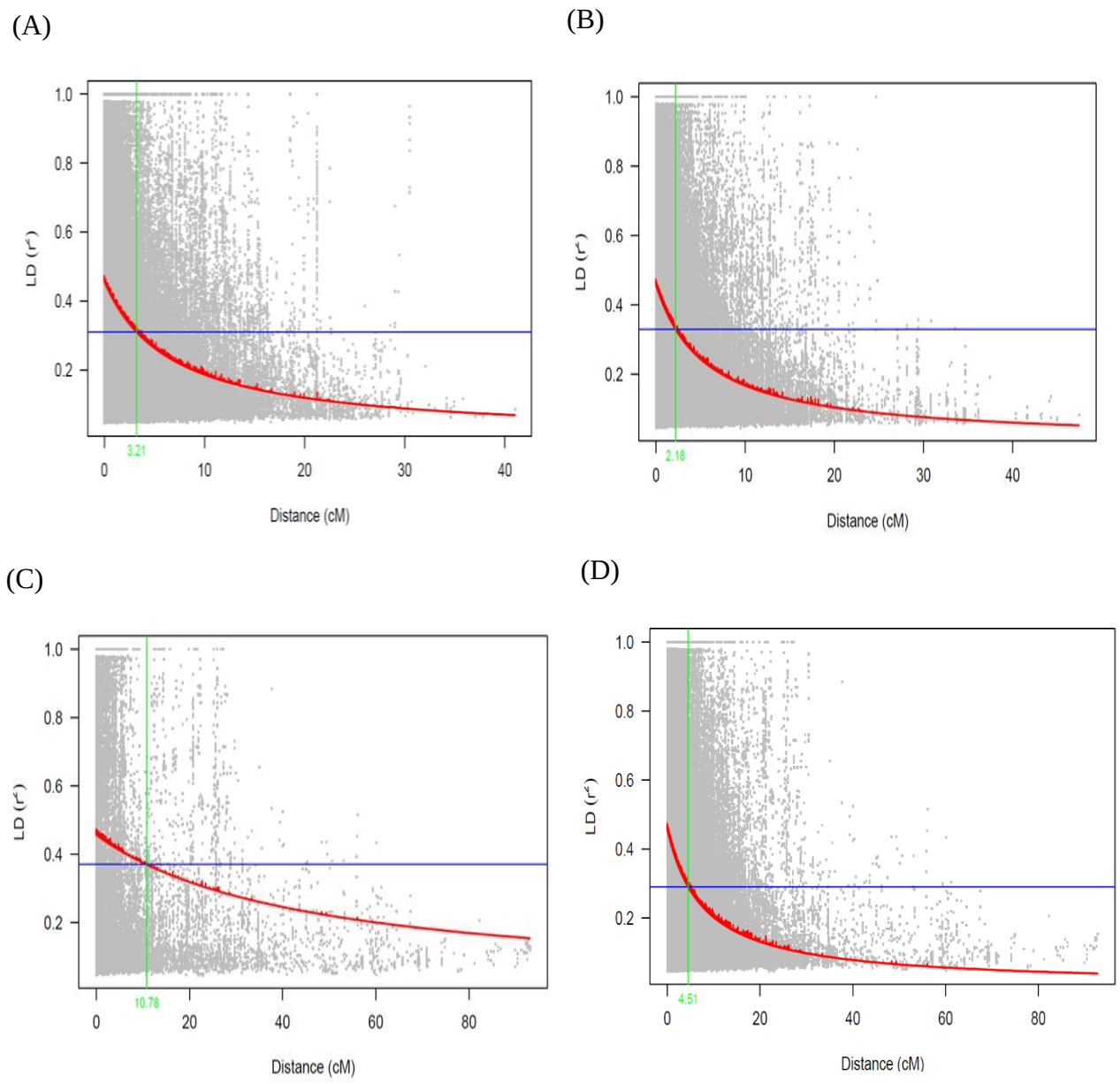
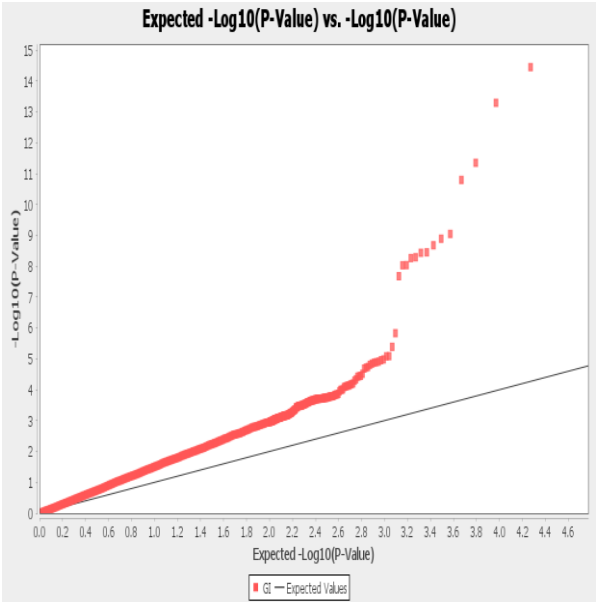


Figure 3.4 Genome-wide linkage disequilibrium (r^2) decays across genetic distance (cM). Based on non-linear regression, the LD degradation is represented by the red curve. The blue line indicates the intersection of the r^2 with the regression line and the green line shows the LD decay as a function of genetic distance in cM. A) LD decay for genome A; B) LD decay for B genome; C) LD decay for D genome; D) LD decay for the whole genome.

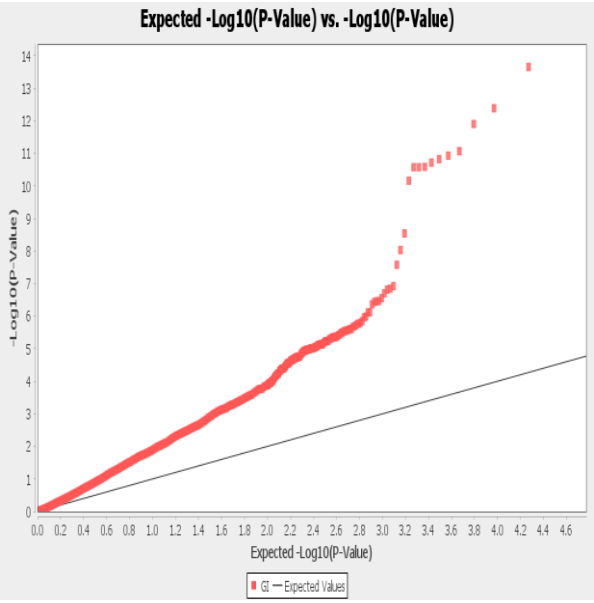
3.3.6 Marker trait association

Genome-wide association analysis was performed with the BLUEs value of the phenotypic data for the individual environments as well as the dataset pooled from all the trail environments. To examine an association between specific SNPs and the dormancy trait, different models such as two GLM models with principal component analysis (GLM-PCA) and Q matrix (GLM-Q) as covariates were run, and two MLM models with K matrix as an additional covariate (MLM-PCA+K), (MLM-Q+K) were run in TASSEL. The QQ plots of all models, GLM (PCA) (Figure 3.5 A), GLM (Q) (Figure 3.5 B), MLM (PCA+K) (Figure 3.5 C) and MLM (Q+K) (Figure 3.5 D) were generated. For each model, the observed and expected -log10 *P* values were compared. In the MLM (PCA+K) and MLM (Q+K) models, a relatively uniform distribution of these *P* values was found which means that both false positives and false negatives are controlled (Figure 3.5 C and D) and both showed significant associations. In the MLM, the model using the (Q+K) and (PCA+K) resulted in lower *P*-value inflation. However, compared to the MLM (PCA+K) model, the MLM (Q+K) model produced a somewhat better uniform distribution and for further analysis we used the MLM model utilizing the kinship and Q matrix as a covariate.

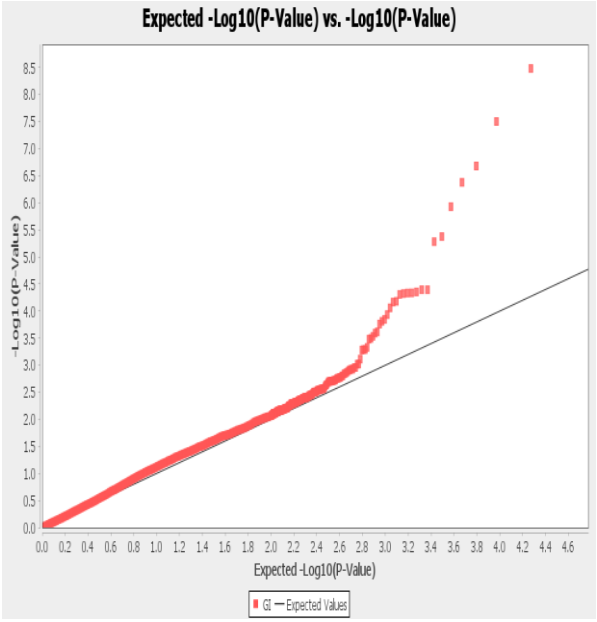
(A)



(B)



(C)



(D)

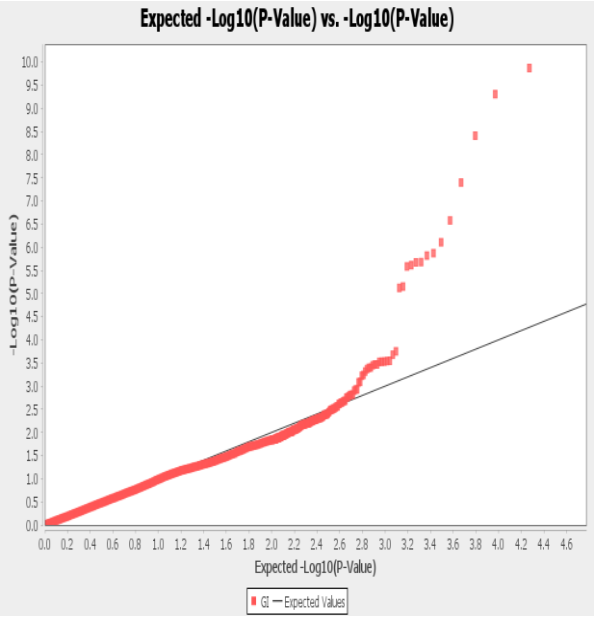


Figure 3.5 Quantile-Quantile plots of the observed -log10 P values against the expected -log10 (*P* values) for the dataset pooled from all the environments using A) GLM model (PCA), B) GLM model (Q), C) MLM model (PCA+K), and D) MLM model (Q+K).

3.3.7 Identification of significant markers

The true marker-trait associations for this study were calculated using Bonferroni correction and the lowest 0.1 percentile method to identify significant markers using a threshold level of > -log 10 (*P*-value). Using the Bonferroni threshold value, all those SNPs with -log 10 (*P*-value) of > 4.27 were identified as significant. As such, 14 significant markers were detected using the pooled dataset and all the 14 SNPs are located on 4A chromosome (Appendix 1, 2). These significant markers explained phenotypic variations ranging from 8.4 to 25.77%. On the other hand, the 0.1 percentile method, which considered SNPs with a threshold *P*-value of 2.938×10^{-4} and a corresponding -log 10 (*P*-value) > 3.52, identified 19 significant markers (Figure 3.6, Table 3.4). These markers are located on chromosomes 2B, 3B, 4A and 5B. The phenotypic variation of these markers ranges from 7.6% to 25.77%. The maximum phenotypic variation was observed with markers located on chromosome 4A. Out of the 19 significant markers identified using the pooled dataset, 11 significant markers were also identified in MOR16 and MOR17. Moreover, wsnp_Ex_c13031_20625900, wsnp_Ex_rep_c66324_64493429 and Jagger_c4331_105 markers are identified as significant markers in all the six trial environments (Appendix 1).

Since the Bonferroni correction method is more stringent, we used the results obtained using 0.1 percentile method for further analysis. The genetic positions of the significant markers were retrieved from the 90K wheat array consensus genetic maps (Wang *et al*., 2014). The physical positions of the significant SNPs were attained by blasting the sequences of the SNPs against wheat Chinese Spring Ref Seq v2.1 assembly (IWGSC, 2018).

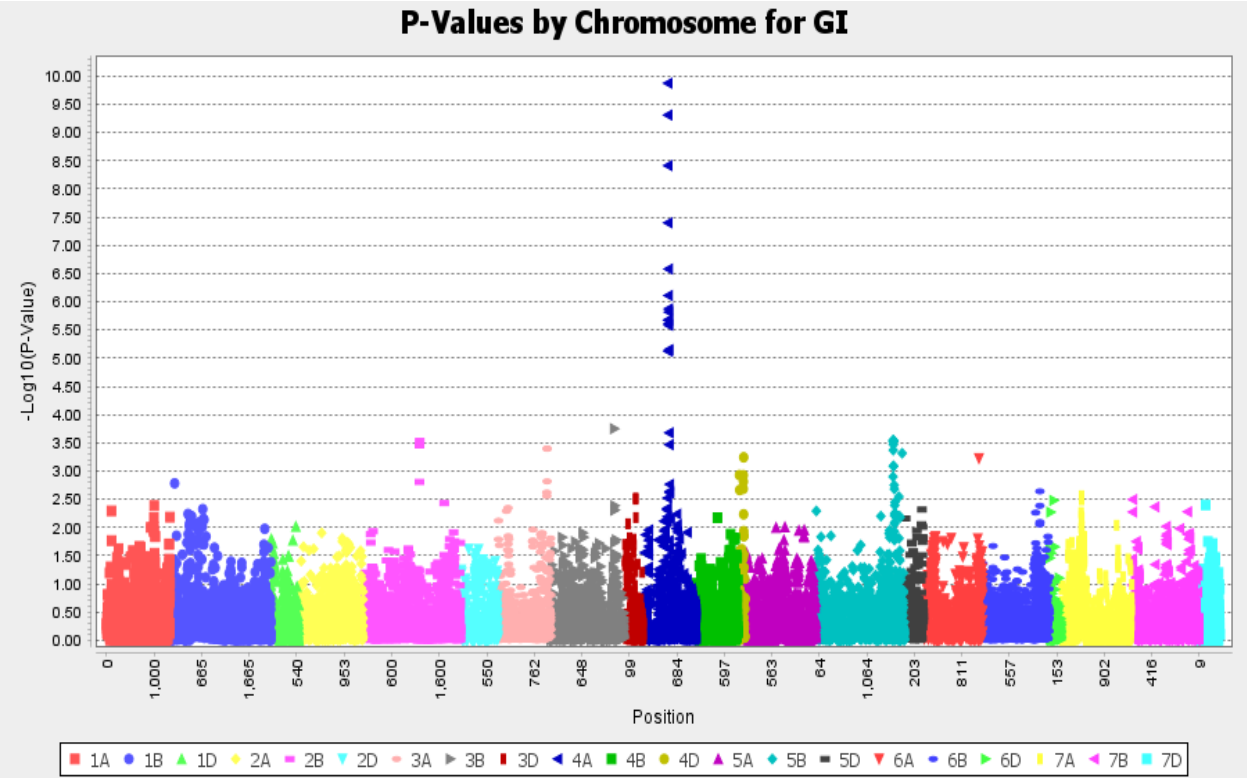


Figure 3.6 Manhattan plots derived for the pooled dataset from all trial environments. The x-axis represents the number of chromosomes in wheat and the y-axis represents the -log 10 (*P*-value). Each coloured dot represents an SNP. The Bonferroni corrected threshold for determining significant relationships between the marker and the trait is shown by the red horizontal line and the 0.1 percentile threshold is shown by the green horizontal line. All the markers beyond the threshold lines are significant.

3.3.8 Detection of QTL for PHS resistance

The GWAS revealed the potential location of QTLs for PHS resistance. Four QTLs associated with the 19 significant markers representing four chromosomes, 2B, 3B, 5B and 4A, were detected. All the significant markers observed within a ± 4.51 cM were considered as part of a single QTL. On chromosomes 2B and 3B, the Tdurum_contig92604_368 (103.33cM) and BS00070210_51(112.63cM) markers represented a single QTL designated as QTL.2B-1 and QTL.3B-1, respectively. The RAC875_rep_c111379_703 (144.25cM) and Kukri_c32825_95

(144.12cM) markers on chromosomes 5B have the same genetic positions, therefore, they represented a single QTL defined as QTL.5B-1. Calculation of the LD decay for subgenomes revealed that the LD decay for subgenome A is 3.21; therefore, the SNPs located within ± 3.21 cM between each other were considered to represent a single QTL. Fifteen significant markers, namely Kukri_c27874_515, wsnp_Ex_c13031_20625900, wsnp_Ex_rep_c66324_64493429, BS00072025_51, Kukri_c12563_52, Jagger_c4331_105, Excalibur_c40202_114, IAAV3132, BS00030838_51, Excalibur_c8658_335, Tdurum_contig14947_611, RAC875_c18090_752, Ra_c22219_350, Ra_c23848_305 and BobWhite_c12128_187 were found to be located within ± 3.21 cM (65.95 cM to 71.25 cM) between each other on Chromosome 4A (Table 3.4). As such, these markers represented a single QTL designated as QTL.4A-1.

Table 3.4 Significant markers identified using dataset pooled from all environments[a]

QTL[b]	Marker name	Chr[c]	Position (bp)[d]	Position (cM)[e]	-log10 (P)[f]	R^2 (%)[g]
QTL.2B-1	Tdurum_contig92604_368	2B	555,098,800	103.33	3.53	7.58
QTL.3B-1	BS00070210_51	3B	607,696,954	112.63	3.75	8.22
QTL.5B-1	RAC875_rep_c111379_703	5B	670,338,306	144.25	3.54	7.56
QTL.5B-1	Kukri_c32825_95	5B	668,473,776	144.12	3.55	7.61
QTL.4A-1	Kukri_c27874_515	4A	604,041,767	65.95	9.31	24.02
QTL.4A-1	wsnp_Ex_c13031_20625900	4A	603,103,256	65.95	9.87	25.78
QTL.4A-1	wsnp_Ex_rep_c66324_64493429	4A	603,104,997	65.95	8.41	22.33
QTL.4A-1	BS00072025_51	4A	607,696,954	66.27	7.4	18.6
QTL.4A-1	Kukri_c12563_52	4A	604,623,882	66.27	5.67	13.3
QTL.4A-1	Jagger_c4331_105	4A	605,291,230	66.68	6.11	14.62
QTL.4A-1	Excalibur_c40202_114	4A	563,490,514	66.95	6.58	17.55
QTL.4A-1	IAAV3132	4A	607,470,479	66.95	5.68	13.42
QTL.4A-1	BS00030838_51	4A	604,555,417	66.95	5.12	12.14
QTL.4A-1	Excalibur_c8658_335	4A	607,692,372	66.95	5.58	13.87
QTL.4A-1	Tdurum_contig14947_611	4A	608,830,565	66.95	5.82	13.89
QTL.4A-1	RAC875_c18090_752	4A	607,696,554	66.95	5.61	13.34
QTL.4A-1	Ra_c22219_350	4A	608,682,947	66.95	5.87	13.84
QTL.4A-1	Ra_c23848_305	4A	609,573,920	69.18	5.15	11.89
QTL.4A-1	BobWhite_c12128_187	4A	609,262,213	71.25	3.68	8.35

[a]Environments are as shown in Table 1
[b]Putative QTL containing one or more SNPs on the basis of LD decay distance
[c]Chromosome
[d]Marker position in bp obtained from the wheat Chinese Spring Ref Seq v2.1 assembly (IWGSC, 2018)
[e]Marker positions in cM derived from SNPs consensus genetic maps (Wang *et al*., 2014)
[f]Markers selected on the basis of 0.1 percentile method with P value threshold of 2.938×10^{-4} [-log10 (*P*) 3.52]
[g]R^2 (%) represents the percentage of phenotypic variation accounted by each significant SNP

3.3.9 Identification of candidate genes

For the identification of candidate genes associated with PHS resistance, the sequences of the significant markers were BLAST searched against IWGSC RefSeq v2.0 using the Ensembl genome database (https://plants.ensembl.org/Triticum_aestivum/Info/Index). Our analysis identified 11 candidate genes associated with significant markers on chromosomes 2B, 2D, 4A, 5A, 5B and 5D, and all 11 genes were within ± 250 kb of the mapped SNPs, based on the physical map of the reference genome of Wheat Chinese Spring IWGSC RefSeq v2.1 (2021) (Table 3.5). Annotation of the identified candidate genes was performed using the names of the respective protein family/subfamily of other species obtained from the protein annotation through evolutionary relationship (PANTHER) database (http://www.pantherdb.org/). Out of the 11 potential candidate genes, the *LRR* gene is associated with the two most important markers, wsnp Ex c13031 20625900 and wsnp Ex rep c66324 64493429, both of which are found on QTL.4A-1. The *SNF4* gene is associated with four significant markers, namely RAC875_c18090_752, BS00030838_51, Excalibur_c8658_335 and Tdurum_contig14947_611, and it belongs to a single QTL (QTL.4A-1). The *KU80* gene is associated with markers Excalibur_c40202_114, IAAV3132 and belongs to QTL.4A-1.

Table 3.5 Putative candidate genes for PHS resistance associated with significant markers.

Short name	Gene ID	Chr	Start (bp)	End (bp)	Associated SNP	Description in PANTHER Family/Subfamily database
DPMS1	TraesCS5B02G500500	5B	668473776	668478123	Kukri_c32825_95	Dolichol phosphate mannose synthase/PTHR43398:SF1
	TraesCS5D02G500800.1	5D	529215471	529219111	Kukri_c32825_95	
	TraesCS5A02G486600.1	5A	657378660	657381315	Kukri_c32825_95	
BPM2	TraesCS2B02G383700	2B	555106333	555107433	Tdurum_contig92604_368	BTB-POZ AND MATH DOMAIN 2/(PTHR26379:SF458)
	TraesCS2D02G363000.1	2D	468945110	468950456	Tdurum_contig92604_368	
GGL13	TraesCS4A02G316500	4A	605302928	605307241	Jagger_c4331_105	GDSL-motif esterase/acyltransferase/lipase/ PTHR45642:SF73)
LRR	TraesCS4A02G313900	4A	603117627	603121477	wsnp_Ex_c13031_20625900/wsnp_Ex_rep_c66324_64493429	Leucine-rich repeat serine/threonine protein kinase/PTHR27008:SF413
RRP46	TraesCS4A02G322500	4A	609274795	609278060	BobWhite_c12128_187	Exosome complex exonuclease RRP46/PTHR11953:SF1
SNF4	TraesCS4A02G320000	4A	607705211	607712888	RAC875_c18090_752/BS00030838_51/Excalibur_c8658_335/Tdurum_contig14947_611	Component of the regulatory subunit of SNF1-related protein kinase/PTHR13780:SF35
MED15a	TraesCS4A02G315600	4A	604569504	604574635	BS00072025_51	Mediator of RNA polymerase II transcription subunit 15a/PTHR33137:SF47
KU80	TraesCS4A02G319800	4A	607480914	607487484	Excalibur_c40202_114/IAAV3132	ATP-dependent DNA helicase 2 subunit KU80/PTHR12604:SF4
MED13	TraesCS4A02G315800	4A	604623980	604639969	Kukri_c12563_52	Mediator of RNA polymerase II transcription subunit 13/PTHR48249:SF3
FRUCT5	TraesCS4A02G321500	4A	608697881	608703540	Ra_c22219_350	6-Fructon exohydrolase/PTHR31953:SF45
FBD	TraesCS4A02G322900	4A	609587781	609591045	Ra_c23848_305	F-box/RNI-like superfamily protein/PTHR31639:SF65

3.4 Discussion

Pre-harvest sprouting is a major concern to wheat producers and causes significant yield losses. Thus, identifying the resistance sources among the existing germplasm and using them in wheat breeding represents a successful strategy against PHS. This study evaluated genome-wide association of the phenotypic trait i.e., seed dormancy/PHS and the genotypic data of 194 diverse spring wheat genotypes over several environments. The wide range of diversity in dormancy/PHS resistance among the genotypes studied over the individual trial environments as well as in the dataset pooled from all environments leads to the identification of QTLs/genomic regions associated with seed dormancy/PHS resistance. Our results revealed a significant and positive phenotypic correlation among the different trial environments, and the broad sense heritability, which is estimated to be 0.81, indicating the significance of genetic factors in regulating PHS trait as reported previously (Kulwal *et al*., 2005; Kumar *et al*., 2006; Imtiaz *et al*., 2008; Mohan *et al*., 2009; Su and Li 2018).

In this study, the three sub-genomes of common wheat, namely A, B, and D, have been evaluated for their LD decay distances in cM. The D sub-genome was found to exhibit a higher LD decay distance of 10.78 cM as compared to that of the A and B sub-genomes, which showed LD decay distances of 3.21 and 2.18 cM, respectively. Similar LD decay distances in wheat have also been reported in previous studies (Zhang *et al*., 2010; Hao *et al*., 2011; Ayana, 2017). Other studies that involved the use of diverse wheat lines have also revealed substantially higher LD. For instance, a study with European common wheat lines has shown LD decay distances of 23 cM (Nielsen *et al*., 2014). In this study, LD decay distance has been used to define a QTL and the LD dropped at an approximate genetic distance of 4.51 cM for the whole genome, therefore, ± 4.51 cM was used to establish confidence intervals for QTL regions.

Our association analysis identified a total of 19 significant marker-trait associations (Table 3.4; Figure 3.5), and the 19 significant SNPs represented four distinct QTLs distributed over chromosomes 2B, 3B, 4A and 5B, namely QTL.2B-1, QTL.3B-1, QTL.4A-1 and QTL.5B-1, respectively (Table 3.4). Consistently, previous genetic studies in wheat identified significant QTLs that are believed to be important in regulating dormancy including Qphs.caas-3AL, Qphs.caas3DL and Qphs.caas-4AL on chromosome 3A, 3D and 4A respectively (Li *et al*., 2021). In order to compare the locations of the QTLs identified by this study with those reported by previous studies, we identified the approximate physical positions of the QTLs using the IWGSC RefSeq v2.1 (IWGSC, 2018). Since different types of markers were used in the previous mapping studies, we were not able to find the precise position of the QTLs. However, QTL.4A, which is detected on chromosome 4A at 563.4 Mbp, appears to be mapped to the same genomic region where Qphs.caas-4AL was detected previously (Liu *et al*., 2004), indicating its potential role in the control of seed dormancy. Among the four QTLs identified by this study, QTL.4A-1 was detected in all trial environments.

Several studies have reported numerous QTLs associated with PHS resistance in wheat using different mapping populations (Kulwal *et al*., 2010; Tyagi and Gupta, 2012; Vetch *et al*., 2019b; Wang *et al*., 2019). The majority of wheat chromosomes, including chromosomes 3A, 3B, 4A, and 4B in particular, have been reported to carry QTLs underlying PHS resistance or seed dormancy (Li *et al*., 2021). In particular, chromosome 4A is crucial for PHS resistance since it contains a major PHS locus designated as *Phs* (Nelson *et al*., 1995; Flintham and Gale, 1996; Bailey *et al*., 1999; Flintham, 2000; Watanabe and Ikebata, 2000; Flintham *et al*., 2002; Gale *et al*., 2002; Himi *et al*., 2002; Nakamura and Toyoma, 2001). As discussed above, a previous study identified a QTL on chromosome 4A at position 489.1–532.2 Mbp (Li *et al*., 2021). This study

also detected QTLs underlying PHS on chromosome 4A, namely QTL.4A-1 (563.4 Mbp to 609.5 Mbp), which explained 8.4% to 25.8% phenotypic variation, indicating the presence of single QTL that control seed dormancy/PHS on chromosome 4. This QTL.4A-1 consists of 15 significant SNPs that were within the LD decay window of ± 3.21 cM (65.95 cM to 71.25 cM).

A total 11 genes are found to be associated with the significant SNP markers identified in our study, and these genes are located within ± 250 kb of the significant markers. Therefore, these genes might play important roles in controlling seed dormancy/PHS resistance and are considered candidate genes underlying the QTLs identified. The significant SNPs representing QTL.4A-1, namely wsnp_Ex_c13031_20625900 and wsnp_Ex_rep_c66324_64493429, accounted for 25.8% and 22.3% of the phenotypic variance, respectively. Consistently, a previous study reported that the markers wsnp Ex c13031 20625900 and wsnp Ex rep c66324 64493429 are associated with PHS resistance in wheat (Zhang *et al*., 2008). Given that the two markers, which are located at 603.1 Mbp, are found to be linked with a leucine-rich repeat (*LRR*) gene, it is likely that *LRR* underlies the QTL and therefore plays important role in regulating seed dormancy/PHS resistance. In support of this result, a previous study in Arabidopsis has reported that overexpression of the *LRR* gene causes delayed seed germination (Wu *et al*., 2017).

The Tdurum_contig92604_368 SNP, which represents QTL.2B-1 and explained 7.6% of the phenotypic variation, is located at 103.33 cM (555.1Mbp), and this SNP is linked with the *btb/poz-math* (*BPM2*) gene, reflecting the role of *BPM2* in the regulation of seed dormancy. The *BPMs* have important roles in seed development and abiotic stress response in plants (Weber and Hellmann 2009, Lechner *et al*., 2011, Chen *et al*., 2013, 2015; Morimoto *et al*., 2017). For example, a previous study in Arabidopsis has shown that *BPM2* and *dehydration-responsive element-binding* (*DREB2A*) form a complex after heat exposure (Morimoto *et al*., 2017).

Upregulation of *BPM2* in response to heat stress induces the degradation of *DREB2A* during heat stress, and this leads to a reduction in the harmful effects of excess *DREB2A* on plant growth (Skiljaica *et al*., 2020). These *BPMs* can also control plant responses to the phytohormone ABA (Lechner *et al*., 2011). Despite these reports, the direct role of *BPMs* in regulating seed dormancy is still unclear.

The Kukri_c32825_95 SNP representing QTL.5B-1 and located at 144.12 cM (668.4 Mbp) is linked with a *dolichol phosphate mannase synthase* (*DPMS1*) gene, reflecting the role of *DPMS1* in seed dormancy/PHS resistance. Previous reports in Arabidopsis have shown that *DPMS1* is localized in the endoplasmic reticulum and controls isoprenyl-linked glycan biosynthesis, which is crucial for regulating plant growth and development and its response to stress factors (Morimoto et al., 2017). Loss-of-function mutations of *DPMS1* causes severe chlorosis, reduced root development, and a wrinkled seed coat (Jadid *et al*., 2011). However, the role of this gene in seed dormancy has not yet been reported.

Guard-cell-enriched GDSL lipase (*GGL*), which encodes GDSL-type esterases/lipases, are widely distributed in plants and have been reported in different species such as Arabidopsis, rice and lycophyte (*Selaginella moellendorffii*). Our data showed that the *GGL13* gene contains a single SNP Jagger_c4331_105, which explained 14.62% of the phenotypic variation, on chromosome 4A at 66.68cM (605.3 Mbp). Given our result and its reported role in germination in *Arabidopsis* (Updegraff, Zhao and Preuss, 2009), it is likely that *GGL13* contributes to the regulation of seed dormancy/PHS resistance in wheat. The SNP BobWhite_c12128_187 of QTL.4A-1 is closely linked with *rRNA-processing protein 46* (*RRP46*) gene, which encodes a RRP46 protein that has been shown to posses phosphorolytic RNase activity in addition to hydrolytic DNase activity in *Oryza sativa* (Yang *et al*., 2010). However, its role in seed dormancy has not been reported so far.

Four SNPs, namely RAC875_c18090_752, BS00030838_51, Excalibur_c8658_335 and Tdurum_contig14947_611 at 66.95 cM (607.7Mbp) of QTL.4A-1 are linked to *sucrose non-fermenting* (*SNF4*) gene. AMP-activated protein kinase (AMPK)/sucrose non-fermenting 1 (SNF1) formed SNF1-related protein kinase 1 (SnRK1) heterotrimeric complexes, which typically consist of a catalytic, a scaffolding and an activating subunit and play crucial regulatory roles in metabolism, stress signalling and plant development, are encoded by the *sucrose non-fermenting* (*SNF*) genes (Broeckx *et al.*, 2016; Li and Sheen, 2016). The *SNF* genes have been shown to play roles in the germination of tomato seeds (Bradford *et al*., 2003). Therefore, this gene might have a role in influencing seed germination in wheat.

The SNP BS00072025_51 located at 66.27 cM (604.5Mbp) on chromosome 4A was identified to be associated with Mediator Complex Subunit 15 (*MED15a*), which encodes a protein that acts as a mediator of RNA polymerase II transcription subunit 15a. *Med15* has been suggested to interact with NAC, AF1/2 and CUC transcription factors that are involved in the process of controlling seed development in rice (Thakur *et al*., 2013). Therefore, this gene might have a role in seed dormancy. Two SNPs designated as Excalibur_c40202_114 and IAAV3132, located at 66.95 cM (607.4 Mbp) of QTL.4A-1 are closely linked to *Ku80*. It has been demonstrated that the transcription levels of *Ku80* increase in response to the formation of DNA damage-specific breaks (DSBs) caused by DNA-damaging agents, indicating its importance for DSB repair via the NHEJ (non-homologous end joining) pathway (Tamura *et al*., 2002). Further studies are required to determine whether *Ku80* plays a role in controlling seed dormancy.

The SNP Ra_c22219_350, which is located at 66.95 cM (608.6 Mbp) of QTL.4A-1, is closely associated with the *fructofuranosidase 5* (*FRUCT5)* gene. *FRUCT5* encodes a protein with fructan exohydrolase (FEH) activity that is involved in the hydrolysis of fructans in plants.

Previous studies have revealed the involvement of this gene in cold and drought tolerance through regulating the remobilization of water-soluble carbohydrates from the stem to the growing grain (Lothier *et al*., 2007; Hincha *et al*., 2003). However, no information about its role in seed dormancy has not been reported so far.

The SNP Ra_c23848_305, which is mapped at 69.18cM (609.5 Mbp) of QTL.4A-1 and explained 11.89% of the phenotypic variation, is closely associated with F-box domain (*FBD*) gene. The *F-box* genes are anticipated to play an important role in ubiquitin–protein transferase activity (Hong *et al*., 2020). Despite their significant roles in various aspects of plant development and its response to abiotic/biotic stress, a functional analysis of wheat *F-box* genes has only ever been performed at the single gene level (Hong *et al*., 2020). As a result, further studies are required to clarify if the *F-box* genes contribute to regulate seed dormancy and germination in wheat.

4.0 GENERAL DISCUSSION AND CONCLUSION

Wheat is the most widely grown cereal crop in the world. Its economic significance is demonstrated by the fact that it accounts for 20% of all human calorie consumption and is the main diet for more than 40% of the world's population (Giraldo *et al*., 2019). Pre-harvest sprouting (PHS), which refers to the germination of grains while on the spike due to the occurrence of wet and humid conditions before harvest, is one of the main factors that restrict wheat productivity. This is because many of the commercial cultivars don't have appropriate levels of dormancy, a trait that prevents seeds from germinating even in conditions that appear to be ideal (Gao and Ayele, 2014). The underlying reason for this is the selection of wheat against dormancy during the domestication of wheat so as to have cultivars with a consistent and quick rate of seed germination and seedling establishment (Tuan *et al*., 2018). It is therefore crucial to develop new cultivars with better PHS resistance, and this has been one of goals of wheat breeding programs around the world. Since it is a complex quantitative trait, PHS is influenced by both genetic and environmental conditions (Gao and Ayele, 2014). Previous genetic studies that used mapping populations have shown that all the 21 chromosomes of wheat consist of PHS resistance/seed dormancy QTLs (Kulwal *et al*., 2010, Vetch *et al*., 2019b, Wang *et al*., 2019). More recently, there is a significant increase in the application of GWAS to investigate the genetic basis of quantitative traits in wheat, including grain characteristics and quality, abiotic and biotic stress tolerance. This is because it uses genetically diverse populations to identify marker-trait associations, which results in high mapping resolution and greater allele frequency.

This study identified QTLs and candidate genes associated with dormancy/PHS using genotypic data as well as seed germination index data obtained from 194 diverse spring wheat genotypes under six environments. A total of 18,672 SNPs acquired via the use of 90 K Illumina

I select genotyping array and were used to find QTL and genes responsible for PHS. The LD decay distances (r^2 = 0.29) were approximately 3.21, 2.18, 10.78, and 4.51 cM for the A, B and D subgenomes and whole genome, respectively, and this shows the presence of considerable genetic diversity in the association mapping panel. Understanding the genetic relatedness of the wheat accessions used in genetic studies requires knowledge of population structure. Our population structure analysis identified two subpopulations. Dormancy phenotype of the genotypes in mapping panel was assessed using mean GI values, and our findings showed the mapping population used in this study consists of genotypes with different level of dormancy, which is categorized as high, moderate and low levels of dormancy. As a result, the mapping panel is considered as useful resource to study the genetic basis of PHS. The best fitting model MLM (Q+K) was used to analyze the relationship between phenotypic and genotypic data because it minimizes the spurious marker-trait associations caused by population structure and familial relatedness. Our analysis detected four putative QTL regions on chromosomes 2B, 3B, 5B and 4A based on 19 significant marker-trait associations (MTAs) identified using MLM (Q+K) model and 0.1 percentile significant threshold of -log10 *(P)* = 3.52. The MTAs identified have the potential to be used as basis for further study or be applied in wheat breeding programs.

This study detected one QTLs on chromosome 4A which is designated as QTL.4A-1, which contains 15 significant SNPs named Kukri_c27874_515, wsnp_Ex_c13031_20625900, wsnp_Ex_rep_c66324_64493429, BS00072025_51, Kukri_c12563_52, Jagger_c4331_105, Excalibur_c40202_114, IAAV3132, BS00030838_51, Excalibur_c8658_335, Tdurum_contig14947_611, RAC875_c18090_752, Ra_c22219_350, Ra_c23848_305 and BobWhite_c12128_187 in the region of 563.4 Mbp to 609.5 Mbp. These results suggest that the 4A locus may play an important role in controlling seed dormancy. Although previous studies have

found QTLs on 4A, we are unable to directly compare the precise location of our QTLs with that of those identified by the others because the mapping approach including the statistical methodologies and markers used by the studies are different. Three significant markers, namely wsnp_Ex_c13031_20625900, wsnp_Ex_rep_c66324_64493429 and Jagger_c4331_105, were identified consistently in each of the six trial environments and dataset pooled from all environments. Therefore, the genomic region (65.95 cM to 66.95 cM) consisting these three markers on Chromosomes 4A can be considered a key genomic locus that may contain combinations of genes for improving PHS resistance. Development of PHS resistant cultivars of wheat is important to reduce the negative effects of PHS on wheat yield and quality. Since seed dormancy/PHS is a quantitative trait, more genetic studies using either bi-parental or association mapping panel are necessary to identify additional genomic regions associated with these important traits.

This study also identified 11 genes linked to the significant markers as candidate genes to control PHS resistance. Six of these genes identified, which include *LRR*, *GGL13, MED13a*, *MED15a*, *SNF4* and *FBD*, have previously been linked to regulate seed germination and dormancy. The involvement of the remaining five genes, namely *BPM2, DPMS1*, *RRP46, Ku80* and *FRUCT*, in germination and dormancy is not well known. Therefore, further studies are needed to investigate the functions of these genes with respect to seed dormancy or germination.

In summary, identification of genomic regions and candidate genes related to seed dormancy and PHS was made possible through the use of a collection of diverse genotype, a high-density DNA marker platform, and a publicly available high-quality reference genome assembly. The genome-wide association analysis enabled to identify functional SNPs that could locate

genomic regions linked to dormancy and be employed in a marker-assisted selection of PHS-resistant genotypes.

LITERATURE CITED

Agarwal M, Shrivastava N, Padh H (2008) Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Rep* **27:** 617-631

Agrama HA, Eizenga GC, Yan W (2007) Association mapping of yield and its components in rice cultivars. *Mol Breed* **19:** 341-356

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. *Gen Res* **12:** 1805-1814

Akhunov E, Nicolet C, Dvorak J (2009) Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina Golden Gate assay. *Theor Appl Genet* **119:** 507-517

Allen AM, Barker GL, Wilkinson P, Burridge A, Winfield M, Coghill J, Uauy C, Griffiths S, Jack P, Berry S, Werner P (2013) Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum L.*). *Plant Biotech J* **11:** 279-295

Alqudah AM, Sallam A, Baenziger PS, Börner A (2020) GWAS: Fast-forwarding gene identification and characterization in temperate Cereals: lessons from Barley–A review. *J Adv Res* **22:** 119-135

Alvarado G, Rodriguez FM, Pacheco A, Burgueño J, Crossa J, Vargas M, Pérez-Rodríguez P, Lopez-Cruz MA (2020) META-R: A software to analyze data from multi-environment plant breeding trials. *The Crop Journal* **8:** 745-756

Anderson JA, Sorrells ME, Tanksley SD (1993) RFLP analysis of genomic regions associated with resistance to preharvest sprouting in wheat. *Crop Sci* **33:** 453–459

Ayana GT, Ali S, Sidhu JS, Gonzalez Hernandez JL, Turnipseed B, Sehgal SK (2018) Genome-wide association study for spot blotch resistance in hard winter wheat. *Front Plant Sci* **9:** 926

Ayoub M, Symons S, Edney M, Mather D (2002) QTLs affecting kernel size and shape in a two-rowed by six-rowed barley cross. *Theor Appl Genet* **105:** 237-247

Bailey PC, McKibbin RS, Lenton JR, Holdsworth MJ, Flintham JE, Gale MD (1999) Genetic map locations for orthologous Vp1 genes in wheat and rice. *Theor Appl Genet* **98:** 281-284

Bajgain P, Rouse MN, Tsilo TJ, Macharia GK, Bhavani S, Jin Y, Anderson JA (2016) Nested association mapping of stem rust resistance in wheat using genotyping by sequencing. *PloS one **11*:** e0155760

Bandyopadhyay BB (2011) Field Evaluation of Cold Sensitive Response in Spring Wheat Genotypes at High Altitude using Multiple Parameters. *Int J Appl Agric Sci* **6:** 253–266

Barabaschi D, Tondelli A, Desiderio F, Volante A, Vaccino P, Valè G, Cattivelli L (2016) Next generation breeding. *Plant Sci* **242:** 3-13

Barrero JM, Cavanagh C, Verbyla KL, Tibbits JF, Verbyla AP, Huang BE, 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:** 1-18

Bason ML, Ronalds JA, Wrigley CW (1991) Screening for pre germination of barley at harvest. In: Westcott T, Williams Y (eds) Proceedings of the 40th Australian cereal chemistry conference. *Royal Australian Chemistry Institute, Melbourne*, pp 120–125

Batley J and Edwards D (2007) SNP applications in plants. In *Association mapping in plants* (pp. 95-102). Springer, New York, NY

Belderok B and Habekotte A (1980) Induction of secondary dormancy in wheat crops by low temperatures and high relative humidities in the field. *Cereal Res Commun* **1:** 167-173

Benech-Arnold RL, Gualano N, Leymarie J, Côme D, Corbineau F (2006) Hypoxia interferes with ABA metabolism and increases ABA sensitivity in embryos of dormant barley grains. *J Exp Bot* **57:** 1423–1430

Benjamini Y and Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc: series B (Methodological)* **57:** 289-300

Bethke PC and Jones RL (2001) Cell death of barley aleurone protoplasts is mediated by reactive oxygen species. *Plant J* **25:** 19-29

Bewley JD (1997) Seed germination and dormancy. *The Plant Cell* ***9*****:** 1055

Bewley JD and Black M (1994) Seeds: physiology of development and germination. Plenum Press, New York

Biddulph TB, Mares DJ, Plummer JA, Setter TL (2005) Drought and high temperature increases preharvest sprouting tolerance in a genotype without grain dormancy. *Euphytica* **143:** 277-283

Biddulph TB, Plummer JA, Setter TL, Mares DJ (2007) Influence of high temperature and terminal moisture stress on dormancy in wheat (*Triticum aestivum L.*). *Field Crops Res* **103:** 139-153

Black M, Bewley JD, Halmer P (2006) The Encyclopedia of Seeds Science, Technology and Uses. Walling ford: CABI Publishing, 528

Bouchet S, Olatoye MO, Marla SR, Perumal R, Tesso T, Yu J, Morris GP (2017) Increased power to dissect adaptive traits in global sorghum diversity using a nested association mapping population. *Genetics* **206:** 573-585

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

Bradford KJ, Downie AB, Gee OH, Alvarado V, Yang H, Dahal P (2003) Abscisic acid and gibberellin differentially regulate expression of genes of the SNF1-related kinase complex in tomato seeds. *Plant Physiol* **132:** 1560-1576

Brady SM and Peter MC (2003) Hormone cross-talk in seed dormancy. *J Plant Growth Regul* **22:** 25-31

Breseghello F and Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum L*.) cultivars. *Genetics* **172:** 1165-1177

Broeckx T, Hulsmans S, Rolland F (2016) The plant energy sensor: evolutionary conservation and divergence of SnRK1 structure, regulation, and function. *J Exp Bot* **67**: 6215-6252

Buckler IV ES, Thornsberry JM (2002) Plant molecular diversity and applications to genomics. *Curr Opin Plant Biology* **5:** 107-111

Bush DS (1996) Effects of gibberellic acid and environmental factors on cytosolic calcium in wheat aleurone cells. *Planta* **199:** 89-99

Canadian Grain Commission (2021) https:// www. grain scanada. gc. ca/ en/ grain- quality/ grain- grading/ wheat- classes.

Castor L, Frederiken F (1977) Seed moulding of grain sorghums caused by *Fusarium* and *Curvularia*. *Proc Annu Phytopathol Soc* **4:** 151

Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S, Akhunov E (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc Natl Acad Sci* **110:** 8057-8062

Chan EK, Rowe HC, Kliebenstein DJ (2010) Understanding the evolution of defense metabolites in *Arabidopsis thaliana* using genome-wide association mapping. *Genetics* **185:** 991-1007

Chang C, Zhang HP, Feng JM, Yin B, Si HQ, Ma CX (2010) Identifying alleles of Viviparous-1B associated with pre-harvest sprouting in micro-core collections of Chinese wheat germplasm. *Mol Breed* **25:** 481-490

Chao S, Sharp PJ, Worland AJ, Warham EJ, Koebner RM, Gale MD (1989) RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. *Theor Appl Genet* **78:** 495-504

Chen CX, Cai SB, Bai GH (2008) A major QTL controlling seed dormancy and pre-harvest sprouting resistance on chromosome 4A in a Chinese wheat landrace. *Mol Breed* **21:** 351-358

Chen H, Xie W, He H, Yu H, Chen W, Li J, Yu R (2014) A high-density SNP genotyping array for rice biology and molecular breeding. *Mol plant* **7:** 541-553

Chen L, Bernhardt A, Lee J, Hellmann H (2015) Identification of Arabidopsis MYB56 as a novel substrate for CRL3BPM E3 ligases. *Mol Plant* **8:** 242-250

Chen L, Lee JH, Weber H, Tohge T, Witt S, Roje S, Fernie AR, Hellmann H (2013) Arabidopsis BPM proteins function as substrate adaptors to a cullin3-based E3 ligase to affect fatty acid metabolism in plants. *The Plant Cell* **25:** 2253-2264

Clavijo BJ, Venturini L, Schudoma C, Accinelli GG, Kaithakottil G, Wright J, Borrill P, Kettleborough G, Heavens D, Chapman H, Lipscombe J (2017) An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome Res* **27:** 885-896

Collard BC, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica* **142:** 169-196

Crossa J, Burgueno J, Dreisigacker S, Vargas M, Herrera-Foessel SA, Lillemo M, Ortiz R (2007) Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. *Genetics* **177:** 1889-1913

Curtis BC (2019) Wheat in the World. Available online: http://www.fao.org/3/y4011e/y4011e04

Dell'Acqua M, Gatti DM, Pea G, Cattonaro F, Coppens F, Magris G, Pè ME (2015) Genetic properties of the MAGIC maize population: a new platform for high-definition QTL mapping in *Zea mays*. *Genome Biol* **16:** 1-23

Derera NF (1989) "Breeding for preharvest sprouting tolerance," in Pre harvest Field Sprouting in Cereals, ed.N.F.Derera (Boca Raton, FL: CRC Press), 111–129

Derera NF, Bhatt GM, McMaster GJ (1977) On the problem of pre-harvest sprouting of wheat. *Euphytica* **26:** 299–308

Derera NF (1990) A perspective of sprouting research. In: International Symposium on Pre-Harvest Sprouting in Cereals **5:** 3-11

Dhakal A, Adhikari C, Manandhar D (2021) Samikshya Bhattarai, Sony Shrestha effect of abiotic stress in wheat: *a review. Reviews In Food and Agriculture* **2:** 69-72

Dixon LE, Bencivenga S, Boden SA (2018) A new opening for wheat seed production. *J Exp Bot* **69:** 341-343

Dubcovsky J, Dvorak J (2007) Genome plasticity a key factor in the success of polyploidy wheat under domestication. *Science* **316:** 1862–1866

Dvorak J, Luo MC, Yang ZL, Zhang HB (1998) The structure of *Aegilops tauschii* gene pool and the evolution of hexaploid wheat. *Theor Appl Genet* **97:** 657–670

El Baidouri M, Murat F, Veyssiere M, Molinier M, Flores R, Burlot L, Alaux M, Quesneville H, Pont C, Salse J (2017) Reconciling the evolutionary origin of bread wheat (*Triticum aestivum*). *New Phytol* **213:** 1477–1486

Emebiri LC, Oliver JR, Mrva K, Mares D (2010) Association mapping of late maturity α-amylase (LMA) activity in a collection of synthetic hexaploid wheat. *Mol Breed* **26:** 39– 49

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14:** 2611-2620

Eversole K, Feuillet C, Mayer KFX, Rogers J (2014) Slicing the wheat genome. *Introduction Sci* **345:** 285-287

Falcão-e-Cunha L, Peres H, Freire JP, Castro-Solla L (2004) Effects of alfalfa, wheat bran or beet pulp, with or without sunflower oil, on caecal fermentation and on digestibility in the rabbit. *Anim Feed Sci Technol* **117(1-2):**131-49

Finkelstein R, Reeves W, Ariizumi T, Steber C (2008) Molecular aspects of seed dormancy. *Annu Rev Plant Biol* **59:** 387-415

Flint-Garcia SA, Thornsberry JM, Buckler IV ES (2003) Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol* **54:** 357-374

Flintham JE (2000) Different genetic components control coat-imposed and embryo-imposed dormancy in wheat. *Seed Sci Res* **10:** 43-50

Flintham JE, Adlam R, Bassoi M, Holdsworth M, Gale M (2002) Mapping genes for resistance to sprouting damage in wheat. *Euphytica* **126:** 39–45

Flintham JE, Gale MD (1996) Dormancy gene maps in homoeologous cereal genomes. In: Noda K, Mares DJ (eds) Preharvest sprouting in cereals 1995. Center for Academic Societies Japan, Osaka, pp 143—149

Food and Agricultural Organization of the United Nations (FAO) (2020) (http://www.fao.org)

Food and Agricultural Organization of the United Nations (FAO) (2022) (http://www.fao.org)

Footitt S, Walley PG, Lynn JR, Hambidge AJ, Penfield S, Finch-Savage WE (2020) Trait analysis reveals DOG1 determines initial depth of seed dormancy, but not changes during dormancy cycling that result in seedling emergence timing. *New Phytol* **225:** 2035-2047

Fragoso CA, Moreno M, Wang Z, Heffelfinger C, Arbelaez LJ, Aguirre JA, Lorieux M (2017) Genetic architecture of a rice nested association mapping population. *G3: Genes Genom Genet* **7:** 1913-1926

Friebe B, Gill BS (1994) C-band polymorphism and structural rearrangements detected in common wheat (Triticum aestivum). *Euphytica* **78:** 1–5

Gale MD, Flintham JE, Devos KM (2002) Cereal comparative genetics and preharvest sprouting. *Euphytica* **126:** 21-25

Gao DY, Xu ZS, He Y, Sun YW, Ma YZ, Xia LQ (2014) Functional analyses of an E3 ligase gene AIP2 from wheat in Arabidopsis revealed its roles in seed germination and pre-harvest sprouting. *J Integr Plant Biol* **56:** 480-491

Gao F, Ayele BT (2014) Functional genomics of seed dormancy and preharvest sprouting in wheat: advances and prospects. *Front Plant Sci* **5:** 458

Gao F, Jordan MC, Ayele BT (2012) Transcriptional programs regulating seed dormancy and its release by after-ripening in common wheat (*Triticum aestivum* L). *Plant Biotechnol J* **10:** 465-476

Gao F, Rampitsch C, Chitnis VR, Humphreys GD, Jordan MC, Ayele BT (2013) Integrated analysis of seed proteome and mRNA oxidation reveals distinct post-transcriptional features regulating dormancy in wheat (*Triticum aestivum* L.). *Plant Biotechnol J* **11:** 921-32

Gatford KT, Hearnden P, Ogbonnaya F, Eastwood RF, Halloran GM (2002) Novel resistance to pre-harvest sprouting in Australian wheat from the wild relative *Triticum tauschii*. *Euphytica* **126:** 67–76

Gerjets T, Scholefield D, Foulkes MJ, Lenton JR, Holdsworth MJ (2010) An analysis of dormancy, ABA responsiveness, after-ripening and pre-harvest sprouting in hexaploid wheat (*Triticum aestivum* L.) caryopses. *J Exp Bot* **61:** 597-607

Giraldo P, Benavente E, Manzano-Agugliaro F, Gimenez E (2019) World-wide research trends on wheat and barley: a bibliometric comparative analysis. *Agron* **9:** 352

Glover NM, Redestig H, Dessimoz C (2016) Homoeologs: what are they and how do we infer them? *Trends Plant Sci* **21:** 609-621

Groos C, Gay G, Perretant MR (2002) Study of the relationship between pre-harvest sprouting and grain color by quantitative trait loci analysis in a white×red grain bread-wheat cross. *Theor Appl Genet* **104:** 39–47

Gu XY, Kianian SF, Foley ME (2004) Multiple loci and epistases control genetic variation for seed dormancy in weedy rice (*Oryza sativa*). *Genetics* **166:** 1503-1516

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 Physiol* **147:** 886-96

Gubler F, Millar AA, Jacobsen JV (2005) Dormancy release, ABA and pre-harvest sprouting. *Curr Opin Plant Biol* **8:** 183-187

Gupta PK, Kulwal PL, Jaiswal V (2019) Association mapping in plants in the post-GWAS genomics era. *Adv Genet* **104:** 75-154

Gupta PK, Varshney RK, Sharma PC, Ramesh B (1999) Molecular markers and their applications in wheat breeding. *Plant Breed* **118:** 369-90

Hagberg S (1961) Note on a simplified rapid method for determining alpha-amylase activity. *Cereal Chem* **38:** 202-203

Hao C, Wang L, Ge H, Dong Y, Zhang X (2011) Genetic diversity and linkage disequilibrium in Chinese bread wheat (*Triticum aestivum L*.) revealed by SSR markers. *PLOS one* **6:** e17279

Hedden P, Phillips AL (2000) Gibberellin metabolism: new insights revealed by the genes. *Trends Plant Sci* **5:** 523-530

Heun M, Scha¨fer-Pregl R, Klawan D, Castagna R, Accerbi M, Borghi B, Salamini F (1997) Site of einkorn wheat domestication identified by DNA fingerprinting. *Sci* **278:** 1312–1314

Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibria in finite populations. *Theor Popul Biol* **33:** 54-78

Himi E, Mares DJ, Yanagisawa A, Noda K (2002) Effect of grain colour gene (R) on grain dormancy and sensitivity of the embryo to abscisic acid (ABA) in wheat. *J Exp Bot* **53:** 1569-1574

Hincha DK, Zuther E, Heyer AG (2003) The preservation of liposomes by raffinose family oligosaccharides during drying is mediated by effects on fusion and lipid phase transitions. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1612:** 172-177

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* **6:** 95-108

Hlaváčová M, Klem K, Smutná P, Škarpa P, Hlavinka P, Novotná K, Rapantová B, Trnka M (2017) Effect of heat stress at anthesis on yield formation in winter wheat. *Plant Soil Env* **63:** 139–144

Hocking PJ and Mason L (1993) Accumulation, distribution and redistribution of dry matter and mineral nutrients in fruits of canola (oilseed rape), and the effects of nitrogen fertilizer and windrowing. *Aust J Agric Res* **44:** 1377-1388

Holm S (1979) A simple sequentially rejective multiple test procedure. *Scand J Stat* **1:** 65-70

Hong L, Brown J, Segerson NA, Rose JK, Roeder AH (2017) CUTIN SYNTHASE 2 maintains progressively developing cuticular ridges in Arabidopsis sepals. *Mol Plant* **10:** 560-574

Hoseney RC and Rogers DE (1990) The formation and properties of wheat flour doughs. *Crit Rev Food Sci Nutr* **29:** 73–93

Huang BE, George AW, Forrest KL, Kilian A, Hayden MJ, Morell MK, Cavanagh CR (2012) A multiparent advanced generation inter-cross population for genetic analysis in wheat. *Plant Biotech J* **10:** 826-839

Huang XQ, Börner A, Röder MS, Ganal MW (2002) Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. *Theor Appl Genet* **105:** 699–707

Humphreys DG, Noll J (2002) Methods for characterization of preharvest sprouting resistance in a wheat breeding program. *Euphytica* **126:** 61–65

Ibrahim AK, Zhang L, Niyitanga S, Afzal MZ, Xu Y, Zhang L, Zhang L, Qi J (2020) Principles and approaches of association mapping in plant breeding. *Trop Plant Biol* **13:** 212-224

Imtiaz M, Ogbonnaya FC, 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:** 1725–1736

Ingvarsson PK and Street NR (2011) Association genetics of complex traits in plants. *New Phytol* **189:** 909-922

International Wheat Genome Sequencing Consortium (IWGSC), Appels R, Eversole K, Stein N, Feuillet C, Keller B, Rogers J, Pozniak CJ, Choulet F, Distelfeld A, Poland J (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361(6403):** eaar7191

International Wheat Genome Sequencing Consortium (IWGSC), Klaus FX Mayer, Jane Rogers, Jaroslav Doležel, Curtis Pozniak, Kellye Eversole, Catherine Feuillet et al (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **345** (**6194):** 1251788

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res* **29:** e25

Jadid N, Mialoundama AS, Heintz D, Ayoub D, Erhardt M, Mutterer J, Meyer D, Alioua A, Van Dorsselaer A, Rahier A, Camara B (2011) DOLICHOL PHOSPHATE MANNOSE SYNTHASE1 mediates the biogenesis of isoprenyl-linked glycans and influences development, stress response, and ammonium hypersensitivity in Arabidopsis. *The Plant Cell* **23:** 1985-2005

Jaiswal V, Mir RR, Mohan A, Balyan HS, Gupta PK (2012) Association mapping for pre-harvest sprouting tolerance in common wheat (Triticum aestivum L.). *Euphytica* **188:** 89-102

Kashiwakura YI, Kobayashi D, Jikumaru Y, Takebayashi Y, Nambara E, Seo M, Kamiya Y, Kushiro T, Kawakami N (2016) Highly sprouting-tolerant wheat grain exhibits extreme dormancy and cold imbibition-resistant accumulation of abscisic acid. *Plant Cell Physiol* **57:** 715-732

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. *Theor Appl Genet* **102:** 980-985

Kawakami N, Miyake Y, Noda K (1997) ABA insensitivity and low ABA levels during seed development of non-dormant wheat mutants. *J Exp Bot* **48:** 1415-1421

Koebner RM and Summers RW (2003) 21st century wheat breeding: plot selection or plate detection? *Trends Biotech* **21:** 59-63

Koornneef M, Bentsink L, Hilhorst H (2002) Seed dormancy and germination. *Curr Opin Plant Biol* **5:** 33-36

Koornneef M, Jorna ML, Brinkhorst-Van der Swan DLC, Karssen CM (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. *Theo Appl Genet* **61:** 385-393

Korzun V (2002) Use of molecular markers in cereal breeding. *Cell Mol Biol Lett* **7:** 811-820

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Mott R (2009) A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS genet* ***5*****:** e1000551

Kucera B, Cohn MA, Leubner-Metzger G (2005) Plant hormone interactions during seed dormancy release and germination. *Seed Sci Res* **15:** 281-307

Kulwal P, Ishikawa G, Benscher D, Feng Z, Yu LX, Jadhav A, Sorrells ME (2012) Association mapping for pre-harvest sprouting resistance in white winter wheat. *Theo Appl Genet* **125:** 793-805

Kulwal PL, Kumar N, Gaur A, Khurana P, Khurana JP, Tyagi AK, Gupta PK (2005) Mapping of a major QTL for pre-harvest sprouting tolerance on chromosome 3A in bread wheat. *Theo Appl Genet* **111:** 1052-1059

Kulwal PL, Mir RR, Kumar S, Gupta PK (2010) QTL analysis and molecular breeding for seed dormancy and pre-harvest sprouting tolerance in bread wheat. *J Plant Biol* **37:** 59–74

Kumar N, Kulwal PL, Gaur A, Tyagi AK, Khurana JP, Khurana P, Balyan HS, Gupta PK (2006) QTL analysis for grain weight in common wheat. *Euphytica* **151:** 135–144

Kumar P, Yadava RK, Gollen B, Kumar S, Verma R K, Yadav S (2011) Nutritional contents and medicinal properties of wheat: a review. LSMR **22:** 1-10

Kumar S, Knox RE, Clarke FR, Pozniak CJ, DePauw RM, Cuthbert RD, 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

Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, Oropeza-Rosas MA, Zwonitzer JC, Kresovich S, McMullen MD, Ware D, Balint-Kurti PJ (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nat Genet* **43:**163-168

Lafiandra D, Riccardi G, Shewry PR (2014) Improving Cereal Grain Carbohydrates for Diet and Health. *J Cereal Sci* **5:** 312-326

Lechner E, Leonhardt N, Eisler H, Parmentier Y, Alioua M, Jacquet H, Leung J, Genschik P (2011) MATH/BTB CRL3 receptors target the homeodomain-leucine zipper ATHB6 to modulate abscisic acid signaling. *Develop cell* **21:** 1116-1128

Lefebvre V, North H, Frey A, Sotta B, Seo M, Okamoto M, Nambara E, Marion-Poll A (2006) Functional analysis of Arabidopsis *NCED6* and *NCED9* genes indicates that ABA synthesized in the endosperm is involved in the induction of seed dormancy. *Plant J* **45:** 309-319

Li L, Sheen J (2016) Dynamic and diverse sugar signaling. *Curr Opin Plant Biol* **33:** 116-125

Li L, Zhang Y, Zhang Y, Li M, Xu D, Tian X, Song J, Luo X, Xie L, Wang D, He Z (2021) Genome-wide linkage mapping for preharvest sprouting resistance in wheat using 15K single-nucleotide polymorphism arrays. *Front Plant Sci* **12:** 749206

Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta SL, Kebrom TH, Provart N, Patel R, Myers CR, Reidel EJ (2010) The developmental dynamics of the maize leaf transcriptome. *Nat Genet* **42:** 1060-1070

Li X, Li G, Li Y, Kong X, Zhang L, Wang J, Li X, Yang Y (2018) ABA receptor subfamily III enhances abscisic acid sensitivity and improves the drought tolerance of Arabidopsis. *Int J Mol Sci* **19:**1938

Lijavetzky D, Carolina Martínez M, Carrari F, Esteban Hopp H (2000) QTL analysis and mapping of pre-harvest sprouting resistance in sorghum. *Euphytica* **112:** 125-135

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. *Theor App Genet* **128:**1385-1395

Linkies A and Leubner-Metzger G (2012) Beyond gibberellins and abscisic acid: how ethylene and jasmonates control seed germination. *Plant cell rep* **31:** 253-270

Liu S, Pumphrey M, Gill B, Trick H, Zhang J, Dolezel J, Chalhoub B, Anderson J (2008) Toward positional cloning of Fhb1, a major QTL for Fusarium head blight resistance in wheat. *Cer Res Comm* **36:** 195-201

Liu S, Sehgal SK, Li J, Lin M, Trick HN, Yu J, Bai G (2013) Cloning and characterization of a critical regulator for preharvest sprouting in wheat. *Genetics* **195:** 263-273

Liu ZH, Faris JD, Meinhardt SW, Ali S, Rasmussen JB, Friesen TL (2004) Genetic and physical mapping of a gene conditioning sensitivity in wheat to a partially purified host-selective toxin produced by *Stagonospora nodorum*. *Phytopath* **94:** 1056-60

Lohwasser U, Röder MS, Börner A (2005) QTL mapping of the domestication traits pre-harvest sprouting and dormancy in wheat (*Triticum aestivum L.*). *Euphytica* **143:** 247-249

Lothier J, Lasseur B, Le Roy K, Van Laere A, Prud'Homme MP, Barre P, Van den Ende W, Morvan-Bertrand A (2007) Cloning, gene mapping, and functional analysis of a fructan 1-exohydrolase (1-FEH) from *Lolium perenne* implicated in fructan synthesis rather than in fructan mobilization. *J Exp Bot* **58:** 969-1983

Lovegrove A and Hooley R (2000) Gibberellin and abscisic acid signalling in aleurone. *Trends Plant Sci* **5:** 102-110

Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S (2012) SNP markers and their impact on plant breeding. *Int J Plant Genomics* **1:**12

Manly KF, Cudmore Jr RH, Meer JM (2001) Map Manager QTX, cross-platform software for genetic mapping. *Mamm Genome* **12:** 930-932

Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, The International Wheat Genome Sequencing Consortium, Jakobsen KS, Wulff BBH, Steuernagel B, Mayer KFX and Olsen OA (2014) Ancient hybridizations among the ancestral genomes of bread wheat. *Science* **345** (6194):1250092

Mares DJ and Mrva K (2014) Wheat grain preharvest sprouting and late maturity alpha-amylase. *Planta* **240:** 1167-1178

Mares DJ, 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. *Theor Appl Genet* **111:** 1357–1364

Marth GT, Korf I, Yandell MD, Yeh RT, Gu Z, Zakeri H, Gish WR (1999) A general approach to single-nucleotide polymorphism discovery. *Nat Genet* **23:** 452-456

Maurer A, Draba V, Jiang Y, Schnaithmann F, Sharma R, Schumann E, Pillen K (2015) Modelling the genetic architecture of flowering time control in barley through nested association mapping. *BMC genomics* **16:** 1-12

McCallum BD, DePauw RM (2008) A review of wheat cultivars grown in the Canadian prairies. *Can J Plant Sci* **88:** 649–677

McCarty DR (1995) Genetic control and integration of maturation and germination pathways in seed development. *Annu Rev Plant Biol* **46:** 71-93

McKibbin RS, Wilkinson MD, Bailey PC, Flintham JE, Andrew LM, Lazzeri PA, Gale MD, Lenton JR, Holdsworth MJ (2002) Transcripts of Vp-1 homeologues are misspliced in modern wheat and ancestral species. *Proc Nat Acad Sci* **99:** 10203-10208

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Buckler ES (2009) Genetic properties of the maize nested association mapping population. *Science* **325:** 737-740

Millar AA, Jacobsen JV, Ross JJ, Helliwell CA, Poole AT, Scofield G, Gubler F (2006) Seed dormancy and ABA metabolism in Arabidopsis and barley: the role of ABA 8′-hydroxylase. *Plant J* **45:** 942-954

Mohan A, Kulwal P, Singh R, Kumar V, Mir RR, Kumar J, Prasad M, Balyan HS, Gupta PK (2009) Genome-wide QTL analysis for pre-harvest sprouting tolerance in bread wheat. *Euphytica* **168:** 319–329

Moot DJ and Every D (1990) A comparison of bread baking, falling number, α-amylase assay and visual method for the assessment of pre-harvest sprouting in wheat. *J Cereal Sci* **11:** 225-234

Mori M, Uchino N, Chono M, Kato K, Miura H (2005) Mapping QTLs for grain dormancy on wheat chromosome 3A and group 4 chromosomes, and their combined effect. *Theor Appl Genet* **110:** 1315–1323

Morimoto K, Ohama N, Kidokoro S, Mizoi J, Takahashi F, Todaka D, Mogami J, Sato H, Qin F, Kim JS, Fukao Y (2017) BPM-CUL3 E3 ligase modulates thermotolerance by facilitating negative regulatory domain-mediated degradation of DREB2A in Arabidopsis. *Proc Nat Acad Sci* **114:** 8528-8536

Munkvold JD, Tanaka J, Benscher D, Sorrells ME (2009) Mapping quantitative trait loci for preharvest sprouting resistance in white wheat. *Theor App Genet* **119:** 1223-1235

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES (2009) Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell* **21:** 2194-2202

Nakamura S and Toyama T (2001) Isolation of a VP1 homologue from wheat and analysis of its expression in embryos of dormant and non-dormant cultivars. *J Exp Bot* **52:** 875-876

Nakamura S, Abe F, Kawahigashi H, Nakazono K, Tagiri A, Matsumoto T, Miura H (2011) A wheat homolog of MOTHER OF FT AND TFL1 acts in the regulation of germination. *Plant Cell* **23:** 3215-3229

Nambara E and Marion-Poll A (2005) Abscisic acid biosynthesis and catabolism. *Annu Rev Plant Biol* **56:**165–185

Nambara E, Okamoto M, Tatematsu K, Yano R, Seo M, Kamiya Y (2010) Abscisic acid and the control of seed dormancy and germination. *Seed Sci Res* **20:** 55-67

Nelson JC, Sorrells ME, Van Deynze AE, Lu YH, Atkinson M, Bernard M, Leroy P, Faris JD, Anderson JA (1995) Molecular mapping of wheat: major genes and rearrangements in homoeologous groups 4, 5, and 7. *Genetics* **141:** 721-731

Neumann K, Kobiljski B, Denčić S, Varshney RK, Börner A (2011) Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum L.*). *Mol Breed* **27:** 37-58

Nielsen NH, Backes G, Stougaard J, Andersen SU, Jahoor A (2014) Genetic diversity and population structure analysis of European hexaploid bread wheat (*Triticum aestivum L.*) varieties. *PloS one* **9:** e94000

Nonogaki H, Barrero JM and Li C (2018) Seed dormancy, germination, and Pre-harvest sprouting. *Front Plant Sci* **9:** 1783

Nonogaki H, Bassel GW, Bewley JD (2010) Germination—still a mystery. *Plant Sci* **179:** 574-581

Nordborg M, Tavaré S (2002) Linkage disequilibrium: what history has to tell us. *Trends Genet* **18:** 83-90

Ogbonnaya FC, Imtiaz M, Ye G, Hearnden PR, Hernandez E, Eastwood RF, Van Ginkel M, Shorter SC, Winchester JM (2008) Genetic and QTL analyses of seed dormancy and preharvest sprouting resistance in the wheat germplasm CN10955. *Theor Appl Genet* **16:** 891-902

Olaerts H and Courtin CM (2018) Impact of preharvest sprouting on endogenous hydrolases and technological quality of wheat and bread: a review. *Compr Rev Food Sci Food Saf* **17:** 698-713

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 App Genet* **106:** 1491-1496

Pan Q, Li L, Yang X, Tong H, Xu S, Li Z, Li W, Muehlbauer G, Li J, Yan J (2016). Genome-wide recombination dynamics are associated with phenotypic variation in maize. *New Phytol* **210:** 1083–1094

Pasam RK, Sharma R, Malosetti M, van Eeuwijk FA, Haseneyer G, Kilian B, Graner A (2012) Genome-wide association studies for agronomical traits in a worldwide spring barley collection. *BMC Plant Biol* **12:**1-22

Paterson AH, Sorrells ME, Obendorf RL (1989) Methods of evaluation for preharvest sprouting resistance in wheat breeding programs. *Can J Plant Sci* **69:** 681–689

Patricia G, Elena B, Francisco MA, Estela G (2019) Worldwide research trends on wheat and barley: A Bibliometric comparative analysis. *Agron* **9:** 352

Peng JH, Sun D, Nevo E (2011) Domestication evolution, genetics, and genomics in wheat. *Mol Breed* **28:** 281

Pérez-Flores L, Carrari F, Osuna-Fernández R, Rodríguez MV, Enciso S, Stanelloni R, Sánchez RA, Bottini R, Iusem ND, Benech-Arnold RL (2023) Expression analysis of a GA 20-oxidase in embryos from two sorghum lines with contrasting dormancy: possible participation of this gene in the hormonal control of germination. *J Exp Bot* **54:** 2071-2079

Piepho HP, Möhring J, Melchinger AE, Büchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* **161:** 209-228

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Gen* **38:** 904-909

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. *Am J Hum Genet* **67:** 170-181

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. *Curr Opin Plant Biol* **5:** 94-100

Rafalski JA (2010) Association genetics in crop improvement. *Curr Opin Plant Biol* **13:** 174-180

Randhawa HS, Puchalski BJ, Frick M, Goyal A, Despins T, Graf RJ, Laroche A, Gaudet D (2012) A Stripe rust resistance among western Canadian spring wheat and triticale varieties. *Can J Plant Sci* **92:** 713722

Reif JC, Gowda M, Maurer HP, Longin CF, Korzun V, Ebmeyer E, Bothe R, Pietsch C, Würschum T (2011) Association mapping for quality traits in soft winter wheat. *Theor Appl Genet* **122:** 961-970

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler IV ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc Nat Acad Sci* **98:**11479-11484

Risch N (1992) Genetic linkage: Interpreting LOD scores. *Science* **255:** 803-804

Rodríguez MV, Barrero JM, Corbineau F, Gubler F, Benech-Arnold RL (2015) Dormancy in cereals (not too much, not so little): about the mechanisms behind this trait. *Seed Sci Res* **25(2):** 99-119

Ross-Ibarra J, Morrell PL, Gaut BS (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc Natl Acad Sci* **104:** 8641-8648

Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Jan, Svensson JT (2006) "Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties." *Proc Natl Acad Sci* **103:** 18656-18661

Roy JK, Prasad M, Varshney RK, Balyan HS, Blake TK, Dhaliwal HS, Edwards KJ, Gupta PK (1999) Identification of a microsatellite on chromosomes 6B and a STS on 7D of bread wheat showing an association with preharvest sprouting tolerance. *Theor Appl Genet* **99:** 336-340

RW King and RA Richards (1984) Water uptake in relation to pre-harvest sprouting damage in wheat: ear characteristics *Aus J Agri Res* **35:** 327 – 336

Sannemann W, Huang BE, Mathew B, Leon J (2015) Multi-parent advanced generation inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a proof of concept. *Mol Breed* **35:** 86

Schmutzer T, Samans B, Dyrszka E, Ulpinnis C, Weise S, Stengel D, Scholz U (2015) Species-wide genome sequence and nucleotide polymorphisms from the model allopolyploid plant Brassica napus. *Sci data* **2:** 1-9

Seki M, Umezawa T, Urano K, Shinozaki K (2007) Regulatory metabolic networks in drought stress responses. *Curr Opin Plant Biol* **10:** 296-302

Seo M and Koshiba T (2002) Complex regulation of ABA biosynthesis in plants. *Trends Plant Sci **7:*** 41-48

Shao M, Bai G, Rife TW, Poland J, Lin M, Liu S, Chen H, Kumssa T, Fritz A, Trick H, Li Y (2018) QTL mapping of pre-harvest sprouting resistance in a white wheat cultivar Danby. *Theor Appl Genet* **131:** 1683-1697

Shewry PR (2009) Wheat. *J Exp Bot* 1537–1553

Shewry PR and Hey S (2015) The contribution of wheat to human diet and health. *Food Energy Secur* **4:**178-202

Shu K, Liu XD, Xie Q, He ZH (2016) Two faces of one seed: hormonal regulation of dormancy and germination. *Mol Plant* **9:** 34-45

Simpson GM and Clutter E (1978) Metabolic regulation of dormancy in seeds–a case history of the wild oat (*Avena fatua*). Dormancy and Developmental Arrest: *Exp Anal Plants Animals* 167-220

Singh AK, Knox RE, Clarke JM, Clarke FR, Singh A, Depauw RM, Cuthbert R D (2014) Genetics of pre-harvest sprouting resistance in a cross of Canadian adapted durum wheat genotypes. *Mol breed: new strategies in plant improvement* **33:** 919-929

Singh RP, Singh PK, Rutkoski J, Hodson DP, He X, Jørgensen LN, Hovmøller MS, Huerta-Espino J (2016) Disease impact on wheat yield potential and prospects of genetic control. *Annu Rev Phytopathol* **54:** 303-322

Škiljaica A, Lechner E, Jagić M, Majsec K, Malenica N, Genschik P, Bauer N (2020) The protein turnover of Arabidopsis BPM1 is involved in regulation of flowering time and abiotic stress response. *Plant Mol Biol* **102:** 359-372

Smith SE, Kuehl RO, Ray IM, Hui R, Soleri D (1998) Evaluation of simple methods for estimating broad-sense heritability in stands of randomly planted genotypes. *Crop Sci* **38:** 1125-1129

Somyong S, Ishikawa G, Munkvold JD, Tanaka J, Benscher D, Cho YG, Sorrells ME (2014) Fine mapping of a preharvest sprouting QTL interval on chromosome 2B in white wheat. *Theor App Genet* **127:** 1843-1855

Statistics Canada (2020) https://www150.statcan.gc.ca/n1/daily-quotidien/201203/dq201203b-eng.htm

Statistics Canada (2021) https://www150.statcan.gc.ca/n1/daily-quotidien/201203/dq201203b-eng.htm

Steinbach HS, Benech-Arnold RL, Sanchez RA (1997) Hormonal regulation of dormancy in developing sorghum seeds. *Plant Physiol* **113:** 149-154

Stich B, Maurer HP, Melchinger AE, Frisch M, Heckenberger M, van der Voort JR, Peleman J, Sørensen AP, Reif JC (2006) Comparison of linkage disequilibrium in elite European maize inbred lines using AFLP and SSR markers. *Mol Breed* **17:** 217-226

Su Q, Li J (2018) QTL detection for kernel size and weight in bread wheat (*Triticum aestivum L.*) using a high-density SNP and SSR-based linkage map. *Front Plant Sci* **9:** 1484

Sun TP (2008) Gibberellin metabolism, perception and signaling pathways in Arabidopsis. The Arabidopsis Book/American Society of Plant Biologists 6

Tamura K, Adachi Y, Chiba K, Oguchi K, Takahashi H (2002) Identification of Ku70 and Ku80 homologues in *Arabidopsis thaliana*: evidence for a role in the repair of DNA double-strand breaks. *Plant J* **29:** 771-781

Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. *Biotech* **7:** 257-264

Thakur JK, Agarwal P, Parida S, Bajaj D, Pasrija R (2013) Sequence and expression analyses of KIX domain proteins suggest their importance in seed development and determination of seed size in rice, and genome stability in Arabidopsis. *Mol Genet Gen* **288:** 329-346

Thomason W, Hughes KR, Griffey CA, Parrish DJ, Barbeau WE (2019) Understanding pre-harvest sprouting of wheat. Virginia Tech, Virginia State University, Publication 424-060, available at https://www.pubs.ext.vt.edu/424/424-060/424-060.html

Thomason WE, Brooks WS, Griffey CA, Vaughn ME (2009) Hulless barley seeding rate effects on grain yield and yield components. *Crop Sci* **49:** 342-346

Tommasini L, Schnurbusch T, Fossati D, Mascher F, Keller B (2007) Association mapping of *Stagonospora nodorum* blotch resistance in modern European winter wheat varieties. *Theor Appl Genet* **115:** 697-708

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, Ogawa T, Takenouchi Y, Tadamura K, Wu J, Ogihara Y (2016) A causal gene for seed dormancy on wheat chromosome 4A encodes a MAP kinase kinase. *Curr Biol* **26:** 782-787

Tuan PA, Kumar R, Rehal PK, Toora PK, Ayele BT (2018) Molecular Mechanisms Underlying Abscisic Acid/Gibberellin Balance in the Control of Seed Dormancy and Germination in Cereals. *Front Plant Sci* **9:** 668

Tuberosa R and Salvi S (2004) QTLs and genes for tolerance to abiotic stress in cereals. In Cereal genomics (pp. 253-315). Springer, Dordrecht

Tyagi S and Gupta PK (2012) Meta-analysis of QTLs involved in pre-harvest sprouting tolerance and dormancy in bread wheat. *Triticeae Genomics Genet* **3:** 9-24

Tye-Din JA , Galipeau HJ, Agardh D (2018) Celiac Disease: A Review of Current Concepts in Pathogenesis, Prevention, and Novel Therapies. *Front Pediatr* **21:** 350

Ullrich SE, Clancy JA, del Blanco IA, Lee H, Jitkov VA, Han F, Matsui K (2008) Genetic analysis of preharvest sprouting in a six-row barley cross. *Mol Breed* **21:** 249-259

Ullrich SE, Hayes PM, Dyer WE, Blake TK, Clancy JA (1993) Quantitative trait locus analysis of seed dormancy in Steptoe barley. In: Walker-Simmons MK, Ried JL (eds) Preharvest spouting in cereals 1992. Am Assoc Cereal Chem, St Paul, Minn, pp, 136-145

Ullrich SE, Lee H, Clancy JA, Del Blanco IA, Jitkov VA, Kleinhofs A, Han F, Prada D, Romagosa I, Molina-Cano JL (2009) Genetic relationships between preharvest sprouting and dormancy in barley. *Euphytica* **168:** 331-345

UN (2019) *World Population Prospects 2019*. New York, NY: United Nations, Department of Economic and Social Affairs, Population Division. Available online at: https://population.un.org/wpp/

Updegraff EP, Zhao F, Preuss D (2009) The extracellular lipase EXL4 is required for efficient hydration of Arabidopsis pollen. *Sex Plant Rep* **22:** 197-204

Vágújfalvi A, Galiba G, Cattivelli L, Dubcovsky J (2003) The cold-regulated transcriptional activator Cbf3 is linked to the frost-tolerance locus Fr-A2 on wheat chromosome 5A. *Mol Genet Genom* **269:** 60-67

Varughese G, Barker T, Saari E (1987) Triticale. CIMMYT, Mexico **11(3):** 225-234

Vetch JM, Stougaard RN, Martin JM, Giroux MJ (2019) Revealing the genetic mechanisms of pre-harvest sprouting in hexaploid wheat (*Triticum aestivum* L.). *Plant Sci* **281:** 180-185

Vos P, Hogers R, Bleeker M, Reijans M, Lee TV, Hornes M, Fritjers A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: A new concept for DNA fingerprinting. *Nucl Acids Res* **23:** 4407-4414

Walker-Simmons M and Sesing J (1990). Temperature effects on embryonic abscisic acid levels during development of wheat grain dormancy. *J Plant Grow Regul* **9:** 51-56

Wang J, Chu S, Zhang H, Zhu Y, Cheng H, Yu D (2016) "Development and application of a novel genome-wide SNP array reveals domestication history in soybean." *Sci Rep* **6:** 1-10

Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M (2014) "Characterization of polyploid wheat genomic diversity using a high-density 90000 single nucleotide polymorphism array." *Plant Biotech J* **12:** 787-796

Wang X, Liu H, Liu G, Mia MS, Siddique KH, Yan G (2019) Phenotypic and genotypic characterization of near-isogenic lines targeting a major 4BL QTL responsible for pre-harvest sprouting in wheat. *BMC Plant Biol* **19:** 1-10

Wang Y, Wang XL, Meng JY, Zhang YJ, He ZH, Yang Y (2016) Characterization of Tamyb10 allelic variants and development of STS marker for pre-harvest sprouting resistance in Chinese bread wheat. *Mol Breed* **36:** 1-2

Watanabe N, Ikebata N (2000) The effects of homoeologous group 3 chromosomes on grain colour dependent seed dormancy and brittle rachis in tetraploid wheat. *Euphytica* **115:** 215-220

Weber H, Hellmann H (2009) *Arabidopsis thaliana* BTB/POZ-MATH proteins interact with members of the ERF/AP2 transcription factor family. *The FEBS journal* **276:** 6624-6635

Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A, Kilian A (2004) Diversity Arrays Technology (DArT) for whole-genome profiling of barley. *Proc Nat Aca Sci.* **26:** 9915-9920

Werner JD, Borevitz JO, Uhlenhaut NH, Ecker JR, Chory J, Weigel D (2005) FRIGIDA-independent variation in flowering time of natural *Arabidopsis thaliana* accessions. *Genetics* **170:** 1197-1207

Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* **18:** 6531-6535

Winfield MO, Allen AM, Burridge AJ, Barker GLA, Benbow HR, Wilkinson PA, Coghill J (2016)"High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotech J* **14:** 1195-1206

Winter P, Kahl G (1995) Molecular marker technologies for plant improvement. *World J Microbiol Biotechnol*.**11:** 438-448

Wu Z, Liang S, Song W, Lin G, Wang W, Zhang H, Han Z, Chai J (2017) Functional and structural characterization of a receptor-like kinase involved in germination and cell expansion in Arabidopsis. *Front Plant Sci* **8:**1999

Xia LQ, Ganal MW, Shewry PR, He ZH, Yang Y, Röder MS (2008) Exploiting the diversity of Viviparous-1 gene associated with pre-harvest sprouting tolerance in European wheat varieties. *Euphytica* **159:** 411-417

Xiao Y, Tong H, Yang X, Xu S, Pan Q, Qiao F, Yan J (2016) Genome-wide dissection of the maize ear genetic architecture using multiple populations. *New Phytol* **210:** 1095-1106

Xu Y, Li P, Yang Z, Xu C (2017) Genetic mapping of quantitative trait loci in crops. *The Crop Journal* **5:** 175-84

Yamaguchi S (2008) Gibberellin metabolism and its regulation. *Annu Rev Plant Biol* **59:** 225-51

Yang CC, Wang YT, Hsiao YY, Doudeva LG, Kuo PH, Chow SY, Yuan HS (2010) Structural and biochemical characterization of CRN-5 and Rrp46: an exosome component participating in apoptotic DNA degradation. *RNA* **16:** 1748-1759

Yang Y, Zhao XL, Xia LQ, Chen XM, Xia XC, Yu Z, He ZH, Röder M (2007) Development and validation of a Viviparous-1 STS marker for pre-harvest sprouting tolerance in Chinese wheats. *Theo App Genet* **115:** 971-980

Yu J and Buckler ES (2006) Genetic association mapping and genome organization of maize. *Curr Opin Biotech* **17:** 155-160

Yu LX (2017) Identification of single-nucleotide polymorphic loci associated with biomass yield under water deficit in alfalfa (*Medicago sativa* L.) using genome-wide sequencing and association mapping. *Front Plant Sci* **29:** 1152

Zeevaart JAD, Creelman RA (1988) Metabolism and physiology of abscisic acid. *Annu Rev Plant Physiol Plant Mol Biol* **39:** 439

Zhang F, Zou YN, Wu QS (2018) Quantitative estimation of water uptake by mycorrhizal extraradical hyphae in citrus under drought stress. *Sci Hort* **229:**132-136

Zhang LY, Marchand S, Tinker NA, Belzile F (2009) Population structure and linkage disequilibrium in barley assessed by DArT markers. *Theor Appl Genet* **119:** 43-52

Zhang X, Chen S, Sun H, Wang Y, Shao L (2010) Water use efficiency and associated traits in winter wheat cultivars in the North China Plain. *Agric Water Manag* **97:** 1117-1125

Zhang XQ, 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.). Mol Breed* **22:** 227-236

Zhang Y, Miao X, Xia X, He Z (2014) Cloning of seed dormancy genes (*TaSdr*) associated with tolerance to pre-harvest sprouting in common wheat and development of a functional marker. *Theor App Genet* **127:** 855-866

Zhang Y, Xia X, He Z (2017) The seed dormancy allele TaSdr-A1a associated with pre-harvest sprouting tolerance is mainly present in Chinese wheat landraces. *Theor Appl Genet* **130:** 81-89

Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH, McCouch SR (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat commun* **2:** 1-10

Zhou T, Pan J, Zhang P, Wei S, Han T (2017) "Mapping winter wheat with multi-temporal SAR and optical images in an urban agricultural region. *Sensors* **17:** 1210

Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* **44:** 821-824

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. *Plant Genome* **1:** 5-20

Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* **20:**176-183

APPENDIX

Appendix 1. Using Bonferroni Correction, a comparison between the number of significant markers in the dataset aggregated from all trial environments and the number in the individual trial environments was made.

Significance thresholds	Number of markers common in trial environments[a]	Number of markers based on the dataset pooled from all trial environments[a]
Bonferroni Correction of *P* value distribution (-log10 (*P*) > 4.27)	4 (MOR16 and MOR17) 6 (GH21and Point21) 4 (CONV21and ORG21) 3 (In all locations)	14

[a]MOR16, Morden 2016; MOR17, Morden 2017; GH21, Winnipeg Greenhouse 2021; Point21, Point 2021 (University of Manitoba Point Field Station); EDM CONV21, Edmonton Conventional 2021; EDM ORG21, Edmonton Organic 2021

Appendix 2. Significant markers identified using dataset pooled from all environments using Bonferroni correction[a]

QTL[b]	Marker name	Chr[c]	Position (bp)[d]	Position (cM)[e]	-log10 (P)[f]	R^2 (%)[g]
QTL.4A-1	Kukri_c27874_515	4A	325,556,942	65.95	9.31	24.02
QTL.4A-1	wsnp_Ex_c13031_20625900	4A	603,103,256	65.95	9.87	25.78
QTL.4A-1	wsnp_Ex_rep_c66324_64493429	4A	603,104,997	65.95	8.41	22.33
QTL.4A-1	BS00072025_51	4A	607,696,954	66.27	7.4	18.6
QTL.4A-1	Kukri_c12563_52	4A	604,623,882	66.27	5.67	13.3
QTL.4A-1	Jagger_c4331_105	4A	605,291,230	66.68	6.11	14.62
QTL.4A-1	Excalibur_c40202_114	4A	563,490,514	66.95	6.58	17.55
QTL.4A-1	IAAV3132	4A	607,470,479	66.95	5.68	13.42
QTL.4A-1	BS00030838_51	4A	604,555,417	66.95	5.12	12.14
QTL.4A-1	Excalibur_c8658_335	4A	607,692,372	66.95	5.58	13.87
QTL.4A-1	Tdurum_contig14947_611	4A	608,830,565	66.95	5.82	13.89
QTL.4A-1	RAC875_c18090_752	4A	607,696,554	66.95	5.61	13.34
QTL.4A-1	Ra_c22219_350	4A	608,682,947	66.95	5.87	13.84
QTL.4A-1	Ra_c23848_305	4A	609,573,920	69.18	5.15	11.89

[a]Environment as described in Table 3.1
[b]Putative QTL containing one or more SNPs on the basis of LD decay distance
[c]Chromosome
[d]Marker position in bp obtained from the wheat Chinese Spring Ref Seq v2.1 assembly (IWGSC, 2018)
[e]Marker positions in cM derived from SNPs consensus genetic maps (Wang *et al*., 2014)
[f]Markers selected on the basis of Bonferroni Correction P value threshold of -log10 (*P*) 4.27
[g]R^2 (%) represents the percentage of phenotypic variation accounted by each significant SNP

ABBREVIATIONS

ABA	abscisic acid
AFLP	amplified fragment length polymorphism
BC	back cross
BC	bonferroni correction
BIM	bayesian interval mapping
BLAST	basic local alignment search tool
BLUE	best linear unbiased estimator
BLUP	best linear unbiased predictor
CaM	calmodulin
CIM	composite interval mapping
cGMP	cyclic guanosine 3',5'-monophosphate
CRLK	calcium/calmodulin-regulated receptor-like kinase
DArT	diversity arrays technology
DH	double haploid
DNA	deoxyribonucleic acid
FAO	food and agriculture organization
FDR	false discovery rate

FHB	fusarium head blight
FN	falling number
GA	gibberellin
GI	germination index
GLM	general linear model
GWAS	genome wide association study
IWGSC	International Wheat Genome Sequencing Consortium
LD	linkage disequilibrium
LOD	logarithm of odds
MAF	minor allele frequency
MAGIC	multiparent advanced generation inter-cross
MAPK	mitogen activated protein kinase
MCMC	Markov Chain Monte Carlo
MIM	multiple interval mapping
MKKK	mitogen activated protein kinase kinase kinase
MLM	mixed linear model
NAM	nested association mapping
NCED	nine-cis-epoxycarotenoid dioxygenase

NIL	near-isogenic line
PA	phaseic acid
PCA	principal component analysis
PHS	pre-harvest sprouting
QQ	quantile-quantile
QTL	quantitative trait loci
RAPD	random amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
RIL	recombinant inbred lines
RNA	ribonucleic acid
mRNA	messanger ribonucleic acid
ROAM	Random-open-parent Association Mapping
SI	sprouting index
SIM	simple interval mapping
SMA	single marker analysis
SNP	single nucleotide polymorphism
SSR	simple sequence repeats
STS	sequence-tagged sites

TASSEL	trait analysis by association, evolution and linkage
UN	United Nations