GENOMIC ANALYSIS OF PRE-HARVEST SPROUTING RESISTANCE IN BARLEY

By

Gurkamal Kaur

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

Copyright © 2021 by Gurkamal Kaur

ACKNOWLEDGEMENT

First and foremost, praises and gratitude to the God, due to his grace I have successfully completed my research project.

I would like to express genuine and sincere thanks to my supervisor, Dr. Belay Ayele for providing me the opportunity to work on this project. His dedication, keen interest and continuous encouragement have always inspired me. I really appreciate him for understanding and supporting me in conducting the research and consistently helping me in presenting my work as clearly as possible. I cannot express enough thanks to my co-supervisor, Dr. Ana Badea for enormous motivation, prompt guidance and suggestions, and also her team for setting up the trials at Brandon and threshing the samples as well as empowering my software skills. I am highly grateful to my committee member, Dr. Curt McCartney for his timely assistance and detailed discussions and his team for setting up the trail at Morden. I extend my appreciation to Dr. Marta Izydorczyk for her valuable feedback and great support. Special thanks to all my past and present fellow lab members, Dr. Anh Pham, Dr. Murali, Dr. Manisha, Dr. Aihua, Uzzal, Ginny, Ramanpreet, Riya, Parneet, Puneet, Rupinder and Nguyen for always guiding and helping me.

I offer my deepest gratitude to my father Mr. Manpreet Singh, mother Mrs. Gurtej Kaur, brother Gursharan Singh and my maternal grandmother Mrs. Swaranjeet Kaur for their infinite love, prayers, moral and emotional support throughout my academic journey. Words cannot express my appreciation to Mr. Gurveer Singh Sidhu who is truly a blessing from God in my life, I could never repay his immense love, care and support. I am highly indebted to Ankita Thapar for her love and ever willing help, and her parents for always being by my side. My heartfelt thanks to my cousins and friends, Jindu, Piyush, Harry, Varun, Monu, Kirat, Parul, Neetu, Gagan, Simran, Tanu, Raman and Handeep for their endless support.

TABLE OF CONTENTS

ACKNOWLEDGEMENT i

TABLE OF CONTENTS ii

ABSTRACT ix

FOREWORD x

1.0 GENERAL INTRODUCTION 1

2.0 LITERATURE REVIEW 4

2.1 Barley: a major cereal crop 4

2.2 Origin and domestication of barley 5

2.3 Barley genome and genomic resources 6

2.4 Barley production and use 7

2.5 Factors affecting barley production 8

2.6 Pre-harvest sprouting 9

2.7 Seed germination 11

2.8 Seed dormancy 12

2.9 Classification of seed dormancy 12

2.10 Factors affecting seed dormancy and germination 14

2.10.1 Role of plant hormones 14

2.10.2 Role of environmental factors 16

2.11 Genetics of dormancy and PHS 17

2.11.1 Use of molecular markers 17
2.11.2 QTL analysis 18
2.11.3 Linkage mapping 19
2.11.4 Association mapping and LD 21
2.11.5 Factors affecting GWAS 23
2.11.5.1 Population Structure 23
2.11.5.2 Heritability 24
2.11.5.3 Sample size 25
2.11.6 Steps involved in GWAS 25
2.11.7 Approaches for association mapping 28
2.11.8 Statistical power in GWAS 29
2.11.9 Advantages and limitations of GWAS 29
2.11.10 Other approaches of association mapping 30
2.11.11 QTLs/genes for PHS 31
3.0 GENOMIC ANALYSIS OF PRE-HARVEST SPROUTING RESISTANCE IN BARLEY 35
Abstract 35
3.1 Introduction 36
3.2 Material and methods 39
3.2.1 Plant material and field trials 39
3.2.2 Phenotyping 40

3.2.2.1 Evaluation of dormancy/PHS phenotype 40
3.2.2.2 Statistical data analysis 40
3.2.3 Genotyping 41
3.2.4 Estimation of LD 42
3.2.5 Population Structure and kinship 42
3.2.6 Association Analysis for dormancy/ PHS 43
3.2.7 QTL detection 44
3.2.8 Identification of candidate genes 44
3.2.9 Expression analysis of candidate genes 45
3.2.9.1 Tissue collection and RNA extraction 45
3.2.9.2 DNAse digestion and cDNA synthesis 46
3.2.9.3 Candidate gene primer sequences 46
3.2.9.4 Real-time qPCR assay 48
3.2.9.5 Statistical analysis 48
3.3 Results 48
3.3.1 Characterization of variation in dormancy level and heritability 48
3.3.2 Analysis of variance and environmental correlation 51
3.3.3 Linkage disequilibrium 52
3.3.4 Population structure analysis 53
3.3.5 Marker-trait associations 55

3.3.6 Detection of significant markers 56

3.3.7 QTL detection 59

3.3.8 Identification of candidate genes 61

3.3.9 Expression analysis of candidate genes 62

3.4 Discussion 66

4.0 GENERAL DISCUSSION AND CONCLUSION 76

LITERATURE CITED 80

APPENDIX 119

Appendix 1. Comparison between number of significant markers among the individual trial environments and the dataset pooled from all trial environments using correction methods FDR at α = 0.05 and lowest 0.1 percentile of the *p* value distribution 119

Appendix 2. Significant SNPs associated with PHS based on the dataset pooled from all trial environments using lowest 0.1 percentile *p* value threshold 120

Appendix 3. Physical map of the 16 significant SNPs located on chromosome 1H, 2H, 5H and 7H based on the genome physical map of Morex v1 122

Appendix 4. Physical map of the 16 significant SNPs located on chromosome 1H, 2H, 5H and 7H based on the genome physical map of Morex v3 123

Appendix 5. Candidate gene IDs in Morex v1 and Morex v2, and the location of Golden Promise v1 contigs overlapping with the candidate genes in Morex v2 124

Appendix 6. Sequence identity of the coding regions of the candidate genes with their homologs in other plant species 125

ABBREVIATIONS 127

LIST OF TABLES

Table 3.1 Primer sequences used for expression analysis of the candidate genes 47

Table 3.2 Germination index and its skewness and heritability in the mapping panel 49

Table 3.4 Correlation coefficients of the germination index of the association mapping panel between the trial environments 51

Table 3.5 Significant SNPs associated with PHS using the dataset pooled from all trial envirnoments based on FDR derived *p* value threshold 60

Table 3.6 Putative candidate genes co-localized with significant markers associated with PHS.. 65

LIST OF FIGURES

Figure 2.1 Steps involved in genome wide association studies in plants 27

Figure 3.1 Frequency distribution of variation in germination index (GI) of the mapping panel evaluated over the four trail environments 50

Figure 3.2 Genome wide linkage disequilibrium (LD) decay plot of intra-chromosomal marker pairs as a function of genetic distance (cM) for the association mapping panel 52

Figure 3.3 Estimation of number of subpopulations.. 54

Figure 3.4 Quantile-quantile plots showing the observed -log10 *P* values against the cumulative -log10 (*P* values) derived from marker-trait association analysis of the dataset pooled from all trial environments.. 56

Figure 3.5 Manhattan plot derived from genome wide association analysis of the dataset pooled from all trial environments... 58

Figure 3.6 Germination index of non-dormant (Himalaya) and dormant (Steptoe) genotypes over a period of seven days.. 62

Figure 3.7 Relative transcript levels of *HvFIE* (A), *HvPM19* (B), *HvGT* (C) and *HvGlyRS (D)* in the embryo of non-dormant (Himalaya) and dormant (Steptoe) seeds.. 63

Figure 3.8 Relative transcript levels of *HvLDL1* (A), *HvMTHD* (B), *HvPRORP1* (C), *HvHUP6* (D), *HvCRLK1* (E), *HvPRX16* (F) *HvTOP1* (G) and *HvPUB8* (H) in the embryo of non-dormant (Himalaya) and dormant (Steptoe) seeds... 64

ABSTRACT

Kaur, Gurkamal. M.Sc. The University of Manitoba, November 2021. Genomic analysis of pre-harvest sprouting resistance in barley. M.Sc. Major Supervisor: Dr. Belay T. Ayele, and co-supervisor: Dr. Ana Badea

Pre-harvest sprouting (PHS), which refers to the germination of seeds before harvest due to humid conditions, is one of the major factors that negatively affects the production of barley. The incidence of PHS is primarily associated with the level of seed dormancy, which refers to the failure of seeds to germinate under favourable conditions. The lack of sufficient level of dormancy leads to PHS whereas high dormancy level is responsible for non-uniform germination and delayed seedling establishment. It is therefore critical to have intermediate level of dormancy. Since seed dormancy and PHS are quantitatively inherited traits, identification of genomic regions/genes associated with dormancy has a paramount significance in breeding barley cultivars that are resistant to PHS. This study carried out genome wide association analysis to detect dormancy/PHS associated loci by using an association mapping panel consisting of diverse barley genotypes grown over multiple environments. The phenotypic data of the mapping panel was recorded as germination index (GI) while their genotypic data was derived using 50k Illumina Infinium iSelect genotyping single nucleotide polymorphism (SNP) array. The findings of the project showed association between the genotypic and phenotypic characteristics of the different barley genotypes exhibiting considerable variation in dormancy, leading to the detection of significant markers and closely linked genomic regions/genes associated with dormancy and PHS resistance. Furthermore, the expression patterns of the candidate genes were studied in two genotypes possessing contrasting dormancy levels, and most of the candidate genes showed differential expression between the two genotypes, indicating their role in dormancy regulation.

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

Barley (*Hordeum vulgare* L.) is one of the most widely grown cereal crops in the world. It is considered as a versatile crop since it can be grown over a wide range of environments including those with high altitudes and latitudes (Newton et al. 2011). Barley is extensively used for animal feed, malt products and human consumption (Baik et al. 2008). However, the yield and quality of barley is adversely affected by a number of factors including PHS, which is defined as the germination of seeds while they are attached to the parent plant due to moist conditions that occurs before harvesting. The occurrence of PHS enhances α-amylase activity which in turn degrades starch and downgrades the seed quality and value of barley grain (Nonogaki et al. 2018). In addition, pre-germinated seeds tend to absorb more water during steeping which can cause mold problems and affects its malting property. Therefore, this problem is of great concern for growers, maltsters and brewers as it causes huge economic losses (Gubler et al. 2005; Rodríguez et al, 2015).

The phenomenon of PHS is significantly correlated with seed dormancy, an adaptive trait that supresses the germination/sprouting of viable seeds under conditions that normally favour germination (Gao and Ayele 2014). Therefore, seeds with low level of dormancy at harvest stage in barley and other cereal crops show high susceptibility to PHS (Li et al. 2004). Dormancy and PHS traits are influenced by a number of endogenous and exogenous factors (Anderson *et al*., 1993). It is well known that seed dormancy and germination are predominantly regulated by phytohormones especially gibberellin (GA) and abscisic acid (ABA). The plant hormone ABA regulates seed dormancy positively and thus it inhibits germination whereas GA regulates dormancy negatively and thus it stimulates dormancy release and germination (Finkelstein et al. 2008).

Seed dormancy/PHS is governed by multiple genetic factors that are also referred to as quantitative trait loci (QTL). Therefore, genetic analysis of seed dormancy/PHS of barley is highly important to identify QTLs or genes that control dormancy levels of seeds in such a way that they become resistant to PHS and at the same time be able to germinate uniformly and quickly after harvest. The development of DNA-based markers has significantly increased the efficiency of selecting PHS resistant genotypes due to the fact that the markers are not influenced by the environmental factors (Singh et al. 2013). Therefore, identifying genetic markers associated with dormancy QTLs is of great importance in marker-assisted breeding. Genome wide association study (GWAS), which is based on linkage disequilibrium (LD), has emerged as a powerful approach to detect QTLs/genes as well as makers associated with the trait of interest (Alqudah et al. 2020). As compared to traditional linkage mapping, which is based on biparental populations, GWAS provides higher mapping resolution as it scans a wide range of germplasm collections, and thereby takes an advantage of ancient recombination events and uses high density markers to cover the whole genome (Korte and Farlow 2013). In cereal crops such as wheat and barley, studies that employed genome wide association analysis have successfully revealed genomic regions underlying plant response to a number of abiotic stress factors such as drought (Abou-Elwafa et al. 2021; Pham et al. 2019), salinity (Alotaibi et al. 2021; Mwando et al. 2020) and waterlogging (Borrego-Benjumea et al. 2021). GWAS has also been shown to be an effective platform in dissecting the genetic bases of dormancy/PHS in cereals including wheat (Kulwal et al. 2012; Martinez et al. 2018), rice (Shi et al.2021; Sohn et al. 2021) and barley (Nagel et al. 2019; Rooney et al. 2021). In barley, the major dormancy QTLs, SD1 and SD2, identified by using biparental mapping populations has also been verified by a recent GWAS of germination traits in malting barley (Rooney et al. 2021). Previous studies have also identified casual genes for dormancy QTLs

in barley. *Alanine aminotransferase* (*AlaAT*) and *Mitogen-activated Protein Kinase Kinase 3* (*MKK3*) genes have been identified as causal genes for the SD1 and SD2 seed dormancy major QTLs of barley and these genes could be used to develop markers to aid selection for PHS resistance in barley breeding programs (Nakamura 2018). Given that GWAS is an effective method for genetic analysis of quantitative traits and detection of important genomic regions this thesis project performed genome wide association analysis of seed dormancy/PHS using an association mapping panel of 255 diverse barley genotypes. The objectives of the project were to:

1. Examine the genotypic and phenotypic characteristics of different barley genotypes for PHS
2. Identify and validate genomic regions responsible for PHS resistance by associating phenotypic and genotypic characteristics
3. Identify causal candidate genes of the genomic regions controlling PHS resistance

2.0 LITERATURE REVIEW

2.1 Barley: a major cereal crop

Barley (*Hordeum vulgare* L.) is one of the earliest cultivated cereal crops, and based on acreage, it is the fourth important cereal crop in the world (FAO 2020). It is a self-pollinating cereal crop and belongs to the plant group Triticeae, a tribe in the Poaceae family (Harwood 2019). It is highly adaptable to diverse environmental conditions and cultivated extensively in different climatic regions of the world, starting from sub-arctic to sub-tropical regions (Zohary and Hopf 2000). Presently, barley is cultivated at high altitudes in the Himalayas, seasonal flooded areas of Southeast Asia, arid Mediterranean regions, as well as in fertile and marginal areas under harsh conditions (Angessa and Li 2016; Von Bothmer et al. 2003).

Barley can be classified as two-row or six-row type based on the spike morphology and fertility (Koppolu et al. 2013). Both types of barley produce three spikelets at each node, however, only the central spikelet is fertile in two-row barley while all the spikelets are fertile in the six-row barley (Ayoub et al. 2002). Barley can also be morphologically categorized as hulled/covered or hulless based on the presence or absence of hull adherence to the grain (Shaveta and Kaur 2019). Furthermore, based on the requirement of cool temperature for vernalization, barley can be classified as winter and spring barley. The winter barley needs exposure to low temperature to be able to produce heads and grains and is often planted in the fall so that it can be exposed to the cool temperatures during winter. On the other hand, spring barley needs no vernalization and is sown in the spring (Kling et al. 2004). In addition to its importance as an agricultural crop, barley has been used as a model plant for studying the Triticeae family because of its true diploidy and availability of high number of genetic stocks (Harwood 2019; Mascher et al. 2017).

2.2 Origin and domestication of barley

Barely was amongst the first crops grown in the old-world agriculture (Zohary and Hopf 2000). Archaeological evidence confirms that barley is domesticated from its two-row wild relative, *Hordeum spontaneum* C. Koch, about 10,000 years ago, in the fertile crescent area in the Middle East (Badr et al. 2000; Zohary et al. 2012). German botanist named Carl Koch, discovered *H. spontaneum* in Turkey (Bothmer et al. 1995). The traces of *H. spontaneum*, a diploid (2n=14) wild progenitor of barley, were also found in Morocco, Algeria, Libya, Egypt, Crete, Ethiopia, and Tibet, indicating multiple geographic origins of barley (Bothmer and Komatsuda 2011; Molina-Cano et al. 2002). The diploid barley *H. vulgare* was domesticated in Ain Ghazal (Rollefson et al. 1985; Willcox 1998) and later distributed from the Mediterranean basin to Africa and Europe, and eastwards via Iran and Afghanistan to India and China (Kilian et al. 2010). Furthermore, Ethiopia and Eritrea are cited as other barley domestication sites (Orabi et al. 2007), and another study reported that the Tibet Plateau and its environs as being one of the domestication sites of hulless barley (Dai et al. 2012). Several phenotypic alterations mainly due to genetic mutations occurred during barley domestication, and these alterations differentiate wild and domesticated barley types (Haas et al. 2019). For example, the formation of brittle rachis in wild barley is highly associated with two closely linked genes, *non-brittle rachis 1* (*btr1*) or *non-brittle rachis 2* (*btr2*). The presence of a recessive allele at either of the two loci of cultivated barley, resulted in non-brittle rachis and led to enhanced grain retention (Pourkheirandish et al. 2015). Furthermore, the presence of a recessive allele of another gene, namely *six-rowed 1* (*vrs1*), resulted in the conversion of two-row barley to six-row barley, leading to tripled spikelets that produce higher number of grains (Komatsuda et al. 2007).

2.3 Barley genome and genomic resources

Barley has a large genome size of approximately 5.3 Gbp, which includes more than 39,000 genes. It has a complex genome with 80.8% of repetitive elements and over 30,000 genes belonging to gene families with multiple members (Mascher et al. 2017). Nevertheless, there are extensive genomic resources available including a high-quality barley reference genome sequence which was made available in the year 2012 by the International Barley Sequencing Consortium (IBSC) (Mayer et al. 2012). Initially, the bacterial artificial chromosome (BAC) clones obtained from Morex (six-row barley) were sequenced using shotgun method and the IBSC V1 was created. Five years later, improved assembly of the barley reference genome (IBSC V2) was generated by Illumina sequencing of BACs along with optical mapping and chromosome conformation capture sequencing (Mascher et al. 2017). Furthermore, a project named barley pan-genomics had been initiated in order to cover the intraspecific diversity within genome content as well as structure in barley germplasm groups. This was performed by assembling high quality genome sequences of multiple accessions (Monat et al. 2019a). IBSC has published the whole genome shotgun sequence assemblies for domesticated two-row barley genotypes including Barke, Bowman, Haruna Nijo and Igri (Mayer et al. 2012) as well as one wild accession, namely B1K-04-12 (Hübner et al. 2009). Over the last few years, draft genome assemblies have been generated using high coverage Illumina sequencing for a number of barley genotypes including Haruna Nijo (Sato et al. 2016), Lasa Goumang (Zeng et al. 2015), Zangqing320 (Dai et al. 2018) and WB1 (Liu et al. 2020). For minimizing the cost associated with genome assembly, a pipeline named TRITEX has been developed (Mona et al. 2019b). Using TRITEX, an enhanced annotated reference genome sequence assembly has been constructed for Morex. The genome sequence

assemblies of barley cvs. Golden Promise (GP) (two-row) and AAC Synergy (two-row) have also been released recently using the TRITEX pipeline (Schreiber et al. 2020; Xu et al. 2021).

2.4 Barley production and use

Global barley production was reported to be 157.9 million metric tons in the 2019/2020 crop year. During the same crop year, the European Union was the main producer of barley, accounting for 55.3 million metric tonnes of the barley produced followed by Russia and Canada, which accounted for 19.9 and 10.4 million metric tons, respectively (USDA 2020). Barley production in Manitoba in the 2020 crop year was estimated to be 0.7 million tonnes, and approximately 42.7% of the total area seeded with barley was allotted for malt barley varieties while 51.7% was allocated for feed and forage barley varieties (Canadian Grain Commission 2020).

Barley was first domesticated for human consumption, but it was later relegated to brewing and animal feed due to the market demand for wheat and rice as primary food crops. Barley is mainly used as animal feed in the form of ground, pelleted, tempered, rolled or flaked barley and also fed to dairy cows as hay, silage or straw. In beef cattle diets, barley grain is the prime source of energy and protein (Ullrich 2010).

In general, approximately 70% of the barley produced is utilized for animal feed, 20-25% for malting and 5-10% for human consumption (Langridge 2018). Although barley is primarily grown for animal feed, the most valuable use of barley is for malting and brewing beer and making wine and strong liquors (Olafson 2018). For malting, barley outperforms other grains since it has high enzymatic content. During malting, these enzymes convert the starch to fermentable sugars including maltose and sucrose, leading to sweet and good taste (Olafson 2018). Barley is considered to be a rich source of fiber, vitamins E and B-complex and β-glucan, as such it is an

important human food (Van Hung 2016) and is consumed in the form of bakery products, cereals, porridge, soups, breads and sattu (roasted barley) (Baik and Ullrich 2008). A barley grain consists of 65 to 68% starch, 11 to 34% total dietary fibre, 10 to 17% protein, 4 to 9% β-glucan, 2 to 3% free lipids and 1.5 to 2.5% minerals (Fastnaught 2001; Gupta et al. 2010). Due to its high nutritional value, previous studies have indicated that consumption of barley can reduce the risks of major diseases such as cancer, heart attacks and also plays a role in enhancing immunity. Moreover, abundance of soluble fibre in barley helps in digestion and cholesterol reduction (Idehen et al. 2017; Minaiyan et al. 2014).

2.5 Factors affecting barley production

The growth and development, and productivity of barley to its maximum genetic potential is negatively affected by several abiotic and biotic stress factors, eventually leading to reduction in its production. The abiotic stresses such as extreme temperatures, drought, salinity, frost, nutrient toxicities or deficiencies and water logging are responsible for significant yield reductions in barley (Borrego-Benjumea et al. 2021; Busconi et al. 2001; Cattivelli et al. 2010; Mahmood 2011; Savin et al. 1997; Setter and Waters 2003). It has been reported that heat stress leads to a reduction in plant height, number of tillers per plant, and thousand kernel weight in barley (Abou-Elwafa and Amein 2016). In addition, the occurrence of drought stress during barley seed development reduces the rate of leaf photosynthetic activity and the duration of grain filling, and enhances plant maturation, resulting in significant losses in grain yield (Forster 2004; Masoud et al. 2005; Samarah 2005).

Biotic stress factors including plant pathogens such as nematodes, viruses, bacterial and fungal pathogens are responsible for causing diseases in barley, thereby affecting its production.

The fungal pathogens that cause leaf diseases such as net blotch, spot-blotch, leaf rust, scald, and powdery mildew (Walters et al. 2012) and the viral diseases such as barley stripe mosaic and yellow dwarf diseases can cause serious yield losses (Shah et al. 2012). Moreover, bacterial diseases of barley such as bacterial blight, which is widespread across the temperate as well as subtropical zones, cause infection on flag leaves that leads to grain weight reduction. Other diseases such as cereal cyst and root lesion nematodes are known to severely affect the roots of barley (Paulitz et al. 2011).

Barley grain quality is a very important factor for the malt industry, and grains that are not good for malting are usually sold for feed at a much lower price. One of the quality factors required for malting barley is the ability of the grains to undergo uniform, complete and rapid germination upon imbibition, that is the ability to exhibit 95% or higher germination within 3 days (Oberthur et al. 1995). To this effect, barley breeders have performed selections against seed dormancy, however, this has a negative consequence as it results in susceptibility to PHS. It is one of the recurring problems in the production of malt barley and causes significant yield and quality losses (Rodríguez et al. 2015).

2.6 Pre-harvest sprouting

Pre-harvest sprouting refers to the germination of grains of cereal crops such as wheat, rice, maize, barley, sorghum, and rye while still on the mother plant due to rainy and high humid conditions (Gualano et al. 2007; Patwa and Penning 2020; Rodríguez et al. 2015; Ullrich et al. 2009). PHS typically occurs when the capacity for germination is high, that is, between physiological maturity (PM) and harvest maturity (Bench and Arnold 2001). During and after germination, the growing embryo requires the mobilization of starch reserves via induction of α-amylase activity (Clarke et

al. 1984). α-amylase catalyzes the hydrolysis of α-1,4 -glycosidic bonds in saccharides and has an important role in determining the rate of germination (Autio et al. 2001; Masojć and Milczarski 2009). Given that the activity of α-amylase is closely associated with the rate of seed germination (Wu et al. 2002), a higher level of α-amylase activity has been used as an indicator of PHS damage in cereals (Schwarz et al. 2004). The activity of α-amylase activity can be determined via falling number and stirring number analysis, which are based on assessment of the starch pasting properties, using a falling number apparatus and Rapid Visco Analyzer (RVA), respectively (Lin et al. 2008). The RVA method is widely accepted by North America's malting barley industry while falling number is a method of choice by the wheat industry. However, the American Society of Brewing Chemists approved both methods as standard protocols to measure PHS damage in barley (Tordenmalm et al. 2004).

Starch degradation due to PHS results in deteriorated end use quality of the grain (Lin et al. 2008), and this in turn leads to lower price of the grain in the market. In barley, sprouting of the grain before harvest can impair the malting process since such grains are less likely to germinate during malting, resulting in high level of β-glucan that is undesirable for the brewing process (Carvalho and Beléia 2019; Kuntz and Bamforth 2007). Moreover, degradation of the starch stored endosperm due to PHS leads to a decrease in seed weight and loss of seed viability (Li et al. 2003). It has also been shown that sprouted grains consist of higher levels of oxidoreductases and hydrolases that cause losses of nutrients including starch, protein and oil, and spoilage of grains (Li et al. 2004). All these effects of PHS overall lead to yield and quality losses and therefore huge economic losses in barley production.

Pre-harvest sprouting is always considered as a serious global issue that negatively affects the production of barley and other cereals around the world including the USA, India, Canada,

China, Germany, Australia, and Japan (Liu et al. 2016; Mares 2019). Annual yield and quality losses of cereal crops due to PHS are estimated to be $1 billion (Black et al. 2006). To counteract these losses, it is highly important to develop barley cultivars that remain dormant enough from their physiological to harvest maturity but lose their dormancy immediately after harvesting to ensure quick and uniform germination during malting (Ullrich et al. 2009).

2.7 Seed germination

Seed germination begins with the absorption of water by mature dry seed (imbibition) and ends with the protrusion of radicle through the seed coat (Bewley and Black 1994). There are three phases of water uptake during seed germination. Phase I involves immediate absorption of water by the seed, phase II consists of gradual and constant rate of water uptake (plateau phase) and phase III involves further increase in uptake of water (Bewley 1997). During phase I, the rapid water uptake by the seeds leads to structural changes in their cell membranes and this causes instant outflow of solutes and small metabolites into the imbibition solution. This phase of water uptake is characterized by the activation of respiration and synthesis of protein from pre-existing mRNA. Repair of damages that occurred to cellular components such as the cell membrane, mitochondria and DNA due to desiccation at the later phases of their maturation is initiated during phase I of water uptake. Once the seeds reach at constant water content, the phase II of water uptake starts in which synthesis of protein takes place from newly synthesized mRNA and the production of new mitochondria begins. Simultaneously, hydrolysis of food reserves and synthesis of the enzymes that are involved in softening the seed covering layers and embryo growth occur (Bewley 1997). This leads to the induction of radicle emergence through the surrounding structures, which marks the completion of germination. Following this, the seed enters into the final phase III of water

uptake during which the seed uptakes more water. This phase involves cell division and expansion, and mobilization of stored reserves, all of which are required for radicle elongation and young seedling growth (Nonogaki et al. 2010).

2.8 Seed dormancy

Pre-harvest sprouting is closely related with the degree of seed dormancy, which is defined as the inability of a viable seed to germinate under ideal environmental conditions (Gao and Ayele 2014; Gubler et al. 2005; Hilhorst 1995). Therefore, it has been identified as a primary factor that controls resistance to PHS (Yang et al. 2011). Resistance to PHS of barley genotypes can be determined by a direct estimation of the level of seed dormancy through seed germination test and subsequent calculation of GI (Strand 1980). In the context to the GI as a weighted index, the early and late germinating seeds are given maximum and less weights, respectively. As a result, the fast-germinating lines can be differentiated from the slower ones via GI (Walker-Simmons 1987). It has been reported previously that GI is the most reliable parameter of germination since it considers the speed as well as the percentage of germination (Kader 2005). Moreover, it has been considered as the key methods used by plant breeders and physiologists to measure dormancy as PHS trait (Biddulph et al. 2008).

2.9 Classification of seed dormancy

Dormancy can be classified as primary or secondary dormancy based on its onset. Primary dormancy is induced during seed maturation while the secondary dormancy is induced in non-dormant mature seeds due to the exposure of conditions that are not favourable for their germination (Finkelstein et al. 2008). Based on the cause, primary seed dormancy is further divided

into two types, seed coat imposed and embryo dormancy. Tissues surrounding the embryo such as the endosperm and seed coat can cause dormancy by restricting the movement of water and gases into the embryo (Bewley 1997). Moreover, the seed coat and other tissues surrounding the embryo can serve as a storage for germination inhibitor compounds such as flavonoids (Debeaujon et al. 2007; Shirley 1998). In agreement, removal of the seed coat and exposure of the embryo to favourable conditions has been reported to lead to germination (Naredo et al. 1998). Susceptibility of seeds to PHS is also attributed to the arrangement of epidermal cells in the seed coat. Loose arrangement of epidermal cells in the seed coat results in PHS susceptibility as the seeds become highly permeable to water and gas while seeds that exhibit tight arrangement of the epidermal cells in the seed coat are resistant to PHS (Gao et al. 2013).

Embryo dormancy is intrinsic to the embryo and is caused due to immaturity and underdevelopment of embryo or due to metabolic blocks that inhibits embryo growth (Finch-Savage and Leubner-Metzger 2006). The presence or absence of embryo dormancy has primarily been attributed to the contents of ABA and GA and sensitivity of seeds to these phytohormones (Bewley 1997). ABA is required for induction and maintenance of seed dormancy while GA is needed to release seed dormancy and promotes germination (Finkelstein et al. 2008; Kucera et al. 2005). Embryo dormancy can be released by prolonged dry storage of seeds at room temperature, which is termed as after-ripening (Bewley 1997). The role of after-ripening in breaking seed dormancy have been explained by changes in the expression of metabolic and signaling genes of various phytohormones, mainly that of GA and ABA (Gubler et al. 2008; Liu et al. 2013). The duration of dry storage needed to break dormancy varies within species as well as within cultivars of a given species. For example, a study in barley reported that seeds of cv. Ashkelon lose dormancy after 28 days of after-ripening while the seeds of cv. Mehola lose dormancy and

germinate after 70 days of after-ripening (Vanhala and Stam 2006). Furthermore, seed dormancy can be released by other treatments, which include chilling or cold treatment (cold stratification), exogenous application of GA, and treatments with light, nitrate and nitric oxide, and smoke (Bethke et al. 2006; Bewley and Black 1994; Kucera et al. 2005).

2.10 Factors affecting seed dormancy and germination

2.10.1 Role of plant hormones

Abscisic acid and GA are two main plant hormones that play important but antagonistic roles in regulating seed dormancy and germination. ABA positively regulates induction and maintenance of seed dormancy whereas GA enhances germination. Therefore, the balance between these two hormones is responsible for dormancy maintenance as well as its release (Shu et al. 2016; Tuan et al. 2018). The presence of low levels of ABA have been reported during embryogenesis and the level increases during the maturation phase before showing a decrease during desiccation (Bewley 1997). Previous studies on ABA deficient mutants of Arabidopsis has demonstrated the role of this hormone in regulating storage reserve deposition and onset of primary dormancy (Holdsworth et al. 1999). For example, mutations in genes that encode ABA biosynthesis enzymes such as 9-*cis*-epoxycarotenoid dioxygenase (NCEDs) and zeaxanthin epoxidase (ZEP) result in a decrease in the level of dormancy (Fang and Chu 2008; Lefebvre et al. 2006). On the other hand, overexpression of these genes lead to the induction of strong dormancy (Lin et al. 2007). Furthermore, overexpression of the gene encoding the ABA catabolic enzyme, ABA-8′-hydroxylases (ABA8′OH) leads to the observation of low level of dormancy (Okamoto et al. 2006). This reflects that an increase in seed ABA level enhances dormancy whereas reduction in

seed ABA level caused a decrease in the level of seed dormancy (Finkelstein et al. 2002; Kushiro et al. 2004).

Gibberellin also plays an important role in regulating seed dormancy as it breaks dormancy and therefore enhance germination (Finch-savage and Leubner-Metzger 2006). The levels of GA in seeds are regulated mainly by two GA biosynthesis enzymes, namely GA 20-oxidase (GA20ox) and GA 3-oxidase (GA3ox), and the GA catabolism enzyme GA 2-oxidase (GA2ox). It has been shown previously that an increase in the expression levels of *GA20ox1* and *GA3ox2* in the wheat embryos is associated with increased production of bioactive GA and enhanced seed germination (Izydorczyk et al. 2018). A study in Arabidopsis seeds has also found that *GA3ox1* and*GA3ox2* are highly expressed during seed germination (Hu et al. 2008). On the other hand, genes encoding the GA catabolism enzyme of Arabidopsis, namely *GA2ox1* and *GA2ox2*, are reported to have roles in reducing the level of bioactive GA in seeds and therefore inhibiting germination (Thomas et al. 1999). Moreover, GA regulates the expression of α-amylase genes and activity of the corresponding enzyme which is involved in starch breakdown during seed germination in cereals such as rice (Kaneko et al. 2002). For example, upregulation of *GA3ox2* during imbibition of barley seeds was found to be associated with enhanced expression of a gene encoding α-amylase (*Amy1*), suggesting the role of *GA3ox2* in influencing GA production and therefore α-amylase activity (Suzuki et al. 2005). Overall, the balance between ABA and GA plays important role in the regulation of dormancy in a seed (Izydorczyk et al. 2018; Shu et al. 2016). In addition to GA and ABA, studies in several plant species have reported that other hormones such as, ethylene, auxins, jasmonates and brassinosteroids play roles in controlling seed dormancy and germination (Kucera et al. 2005; Linkies and Leubner-Metzger 2012; Liu et al. 2013).

2.10.2 Role of environmental factors

Seed dormancy and germination are influenced by the environment experienced by the mother plant during seed development (Fenner 1991; Wulff 1995). The level of dormancy is negatively correlated with a number of environmental factors that occur during seed development including high temperatures, short days and drought (Fenner 1991). The degree of dormancy in barley seeds is known to be affected by temperature only within a specific period of seed filling during which dormancy as a trait is sensitive to temperature (Rodríguez et al. 2001). It has been reported previously that exposure to a low temperature during the first half of the seed filling period followed by exposure to a high temperature during the second half of seed filling period leads to reduction of seed dormancy in barley (Reiner and Loch 1976). Another factor responsible for influencing the level of seed dormancy during seed development is water stress (drought). A previous study in barley has demonstrated that water deficits during grain filling leads to lower dormancy levels (Gualano and Benech-Arnold 2009). However, another report showed that dormancy level in barley seeds increases due to the occurrence of water stress during the grain filling period (Aspinall 1965). Therefore, further studies are required to validate the effects imposed by drought on seed dormancy during seed maturation.

Environmental factors such as temperature, light and water supply during seed incubation/imbibition can have effect on seed dormancy and germination (Finch-Savage and Footitt 2017). It has been observed that imbibition of barley seeds at lower temperatures (5-20°C) results in reduction in the expression of dormancy while its expression increases with rising temperatures (above 25°C) (Corbineau and Côme 1996; Leymarie et al. 2008). Temperature effects can also be attributed to changes in oxygen availability in the embryo since oxygen solubility reduces at high temperatures (Hoang et al. 2013). Light is another environmental factor

that plays a role in controlling seed dormancy and germination during imbibition. While seed germination in cereals such as barley is promoted by darkness during their imbibition (Hoang et al. 2014), pervious studies have shown that the germination of seeds of these crops is repressed by white or blue light (Gubler et al. 2008; Hoang et al. 2014). Furthermore, other environmental factors such as water supply during imbibition influence the level of seed dormancy and germination (Copeland and McDonald 1999).

2.11 Genetics of dormancy and PHS

Dissection of the genetic architecture of dormancy and PHS is challenging due to the fact that they are quantitative and complex traits controlled by multiple genes (Li et al. 2004). Intergenic interactions along with interactions between environment and genetics have significant effect on dormancy and PHS. Therefore, QTL analysis has been used as a method of choice to study seed dormancy and PHS since these traits exhibit high variability in natural populations, and germplasm collections and accessions in many plant species (Gubler et al. 2005).

2.11.1 Use of molecular markers

The discovery of molecular markers in the 1980s has opened up new avenues for QTL analysis in crop species (Collard et al. 2005). Genetic mapping of barley with the use of molecular markers was first performed using restriction fragment length polymorphism (RFLP) markers (Melzer et al 1988). The RFLPs have been used in genetic mapping studies of several agronomic and malting quality traits as well as disease resistance (Hayes et al. 2003; Wagner et al 2008). Afterwards, barley genetic maps were developed using polymerase chain reaction (PCR) based molecular markers including randomly amplified polymerase DNA (RAPD) markers (Giese et al. 1994),

microsatellites or simple sequence repeats (SSRs) (Varshney et al. 2007) and amplified fragment length polymorphism (AFLP) markers (Qi et al. 1998). Subsequently, generating high-density linkage maps was made possible through the use of diversity array technology (DArT) (Wenzl et al. 2004) and expressed sequence tag (ESTs) markers (Stein et al. 2007). Recently, SNP markers have been used widely for genetic/genomic studies of population because of their reliability, abundance and low-cost (Jehan and Lakhanpaul 2006). Due to the development of new sequencing technologies such as next-generation sequencing (NGS) for genotyping over the past decade, analysis of SNP markers and precise mapping of genes have become more efficient (Davey et al. 2011).

The development of Illumina GoldenGate assay with 1,536 SNP markers, has for the first time allowed the use of high throughput genotyping in barley (Close et al. 2009). Subsequently, a 9k iSelect Illumina Infinium array containing 7,842 SNPs was developed (Comadran et al. 2012). Recently, a new 50k Illumina Infinium iSelect genotyping array containing 44,040 SNPs has been developed (Bayer et al. 2017). This was made possible due to the availability of high-resolution barley genome assembly (Mascher et al. 2017), which provided detailed gene annotation and thereby led to accurate placement and interpretation of markers in the context of linked genes.

2.11.2 QTL analysis

Detection of QTL plays an important role in the identification of genomic regions/genes associated with economically significant quantitative or polygenic traits controlled by multiple genes (Collard et al. 2005). The phenotypic variation and genetic architecture of such complex traits can be determined by identifying allelic variation at particular loci. Through the use of mapping tools, the underlying genomic regions influencing phenotypic variations in the traits of interest can be

identified (Alqudah et al. 2020). There are two widely used approaches for QTL mapping, namely linkage mapping and association mapping. Both techniques utilize the concept of recombination's capability of breaking up the genome into DNA fragments that can be linked to variation in phenotype (Myles et al. 2009). Linkage and association mapping have different degrees of influence over recombination. In linkage mapping, the experiment is generally carried out using biparental populations such as recombinant inbred lines (RILs), near isogenic lines (NILs), double haploid (DH), and F2 and backcross mapping populations with known relatedness. However, such populations are characterized by the very limited genetic variation displayed by the parental lines and limited recombination events (Xu et al. 2017). On the other hand, association mapping is an approach that identifies QTL regions using diverse germplasm or natural population exhibiting phenotypic diversity for particular traits (Alqudah et al. 2020). This approach examines large genetic variability and exploits the recombination events from several generations such that each generation provides information regarding the recombination between QTL and a marker or between two markers (Gupta et al. 2019).

2.11.3 Linkage mapping

Linkage mapping also referred as QTL mapping has been shown to be a powerful method for identifying loci that co-segregate with target traits in biparental populations (Alqudah et al. 2020). A suitable mapping population is often developed by selecting two parents showing contrasting phenotype for the trait of interest. Subsequently, the mapping population is genotyped using relevant molecular markers (Dhingani et al. 2015). For example, QTLs associated with physiological, biochemical and yield traits in barley RIL population have been identified using SSR markers (Gudys et al. 2018; Mikolajczak et al. 2016). Afterwards, linkage maps are

constructed using coding information of genetic markers for each individual/unit of the population. Linkage among markers is generally determined by using logarithm of odds (LOD) scores, which is estimated by calculating the ratio of linkage versus no linkage (Risch 1992). LOD score of >3 is typically used for constructing linkage maps. If the LOD value between two markers is found to be 3 , then it shows that the odds ratio is 1000:1, which means probability of linkage is 1000 times more as compared to no linkage (Jamil et al. 2016).

Linkage analysis between large number of molecular markers is conducted using powerful programs such as Mapmaker/EXP (Lander et al. 1987), JoinMap (Stam 1993), MapManager (Manly et al. 2001), MultiPoint (Mester et al. 2003), RECORD (Van Os et al. 2005), MSTmap (Wu et al. 2008) and MapDisto (Lorieux 2012). There are several methods that are used for QTL mapping including single marker analysis (Soller et al. 1976), simple interval mapping (Lander and Botstein 1989) and composite interval mapping (Zeng 1994).

A large number of linkage studies have been conducted in many plant species to identify QTLs associated with agronomically important traits such as flowering time in Arabidopsis (Werner et al. 2005), frost tolerance in wheat (Vagujfalvi et al. 2003) and photoperiod response in barley (Turner et al. 2005). With respect to seed dormancy and PHS, several QTLs have been detected in the major cereal crops, which includes wheat (Kulwal et al. 2005; Mori et al. 2005), barley (Hickey et al. 2012; Takeda and Hori 2007) and rice (Miznou et al. 2018). However, due to the occurrence of limited recombination events in the mapping populations, linkage mapping usually identifies QTLs within large chromosomal regions, 10 to 20 cM intervals (Holland 2007), and this causes difficulties in identifying candidate genes. As a result, only genes associated with a few of the detected QTLs have been cloned (Price et al. 2006). Moreover, linkage mapping

enables detection of only specific alleles as the information is obtained from biparental cross and therefore results in lower resolution (Rakshit et al. 2012).

2.11.4 Association mapping and LD

Association mapping often termed as LD mapping identifies correlation between phenotype and genotype in unrelated individuals by detecting polymorphisms associated to functional alleles (Myles et al. 2009). Association mapping is often categorized as candidate gene association mapping and genome wide association mapping. The candidate gene association mapping is used to identify polymorphisms in chosen putative candidate genes responsible for the phenotypic variation of target traits whereas GWAS involves the entire genome scan for identifying genetic variation associated with different complex traits (Myles et al. 2009).

In association mapping, the closeness of marker-trait associations is identified using LD between the genes/QTLs and markers across a diverse set of germplasm (Thornsberry and Buckler 2003). Thus, it is a type of association analysis that looks for genetic variations throughout the whole genome by relying on the extent of LD between the marker and the functional variant (Myles et al. 2009). LD is described as the non-random association of alleles at different loci (Flint-Garcia et al. 2003). The concept of LD mainly relies on how the linkage disequilibrium is affected by the genetic or physical distance between loci as the closeness between markers reflects strong LD (Thornsberry and Buckler 2003; Zhu et al. 2008). Therefore, association mapping is aimed at identifying markers that are tightly linked to the desired genes and can explore all the recombination events and mutations unlike conventional mapping approaches such as linkage mapping (Yu and Buckler 2006). The accuracy of detecting a QTL depends on how quickly the LD decays over distance, therefore, using a wide variety of germplasm covering all the historical

recombination events insures high precision of mapping (Rafalski 2010). LD is influenced by both genetic and phenotypic factors such as population size, demographic histories, admixture, mating system, recombination rate as well as selection effects (Gaut and Long 2003; Mohammadi et al. 2015; Nordborg et al. 2002).

The number and density of markers required to perform association analysis vary among different species as it depends on the level of LD. In self pollinating species, LD is higher as compared to cross pollinating species, thus a smaller number of markers is required for the genome coverage (Alqudah et al. 2020; Varshney and Tuberosa 2007). Several studies have indicated that the number of markers required for association analysis varies with the genetic diversity and size of the genome. For example, it has been stated that 140,000 markers are required to cover the genome of Arabidopsis whereas in maize (outcrossing species) around 10 to 15 million markers are necessary for a sufficient genome coverage (Kim et al. 2007; Myles et al. 2009). With respect to barley, which is a self-pollinated crop, if the LD decays at a distance of 5 kb, 1 million markers (usually SNPs) are required while if the decay occurs at 100 kb, only 57,000 markers are needed to cover entire barley genome (Semagn et al. 2010).

The probability of linkage between two loci is measured by either standardized disequilibrium coefficient (D′) (Herdrick 1987) or squared allele frequency correlation (r^2) (Weir 1996), which calculates the pairwise correlation between loci (Remington et al. 2001). The high r^2 value among the loci of the same chromosome (syntenic r^2) signifies high probability of markers to be inherited together (Alqudah et al. 2020). To estimate the distance of LD decay, scatter plots are plotted between syntenic r^2 and distance (bp/cM) among all pairs of alleles and LD decay curve is fitted in the scatter plot using Hill and Weir (1988) method or a smooth curve is fitted to the data using locally estimated scatterplot smoothing (LOESS) (Cleveland 1979). Furthermore, two

methods of selecting threshold r^2 have been examined to describe the relationship between LD decay and distance. Some studies demonstrated the use of critical r^2 value based on fixed arbitrary threshold values (Nordborg et al. 2002; Remington et al. 2001) while other studies calculate threshold r^2 from the 95th percentile of square root transformed r^2 values of unlinked markers (Breseghello and Sorrells 2006). Subsequently, LD decay distance is estimated as the point of intersection between the fitted curve and threshold r^2 value. The distance at which LD decays has a crucial role in determining whether a group of SNPs represents the same or different QTLs on a chromosome, as the SNPs lying within the LD decay distance has been considered as a part of single QTL (Abdel-Ghani et al. 2019).

2.11.5 Factors affecting GWAS

2.11.5.1 Population Structure

Genome wide association study involves the use of a collection of diverse genotypes of a given species, as a result, all accessions are not equivalently distantly related at genetic level as the individual genotypes could have similar geographical origins, growth habitat or history of breeding (Hamblin et al. 2010; Malysheva-Otto et al. 2006). This leads to the formation of subgroups of related individuals within a population. Thus, SNPs with high frequency in a given subgroup due to population structure or genetic relatedness may appear to show false associations with the trait in cases when the phenotype is highly correlated to the same subgroup (Flint-Garcia et al. 2005; Myles et al. 2009). For instance, spurious associations due to population structure have been detected in association analysis of flowering time in Arabidopsis (Aranzana et al. 2005). As a result, several approaches have been developed to correct genotype-phenotype covariance in GWAS. The first statistical approach for defining population structure has been implemented using

STRUCTURE program (Pritchard et al. 2000a). It estimates the correlation between the genotypes using unlinked markers, which is expected to be the outcome of population structure. The number of subpopulations is then inferred by the generation of Q-matrix and the individuals are assigned to their respective clusters (Pritchard et al. 2000a). The output from STRUCTURE is visualized in STRUCTURE HARVESTER program, which presents the most likely number of subpopulations by delta k against the number of successive subgroups (Earl 2012). The STRUCTURE program can accurately determine the population structure; however, it is a time-consuming process and requires advanced data analysis skills. Another approach to determine population structure has also been developed, designated as EIGENSTRAT approach, which is based on principal component analysis (PCA), and it is a relatively fast method of inferring population structure (Price et al. 2006). EIGENSTRAT is a dimension reduction method that uses genotypic data to determine genetic variation, which is explained by a minimal number of dimensions. The results from PCA includes the major principal components (PCs), PC1 and PC2, explaining maximum variation and are presented in a scatter plot (Alqudah et al. 2020). The approach used to correct false associations is a mixed model method, which generates a pairwise genetic relatedness matrix (kinship matrix; K) by estimating the degrees of relatedness among all pairs of individuals using random genetic markers (Thabet et al. 2018; Yu et al. 2006).

2.11.5.2 Heritability

Broad-sense heritability is an important factor to be considered in association studies since it is a useful measure of the contribution of genetic variation to the phenotype and association between genotype and phenotype (Zaitlen and Kraft 2012). Low heritability is responsible for reducing the capability of GWAS for detecting true associations. The heritability of a particular trait is

decreased by strong interactions between genotype and environment (G×E), which may occur during repeated field trials of genotypes across several sites and/or years. There are a number of ways of adjusting phenotypic data such as best linear unbiased predictors (BLUPs) and best linear unbiased estimators (BLUEs) that results in improved estimates for phenotypic values with respect to G×E interactions (Alqudah et al. 2020; Hess et al. 2006; Milner et al. 2019; Saade et al. 2017).

2.11.5.3 Sample size

In order to attain sufficient statistical power during association studies, a suitable sample size is required since GWAS involves the assessment of thousands of SNPs (Klein 2007; Park et al. 2010). Several studies have reported that a sample size of at least 100 is required to carry out GWAS as the statistical power of association analysis decreases significantly when the sample size is less than 100 (Hayes 2013; Kumar et al 2012; Pfeiffer and Gail 2003).

2.11.6 Steps involved in GWAS

Genome wide association study procedure involves several steps, which begins with the selection of genetically diverse set of individuals from natural population forming association germplasm panel (Figure 2.1). The phenotype of the germplasm panel is then evaluated for a particular trait of interest. Error-free phenotyping is considered as an important factor for GWAS, and its efficiency is enhanced by performing replicated field trails over varied locations and years. The overall heritability of the data sets should be assessed via considering G×E interaction. Since the phenotypic data of plants is highly unbalanced, the mean (BLUEs or BLUPs) is estimated to get accurate measurements (Milner et al. 2019; Nagel et al. 2019).

The next step of GWAS includes genotyping of the sample using appropriate molecular markers and SNP markers are the most widely used markers. Genotyping of diverse samples produces a large number of SNPs throughout the genome, and this allows precise mapping of the variation in complex traits via exploiting recombination events. Genotype-by-sequencing is the most common approach employed in genotyping as it generates abundant SNPs that span the crop genome at low cost. Subsequently, LD is analysed, and population structure and kinship are estimated to control false associations. Finally, the phenotypic and genotypic results of the mapping panel are analysed for their relatedness through the best fitting model using an appropriate software (e.g., Trait Analysis by Association, Evolution and Linkage; TASSEL). Significant markers are then selected based on correction methods such as false discovery rate (FDR) and bonferroni correction (BC), leading to the detection of markers and QTL/genes that are associated with the trait of interest (Alqudah et al. 2020). Over the past decade, several GWAS have been performed in barley to identify QTLs/genes closely associated with complex traits such as drought resistance (Jabbari et al. 2018), seed dormancy (Nagel et al. 2019) and salinity and cold tolerance (Mwando et al. 2020; Visioni et al. 2013).

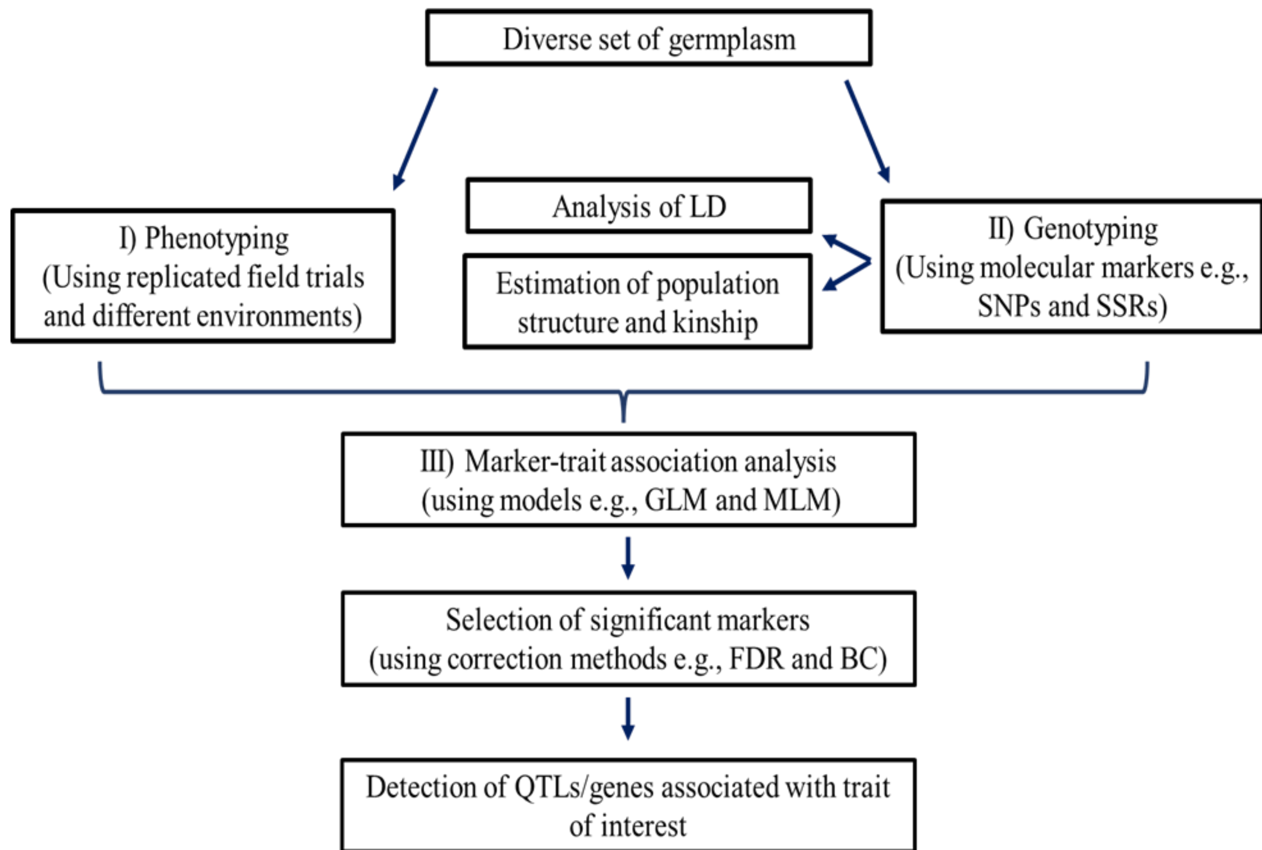


Figure 2.1 Steps involved in genome wide association studies in plants, namely phenotyping (stage I), genotyping (stage II) and genome wide association analysis (stage III).

2.11.7 Approaches for association mapping

Several software/tools have been developed in order to conduct GWAS such as TASSEL (Al-Abdallat et al. 2017; Bradbury et al. 2007), PLINK (Purcell et al. 2007), R (GAPIT) (Lipka et al. 2012), JMP Genomics (SAS Institute 2013) and GenStat (Alqudah et al. 2014; Payne 2009). For association analysis in plants, one of the most extensively used software is TASSEL (Matthies et al 2012; Pasam et al. 2012), which is applied to investigate marker-trait associations for specific traits through employing general linear model (GLM) and mixed linear model (MLM) (Bradbury et al. 2007). These models have been implemented to control false positives by considering the population structure and familial relatedness within a given population as covariates (Yu et al. 2006). The GLM model calculates the correlation between the genotypic and phenotypic data using Q-matrix or PCA as a covariate. Whereas MLM includes the use of both Q and K matrix or PCA and K matrix as covariates (Bradbury et al. 2007). Previous studies have reported that MLM outperforms GLM by reducing type I error (false positives) as well as type II error (false negatives). K matrix in the MLM corrects for relatedness among each possible pair of individual while Q and PCA covers relatively limited axis of variations (Yu et al. 2006). The latest version of TASSEL (v5) includes more features such as displaying the association results in the form of manhattan plots and quantile-quantile (QQ) plots, generation of phylogenetic tree by means of archaeopteryx tree viewer, and visualisation of LD and PCA through scatter plots (Alqudah et al. 2020). TASSEL (v5) involves the use of SNP data, and it is easy to access and freely available at http://www.maizegenetics.net/tassel.

2.11.8 Statistical power in GWAS

Significant marker-trait associations are selected on the basis of *p* value threshold above which the null hypothesis of no association is rejected. The significance of markers is generally assessed by BC (Weisstein 2004) or FDR (Storey and Tibshirani 2003). The BC method has been used intensively to determine the significant marker threshold (Maurer et al. 2016; Saade et al. 2017) and is calculated by dividing 0.05 (α level) by the total number of SNPs used in the analyses. The resulting value is a fixed threshold for all traits and is generally considered as too strict (Spencer et al. 2009). While the FDR approach is less stringent but more efficient as the *p* value thresholds for each trait are calculated independently (Storey and Tibshirani 2003). The FDR level for testing significance can be set at 0.01 and 0.05 depending on the study (Alqudah et al. 2016; Nagel et al. 2019). Alternatively, some association studies have used the bottom 0.1 percentile distribution of the *p* values as a threshold level where the FDR was inherently high (Chan et al. 2010; Pasam et al. 2012).

2.11.9 Advantages and limitations of GWAS

Association mapping has many advantages over linkage mapping. It is a relatively faster approach than linkage mapping as there is no need of generating mapping populations. GWAS enables the detection of more alleles in the panel whereas QTL mapping could only detect two alleles. It is a highly precise mapping approach and has a greater resolution than linkage mapping (Yu and Buckler 2006; Zhu et al. 2008). However, it has a major limitation of low power to detect rare alleles even if their effects are substantial (Cockram and Mackay 2018). This is because the threshold of minor allele frequency (MAF) is typically set at < 0.05 in order to correct false positives (Cockram and Mackay 2018). Since, in some species, rare alleles represent a significant

proportion of natural diversity, linkage mapping can better detect the rare alleles and consequently the allele frequency can be artificially enhanced by generating relevant mapping population (Stich and Melchinger 2010). Moreover, the power of association mapping to identify QTLs is dependent on various factors such as number of molecular markers, population structure, sample size and phenotyping of population (Xu et al. 2017).

2.11.10 Other approaches of association mapping

Both linkage and association mapping approaches have their own advantages. Through combining the great statistical power of linkage mapping for detecting rare alleles and the higher mapping resolution and wider allele coverage of association mapping, a new approach called nested association mapping (NAM) has been introduced (Gireesh et al. 2021; Yu et al. 2008). Nested association mapping comprises of multiparent mapping populations, which are generated by crossing multiple diverse donor parents with a single parent, and this strategy exploits historic as well as recent recombination events (Yu et al. 2008). For instance, the first NAM population in barley, named Halle Exotic Barley 25' (HEB-25), was developed in 2015 (Maurer et al. 2015). The single parent Barke (*Hordeum vulgare* ssp. *vulgare, Hv*) was crossed with 25 diverse wild barley accessions [24 accessions of *H. vulgare* ssp. *Spontaneum* (*Hsp*) and one accession of *H. vulgare* ssp. *Agriocrithon* (*Hag*)]. The HEB-25 offers an exceptional biodiversity which is highly advantageous in dissecting the genetic basis of complex agronomic traits such as flowering time and drought stress in barley (Pham et al. 2019). Moreover, another mapping population, which has been aimed at increasing the allelic diversity, was developed by crossing genetically diverse multiple parents, termed as Multiparent advanced generation inter-cross (MAGIC) (Cavanagh et al. 2008). The first eight-parent MAGIC DH population was generated to study genetic control of

flowering time in barley via QTL mapping (Sannemann et al. 2015). The population was developed by intermating eight spring barley accessions (7 accessions of old German barley landraces and one accession of modern German barley) in an eight-way cross.

2.11.11 QTLs/genes for PHS

The QTLs related to barley seed dormancy have initially been reported in the early 90's by means of QTL mapping (Ullrich et al. 1993). Four key QTLs associated with seed dormancy were identified via linkage mapping using biparental mapping population derived by crossing Steptoe (dormant) and Morex (non-dormant) genotypes (Han et al. 1996). The QTLs were designated as SD1, SD2, SD3 and SD4 explaining 50%, 15%, 5% and 5% phenotypic variability, respectively. Out of these, SD1 and SD2, which are located near the centromeric and telomeric region of the long arm of chromosome 5H, respectively, have been consistently reported to have large effects on dormancy (Han et al. 1996; Sato et al. 2009; Ullrich et al. 1993). The SD3 and SD4 showed only minor effects and were mapped on chromosome 7H and 4H, respectively (Gao et al. 2003; Han et al. 1996).

QTLs related to dormancy have also been identified in another barley DH population, generated by crossing Triumph (European, two-row, dormant) and Morex (North American, six-row, non-dormant) genotypes (Prada et al. 2004). The QTLs identified from this population include a major QTL associated with seed dormancy establishment during seed development, which is located around the centromere of chromosome 5H (equivalent to SD1). Two more QTLs were also mapped around the telomeric region of chromosome 5H (equivalent to SD2) and near *vrsl* locus on chromosome 2H, both possessing a role in dormancy release by after ripening. Similarly, a GWAS in barley revealed the role of SD1 in dormancy maintenance and SD2 in

dormancy breakdown (Nagel et al. 2019). It has been hypothesised that the regions at the centromere and telomere of the long arm of chromosome 5H are conserved hotspots for barley seed dormancy, which is consistent with findings of previous studies that investigated other barley mapping populations (Han et al. 1996; Lohwasser et al. 2013; Oberthur et al. 1995; Prada et al. 2004; Romagosa et al. 1999; Sato et al. 2009). Although several genetic studies in barley have detected a number of QTLs for seed dormancy on all seven chromosomes other than SD1 and SD2, these QTLs are reported to have minor effects on dormancy as compared to SD1 and SD2 (Bonnardeaux et al. 2008; Gao et al. 2003, Han et al.1996; Li et al. 2004; Lohwasser et al. 2013; Oberthur et al. 1995; Prada et al. 2004; Ramagosa et al. 1999; Ullrich et al. 1993, 2009).

Identification of causal genes for known dormancy-related QTLs plays a pivotal role in marker-assisted selection for the development of PHS resistant cultivars. Since seed dormancy is a complex trait, only a few number of causal genes associated with dormancy QTL have been detected in some plant species including *Delay of germination1* (*DOG1*) in Arabidopsis, *MOTHER OF FT AND TFL1* (*MFT*) in wheat and *Seed dormancy 4* (*Sdr4*) in rice (Bentsink et al. 2006, Nakamura et al. 2011; Sugimoto et al. 2010). In barley, two genes, *AlaAT* (Sato et al. 2016) and *MKK3* (Nakamura et al. 2016), have been reported to be the causative genes for the barley seed dormancy QTLs SD1 and SD2, respectively (Nakamura 2018). Furthermore, the *MKK3* is reported to be a causal gene for wheat seed dormancy QTL *Phs1* (Torada et al. 2016).

It has been shown previously that the *AlaAT* gene plays an important role in the assimilation of nitrogen, carbon metabolism and protein synthesis (Duff et al. 2012). The *AlaAT* gene has recently been suggested to have a function in ABA signaling, however, the precise mechanism is unknown (Wei et al. 2019). The *MKK3* gene is an important part of mitogen activated protein kinase (MAPK) cascade mediating signal transduction pathways that influence a wide range of

plant growth and developmental processes. Previous reports have shown that *MKK3* participates in several signaling pathways including those involving ABA, jasmonates, blue light, pathogen resistance, osmotic stress, and wounding (Dóczi et al. 2007; Sethi et al. 2014; Takahashi et al. 2007, 2011; Zhang et al. 2012). It has been reported that *MKK3* is involved in the control of seed dormancy by negatively regulating ABA signaling (Nakamura et al. 2016; Nakamura 2018). Moreover, an important candidate gene that encodes a key GA biosynthesis enzyme and was suggested to have role in regulating PHS and germination rate in barley and rice, respectively (Abe et al. 2012; Li et al. 2004; Nagel et al. 2019), namely *Gibberellin 20 oxidase 1* (*GA20ox1*), has been mapped 1.7 Mbp proximal to *MKK3* in the SD2 region on chromosome 5HL (Sweeney et al. 2021).

Association mapping of various traits has been performed in several plant species including *Arabidopsis thaliana* (Aranzana et al. 2005), wheat (Arif et al. 2012), rice (Agrama et al. 2007) and barley (Nagel et al. 2019). Genome wide association mapping in barley has been demonstrated to be a highly efficient method to identify marker-trait associations for quantitative traits related to grain quality (Hassan et al. 2017), abiotic stress tolerance (Fan et al. 2016), morphological characteristics (Wang et al. 2012) and disease resistance (Gutiérrez et al. 2013; Turuspekov et al. 2016). GWAS provided several candidate loci associated with dormancy and PHS resistance in wheat (Albrecht et al. 2015; Jaiswal et al. 2012; Kulwal et al. 2012; Lin et al. 2016; Martinez et al. 2018; Zhou et al. 2017). A recent association analysis in barley, which used 5,156 SNPs markers to scan 184 spring barley genotypes, also identified new loci for seed dormancy (on chromosome 1H and 3H) and PHS (on chromosome 5H). In addition, the study identified specific genes including *GA20ox1*, *GA2ox4*, *lysyl oxidase-like2* (*LOXL2*), *cytochrome P450* (*CYP734A7*) and *DOG1L1* as candidate genes responsible for seed dormancy and PHS (Nagel et al. 2019). In

general, GWAS is a valuable tool for efficient detection of marker-trait associations and thereby localization of novel loci through studying germplasm lines exhibiting considerable variation in quantitative traits such as seed dormancy and PHS resistance.

3.0 GENOMIC ANALYSIS OF PRE-HARVEST SPROUTING RESISTANCE IN BARLEY

Abstract

Pre-harvest sprouting (PHS) causes significant reduction in yield and quality of barley. It is a quantitatively inherited trait closely related to the level of seed dormancy. Therefore, identification of dormancy associated genomic regions/genes plays an important role in enhancing PHS resistance. To this end, this study performed genome wide association study (GWAS) of a diverse mapping panel containing 255 barley genotypes over four environments. The mapping panel was phenotyped through monitoring germination index (GI) and genotyped using 50k Illumina Infinium iSelect genotyping array. The genotypes showed substantial variation in GI or dormancy levels, which is considered as PHS trait. Analysis of marker-trait association using MLM (Q+K) identified a total of 16 significant single nucleotide polymorphism (SNP) markers associated with dormancy/PHS, based on FDR ($\alpha = 0.05$) derived threshold of $-\log10(P) = 4.6$. Linkage disequilibrium (LD) decay analysis detected seven QTLs on chromosome 1H, 2H, 5H and 7H. The significant SNPs located in these QTLs explained 7.53% to 11.69% of the phenotypic variation and were contained within 13 candidate genes. Our analyses revealed that three SNPs explaining the highest phenotypic variations (>10%) are located within *HvFIE* and *HvLDL1* on chromosome 7H, and *HvPRORP1* on chromosome 2H, indicating their significance in controlling dormancy/PHS. Ten of the 13 candidate genes identified including the *HvFIE*, *HvLDL1* and *HvPRORP1* showed differential expression between dormant and non-dormant genotypes indicating their role in regulating dormancy/PHS. Overall, the study detected novel loci/genes and validated previously identified genomic regions associated with dormancy/PHS, which might play essential roles in enhancing PHS resistance in barley through marker-assisted breeding.

3.1 Introduction

Barley (*Hordeum vulgare* L.) is one of the most important cereal crops globally, and it is grown mainly for animal feed (~70%), malt products (20-25%) and human consumption (5-10%) (Langridge 2018). The phenomenon of pre-harvest sprouting (PHS), which is defined as the germination of physiologically matured seeds on the parent plant prior to harvest due to wet and humid weather conditions, is among the major factors that cause barley yield and quality losses. The incidence of PHS is closely related with the level of seed dormancy (Kulwal et al. 2012; Rodríguez et al. 2015), which refers to a trait that blocks germination even under ideal environmental conditions (Finch-Savage and Leubner-Metzger 2006; Gao and Ayele 2014). Seeds with low level of dormancy are more susceptible to PHS while high level of dormancy affects uniformity of germination and subsequent seedling establishment (Ullrich et al. 2009). Moreover, high level of seed dormancy increases storage costs since such seeds must be stored for a longer duration before it can be used for malting. Therefore, an appropriate level of dormancy is highly desirable for barley seeds. Several studies have shown that the level of seed dormancy is determined by the balance between two phytohormones, ABA and GA (reviewed in Tuan et al. 2018). Abscisic acid is a major regulator of seed dormancy as it induces and maintains seed dormancy. On the contrary, GA promotes dormancy release and seed germination (Finkelstein et al. 2018; Tuan et al. 2018).

Seed dormancy is a polygenic trait; therefore, QTL analysis is one of the most efficient approaches to elucidate genomic regions associated with seed dormancy, and the markers identified using this analysis can potentially be used in marker-assisted breeding for PHS resistance (Zhu et al. 2016). Previous studies have reported several QTLs controlling dormancy or PHS in a number of cereal crop species including sorghum (Cantoro et al. 2016; Carrari et al.

2003), wheat (Kulwal et al. 2012; Martinez et al. 2018; Zhu et al. 2019) and rice (Sugimoto et al. 2009). With respect to barley, QTLs for seed dormancy have been detected on all of the seven chromosomes (Bonnardeaux et al. 2008; Gao et al. 2003, Han et al. 1996; Li et al. 2004; Lohwasser et al. 2013; Oberthur et al. 1995; Prada et al. 2004; Ramagosa et al. 1999; Ullrich et al. 1993, 2009). However, only two major dormancy QTLs, designated as SD1 and SD2, have been reported to date. These two QTLs are detected near centromeric and telomeric region of chromosome 5H (Ullrich et al. 1993) and are found to be the two most consistent QTLs in several linkage mapping studies of biparental populations (Han et al. 1996; Oberthur et al. 1995; Ramagosa et al. 1999; Sato et al. 2009).

Qunatitaive trait loci mapping studies are generally carried out using biparental populations that are characterized by limited genetic variations exhibited by the parental lines and limited recombination events (Xu et al. 2017). Moreover, linkage mapping enables detection of only specific alleles, and it is characterized by low resolution of the QTL chromosomal location. Because of these limitations of linkage mapping, association mapping, which is based on LD, has become an approach of choice to identify QTL regions associated with polygenic traits of interest. This is because association mapping involves the use of using diverse germplasm exhibiting high degree of genetic diversity and historical recombination events, leading to enhanced ability to precisely locate genes and QTLs linked to the specific markers (Alqudah et al. 2020; Rafalski 2010). In addition, rapid advancement in the development of genotypic arrays, from SSR to high-density SNP arrays, has enabled GWAS to efficiently detect genomic regions/QTLs underlying quantitative traits (Zhu et al. 2019). Association mapping has been used for genetic analysis of a number of quantitative traits in barley including grain quality (Hassan et al. 2017), abiotic stress tolerance (Fan et al. 2016; Nagel et al. 2019), morphological characteristics (Wang et al. 2012)

and disease resistance (Gutiérrez et al. 2015; Turuspekov et al. 2016). GWAS has also been widely used to identify genetic loci responsible for controlling PHS and dormancy in wheat (Kulwal et al. 2012; Lin et al. 2016; Martinez et al. 2018; Zhu et al. 2019) and rice (Lu et al. 2018; Magwa et al. 2016; Shi et al. 2021). To date, only two GWAS of dormancy and PHS/germination in barley have been reported, indicating that the application of GWAS for studying the genetics/genomics of seed dormancy/PHS in barley is still very limited. One of these studies detected loci associated with dormancy and PHS on chromosome 1H and 3H, and chromosome 5H, respectively (Nagel et al. 2019). A recent GWAS of germination/PHS in malting barley has identified a novel locus associated with germination trait on chromosome 6H, in addition to identifying the previously reported SD1 and SD2 QTL regions (Rooney et al. 2021).

Over the past decade, two casual genes, designated as *AlaAT* and *MKK3*, underlying the SD1 and SD2 dormancy QTLs of barley, respectively, have been cloned (Nakamura et al. 2016; Sato et al. 2016). A study in wheat has also reported *MKK3* as a causal gene for major dormancy QTL, *Phs1* in wheat, demonstrating that this QTL is an ortholog of SD2 in barley (Torada et al. 2016). It is well known that the plant hormone ABA plays an important role in repressing seed germination (Nambara et al. 2010; Tuan et al. 2018). Both *AlaAT* and *MKK3* genes has been reported to have a role in ABA signaling (Nakamura et al. 2016; Wei et al. 2019). While the exact mechanism by which *AlaAT* affects ABA signaling is not yet known, *MKK3* has been reported to negatively regulate ABA signaling and thereby promoting seed dormancy release (Ishikawa et al. 2019; Nakamura et al. 2016). In addition, a gene encoding the GA biosynthesis enzyme *GA20-oxidase1* has been mapped 1.7 Mbp proximal to *MKK3* on chromosome 5HL (Sweeney et al. 2021), and previous studies in barley have demonstrated the role of *GA20-oxidase1* in PHS (Li et al. 2004; Nagel et al. 2019). This is consistent with the role of GA in enhancing seed dormancy

breakage and germination (Tuan et al. 2018). A recent GWAS in barley has also identified *GA2ox4*, *GA20ox1*, *LOXL2*, *CYP734A7* and *DOG1L1* as candidate genes associated with the dormancy and PHS trait (Nagel et al. 2019).

The aims of this study were to detect genomic regions responsible for dormancy and PHS resistance through genome-wide association mapping of a diverse barley germplasm collection grown over multiple environments and determine the expression patterns of candidate genes identified to be associated with seed dormancy.

3.2 Material and methods

3.2.1 Plant material and field trials

An association panel consisting of 255 diverse barley genotypes was used for this study. The genotypes comprising the panel have diverse geographical origins. The genotypes were grown in replicated field trials in two different locations over three growing seasons (2018, 2019 and 2020); at Agriculture and Agri-Food Canada, Brandon Research and Development Center, Brandon, Manitoba (MB) (2018, 2019 and 2020) and Agriculture and Agri-Food Canada, Morden Research and Development Center, Morden, MB (2019). The field trials were carried out using randomized complete block design (RCBD) with three replications. The grains were harvested at maturity when the peduncle of most the spikes became yellow.

Two genotypes with contrasting dormancy phenotype, namely Steptoe and Himalaya, were selected for studying the expression patterns of candidate genes. Plants of these genotypes were grown in a growth room at 22°C/18°C (day/night temperatures) with a 16/8-hour photoperiod. The grains were harvested at maturity and then stored at room temperature for approximately a month followed by a storage at -80˚C until further use.

3.2.2 Phenotyping

3.2.2.1 Evaluation of dormancy/PHS phenotype

The sprouting phenotype of seeds collected from each trail environment was determined by performing germination test, which was carried out by imbibing 50 seeds per plate (one plate per replication, a total of 3 replication) from each barley genotype in a Petri Plate system described previously (Gao et al. 2012) over a period of 15 days. To verify their dormancy level, germination test was also performed with the two genotypes (Himalaya and Steptoe) grown in a growth room but with 20 seeds per replication and over a period of 7 days. Seed germination, which is marked by the penetration of coleorhiza through the layers covering the seed, was scored daily. Germination index 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 stands for the number of seeds germinated on first (g1), second (g2), third (g3), …. and 15th (g15) day of imbibition, and n stands for the number of total seeds imbibed.

3.2.2.2 Statistical data analysis

Analysis of variance (ANOVA) and descriptive statistics were carried out using IBM SPSS (V.26) statistics software (https://www.ibm.com/analytics/spss-statistics-software). The GLM, with univariate model and type III sum of squares, was used to examine differences in GI among genotypes and environments and to determine the presence of interactions between genotype and environment. The GLM model is based on the following equation:

$$GLM_{yijkl} = \mu + G_i + E_j + Bk_{(j)} + GE_{ij} + eijkl$$

Where G_i is the effect of the ith genotype, E_j is the effect of the jth environment, and $Bk_{(j)}$ is the blocking effect. The genotype-environment interaction effect is denoted by GE_{ij}, and eijkl represents the random error (Imtiaz et al. 2008; Shao et al. 2018).

The best linear unbiased estimators, broad-sense heritability (H^2) and correlation between the environments were estimated by using the residual maximum likelihood (REML) method in multi-environment trail analysis with R version 6.0 (META-R) (Alvarado et al. 2020). The BLUEs were derived by considering genotype as fixed effect and environments as random effect. The resulting BLUEs were used as the mean values for the dormancy/sprouting phenotype in the association analysis. The H^2 by the environment was estimated using the following equation (Imtiaz et al. 2008; Shao et al. 2018):

$$H^2 = V_g / (V_g + V_{ge} / e + V_e / re)$$

Where V_g, genotype variance; V_{ge}, genotype-environment variance; V_e, error component's variance; r, number of replications; e, number of environments.

3.2.3 Genotyping

The 255 barley genotypes selected for this study were genotyped using 50k Illumina Infinium iSelect genotyping array containing 44,040 SNPs (Bayer et al. 2017). The markers were filtered by excluding SNPs with heterozygosity, MAF < 5% and missing values > 20%, resulting in a total of 31,899 SNPs that were used for the final analysis. The genetic positions of the SNPs were

determined using population sequencing (POPSEQ_2017) genetic map (Mascher et al. 2013) and the physical positions of the SNPs were obtained from Morex v1 (short-read BACs assembly) and Morex v3 (the latest long-read whole genome assembly) physical map (Mascher et al. 2017, 2021) using Barleymap tool (Barleymap (csic.es); Cantalapiedra et al. 2015).

3.2.4 Estimation of LD

The pairwise LD was estimated by calculating a squared allele frequency correlations coefficient (r^2) between intra-chromosomal SNP pairs for the 26,837 SNP markers using TASSEL v.5.2.72 software (Bradbury et al. 2007; https://tassel.bitbucket.io/). The LD decay in the whole genome over genetic distance (cM) was assessed by considering only significant ($p<0.001$) syntenic r^2 values. The non-linear regression was plotted against r^2 and distance between marker pairs using R v.4.1.1 software (R: The R Project for Statistical Computing; https://www.r-project.org/) and following the method described by Hill and Weir (1988). The critical r^2 value as a threshold was calculated by square root transformation of unlinked r^2 values (marker pairs that are >50 cM apart on the same chromosome or loci on different chromosomes). The 95th percentile of the resulting distribution was then considered as the critical r^2 value. The distance at which LD decayed was finally estimated by selecting the point of intersection between the fitted curve and threshold r^2 value.

3.2.5 Population Structure and kinship

A set of 189 SNP markers uniformly distributed over all seven barley chromosomes was used to characterize the population structure using the Bayesian cluster analysis in STRUCTURE v.2.3.4 software (Pritchard et al. 2000;

https://web.stanford.edu/group/pritchardlab/structure_software/release_versions/v2.3.4/html/structure.html).The approach used admixture ancestry model and correlated allele frequencies across population. The number of subpopulation (K) values ran from 1 to 7 with 25 iterations set for each K value. The burn-in time and Markov Chain Monte Carlo (MCMC) replication numbers were each set to 1×10^5 in order to generate the population structure (Q matrix). The optimal number of subpopulations was estimated by the delta K approach based on the rate of change in data's logarithmic likelihood between successive K values (Evanno et al. 2005) using STRUCTURE HARVESTER (Earl 2012; http://taylor0.biology.ucla.edu/structureHarvester/). The number of subpopulations was verified by performing PCA and phylogenetic tree-based analysis using TASSEL. In addition, familial relatedness was assessed by identity-by-state method (IBS) via generating kinship (K) matrix using TASSEL. The two PCs, Q and K matrices were used as covariates to avoid false positives in the linear models.

3.2.6 Association Analysis for dormancy/ PHS

Genome-wide association mapping was conducted using TASSEL v.5.2.72 for the identification of marker-trait associations between 31,899 SNPs and the phenotypic data (BLUEs). The GLM and the MLM were used, with the GLM model only considering fixed effects and the MLM model considering both fixed and random effects. Two GLM models, GLM_PCA and GLM_Q, where PCA and Q matrix are used as covariates, were run with 1000 permutations. Two MLM models along with K matrix as an additional covariate, namely MLM_PCA + K and MLM_Q + K, were run using optimum compression level in order to reduce the dimensionality of K matrix, and the variance components were computed using the population parameters previously determined (P3D) approach in TASSEL. Manhattan and QQ plots were obtained from each model and the

significant markers from the best model (MLM_Q+K) were considered for the analysis. The markers were considered statistically significant when the observed *P* value is less than the fixed *P* value threshold above which the null hypothesis of no marker-trait association is rejected. In this study, the cut-off *P* value was selected using FDR (α = 0.05) (Storey and Tibshirani 2003). Since the FDR derived threshold *P* value was found to be stringent to detect significant markers that are common among all the trial environments, marker significance was also determined using a less stringent method, which considers the lowest 0.1 percentile of the *P* value distribution as a cut-off *P* value (Chan et al. 2010). However, the FDR derived *P* value threshold was selected for further analysis as it narrowed down the overall number of significant markers in the dataset pooled from all trial environments.

3.2.7 QTL detection

The QTLs were designated according to their location on each chromosome and consisted of either single SNP or more than one SNP. Based on LD decay calculation, the significant SNPs lying within the window of LD decay distance on a particular chromosome were considered as a part of single QTL.

3.2.8 Identification of candidate genes

To determine if the significant SNPs are physically located within the candidate genes, the significant markers were anchored to the reference genome sequence of Morex v1 (short-read assembly of BACs) (Mascher et al. 2017) using Barleymap. To further verify the candidate genes associated with the significant markers, sequences of the markers derived from Germinate Barley SNP Platforms (hutton.ac.uk) were BLAST searched against the Barley Morex v2 [short-read

whole-genome shotgun assembly via TRITEX method (Monat et al. 2019)] using GrainGenes (GrainGenes | A Database for Triticeae and Avena (usda.gov)). Candidate gene sequences were also BLAST searched against the reference genome sequence assembly of cv. Golden Promise (GP) v1 using EnsemblPlants (https://plants.ensembl.org/Hordeum_vulgare/Tools/Blast?db=core).

Following this, the identification, IDs, annotations and positions (bp) of the candidate genes linked with the significant markers were determined from Barleymap based on the physical map of cv. Morex (Morex v1). Genes were annotated based on the names of the corresponding proteins in the protein annotation through evolutionary relationship (PANTHER) database of protein family/subfamily (http://www.pantherdb.org/) in *H. vulgare* or orthologs in Arabidopsis and other cereal crops. The coding nucleotide and the corresponding amino acid sequences of the putative candidate genes were obtained using the barley draft genome explorer (BARLEX) (https://apex.ipk-gatersleben.de/apex/f?p=284:10). Subsequently, nucleotide BLAST was used to identify homologs of the candidate genes in other species and determine their sequence identity with the respective candidate gene.

3.2.9 Expression analysis of candidate genes

3.2.9.1 Tissue collection and RNA extraction

For gene expression analysis, mature seeds of the two genotypes with contrasting dormancy phenotype were surface sterilized and imbibed in three replicates (25 seeds each) for 24 h following the method described previously (Gao et al. 2012). Embryos tissues were isolated from the imbibed seeds and frozen immediately in liquid nitrogen and then stored at -80˚C until RNA

extraction. Total RNA was isolated from the embryo tissues as described previously (Mukherjee et al. 2015). The RNA samples were stored at -80˚C until they are needed for further analysis.

3.2.9.2 DNAse digestion and cDNA synthesis

Elimination of genomic DNA contamination from the RNA samples was carried out via digestion of the RNA with DNase (DNA-*free* kit; Ambion, Austin, TX, USA) as described previously (Izydorczyk et al. 2018). The DNase digested RNA was used to synthesise complementary DNA (cDNA) using iScript cDNA Synthesis Kit (Bio-Rad, Hercules, CA, USA) as previously described (Nguyen et al. 2016). The resulting cDNA was then diluted 20X for qPCR analysis of the candidate genes.

3.2.9.3 Candidate gene primer sequences

Primer sequences specific to the candidate genes (Table 3.1) were designed using Primer 3 software (Primer3 Input, version 0.4.0). Specificity of the primers for each candidate gene was determined by BLAST searching the primers sequences against GenBank database (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and melting curve and gel electrophoresis.

Table 3.1 Primer sequences used for expression analysis of the candidate genes.

Name	Gene IDs	Type	Primer Sequence (5´ to 3´)
MTHD	HORVU1Hr1G082150	FP	AGCCAGTCCCTCAAGTTCG
		RP	CGTCGATGACCAGAGTCTCC
HUP6	HORVU2Hr1G084190	FP	AGAACTGCAGTGGGAGGAGA
		RP	GTGCCATATCCGGCAAATAC
PRX16	HORVU2Hr1G111610	FP	CGGATCTACACGTTCAAGCA
		RP	GAAGTAGGCGTTGTCGAAGG
PRORP1	HORVU2Hr1G127250	FP	TTCGTTGCTTGTTACGAACG
		RP	TAGCCTGACCTGGTGCTTCT
JMT	HORVU5Hr1G124360	FP	AGCAACCACCACCAGAAGTC
		RP	AAGAGCCGCTCATAAAACGA
PM19	HORVU5Hr1G125460	FP	CTGGTGCTGAACCTCATCAT
		RP	ATGGCGAAGACGAGGAAGTA
GlyRs	HORVU7Hr1G055340	FP	CTTGGAAAGGACGTGGTGAT
		RP	TCCTCCCTATGCCAAATGAG
TOP1	HORVU7Hr1G065200	FP	CCGTTATTCTGAGGGTGCAT
		RP	AGACCATTCGACCACGAAAC
LDL1	HORVU7Hr1G073020	FP	AGGCTGCTTTGGAGTTTGAA
		RP	AGCTGATTGAATGGGGTCAG
PUB8	HORVU7Hr1G073100	FP	GACCTGTCCAGAATCCCAGA
		RP	TCTTGCAGTTGTCGTCGTTC
CRLK1	HORVU7Hr1G073300	FP	CTAGAAACGGGGCCAACATA
		RP	CTGCCAGCAATCAGTTCAAA
GT	HORVU7Hr1G073370	FP	AAGAGGTCGCAATGTTACGG
		RP	AGTTTGGCAAGCATTTTTGG
FIE	HORVU7Hr1G073410	FP	CATTCGGGTCATCAACTGTG
		RP	TGCAGATCCCTGTATGGACA

FP: Forward Primer; RP: Reverse Primer

3.2.9.4 Real-time qPCR assay

The qPCR assays were conducted using CFX96 Real-Time PCR system and SsoFast EvaGreen Supermix (BioRad, Hercules, CA, USA). The qPCR reaction mixture consisted of 5 µl diluted cDNA, 1.25 µl forward primer (5 µM; final concentration 300 nM), 1.25 µl reverse primer (5 µM; final concentration 300 nM), 10 µl Eva Green Supermix and 2.6 µl sterile water, with a total reaction volume of 20 µl. The thermal cycling conditions were: initial denaturation and enzyme activation at 95°C for 5 min, followed by 40 cycles of denaturation at 95°C for 15 s, annealing at 60°C for 30 s, and extension at 72°C for 30 s. After normalisation with *β-Actin* as a reference gene, the transcript levels of the target genes were calculated using Livak and Schmittgen (2001) method.

3.2.9.5 Statistical analysis

Student's t-test at $P<0.05$ was used to determine statistically significant differences between samples.

3.3 Results

3.3.1 Characterization of variation in dormancy level and heritability

The level of dormancy in the association panel was evaluated as PHS trait by calculating GI over four environments. High degree of variation in GI scores was observed among the 255 genotypes in all trial environments, ranging from 0.15 to 0.97 (Brandon 2018; BRD18), 0.07 to 0.88 (Brandon 2019; BRD19), 0.06 to 0.90 (Brandon 2020; BRD20) and 0.02 to 0.94 (Morden 2019; MOR19) (Table 3.2). The lowest mean GI (0.42) of the association panel was recorded in BRD19, while the highest mean GI (0.69) was recorded in BRD18. The mean GI value pooled from all trial

environments was found to be 0.52. Based on the calculated GI values, genotypes were divided into three groups: highly dormant ranging from 0 to 0.40 GI, moderately dormant ranging from 0.41-0.70 GI and low dormant ranging from 0.71-1.00 GI.

Table 3.2 Germination index and its skewness and heritability in the mapping panel.

Environment[a]	Minimum GI	Maximum GI	Mean GI	Skewness	Heritability
BRD18	0.15	0.97	0.69	-0.61	0.89
BRD19	0.07	0.88	0.42	0.17	0.71
BRD20	0.06	0.90	0.53	-0.06	0.88
MOR19	0.02	0.94	0.45	0.23	0.93
Pooled	0.10	0.86	0.52	-0.09	0.87

[a]BRD18, Brandon 2018; BRD19, Brandon 2019; BRD20, Brandon 2020; MOR19, Morden 2019

The frequency distribution of the GI values pooled from all trial environments exhibited normal distribution with 53 genotypes categorized under the low dormancy group followed by 115 genotypes under moderate dormant group and 87 genotypes under high dormant group (Figure 3.1E). With respect to the individual environments, frequency distribution of the GI values was normally distributed in all environments (Figure 3.1B, C, D) except BRD18 (Figure 3.1A), which was moderately negatively skewed towards PHS susceptibility as 134 genotypes exhibited low dormancy level. The broad-sense heritability for PHS ranged from 0.71 to 0.93 in the individual environments and was estimated to be 0.87 based on the dataset pooled from all trial environments, indicating that the phenotypic variation was mostly genetically controlled (Table 3.2).

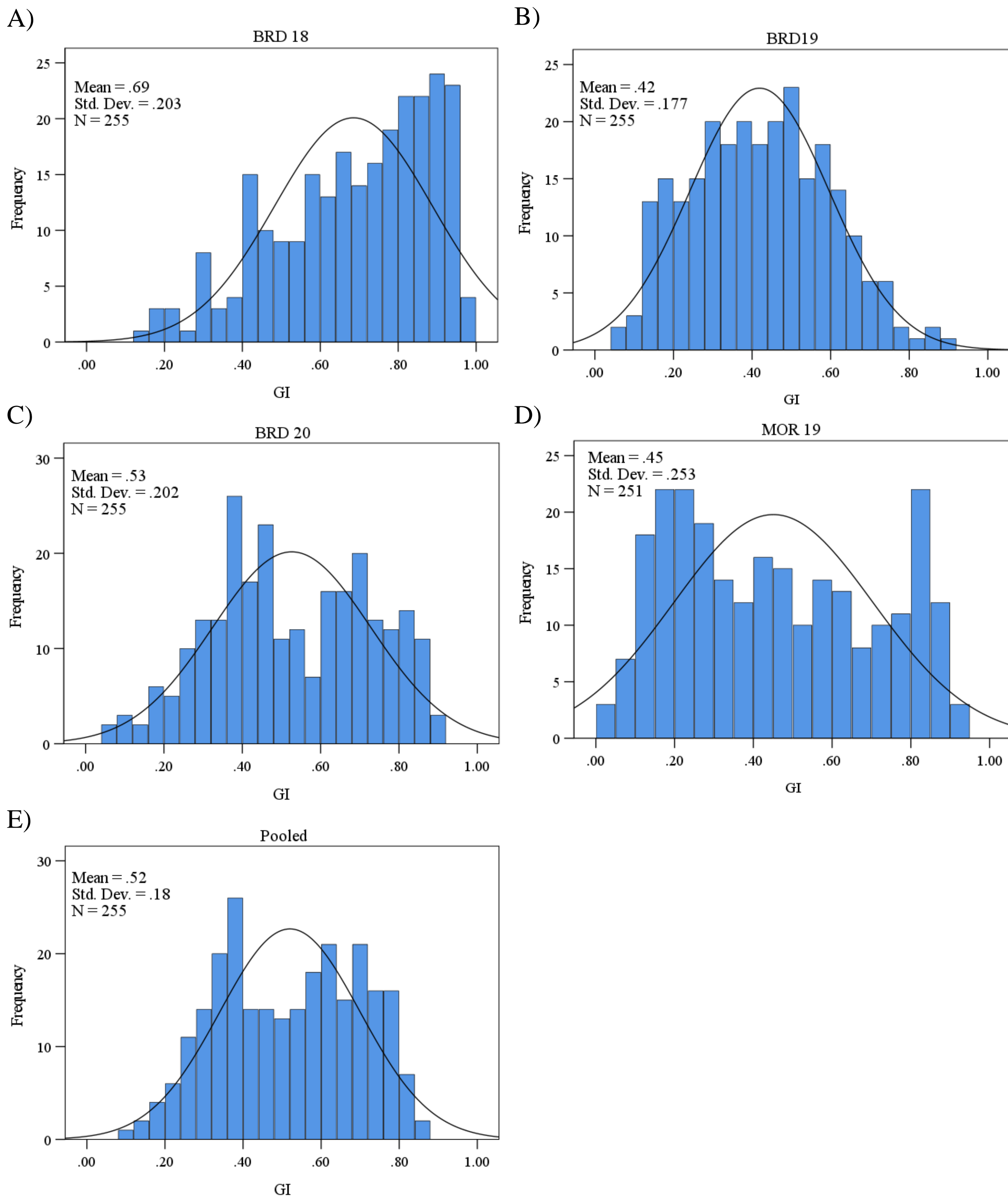


Figure 3.1 Frequency distribution of variation in germination index (GI) of the mapping panel evaluated over the four trail environments BRD18, Brandon trial in 2018 (A); BRD19, Brandon trial in 2019 (B); BRD20, Brandon trial in 2020 (C); MOR19, Morden trial in 2019 (D); pooled, data pooled from all trial environments (E). The black curves represent the normal distribution. The mean GI, standard deviation and total number of genotypes (N) tested in each environment are shown in each histogram.

3.3.2 Analysis of variance and environmental correlation

Analysis of variance of the GI values of the association panel revealed statistically significant differences ($P < 0.0001$) due to genotype, environment, and genotype and environment interaction (Table 3.3).

Table 3.3 Summary of analysis of variance of the germination index in the association panel.

Source	DF	Type III SS	Mean square	Expected Mean square	F	P > F
Environment[a] (E)	3	32.206	10.735		564.559	< 0.0001
Genotype (G)	254	97.217	0.383	Ve+rVge+reVg	20.128	< 0.0001
G×E	758	36.659	0.048	Ve+rVge	2.543	< 0.0001
Error	2026	38.525	0.019	Ve		
Total	3042	1030.717				

[a]Environment as described in Table 3.2
DF, degrees of freedom; SS, sum of square, Ve, variance due to error; Vg, variance due to genotype; Vge, variance due to genotype and environment interaction; r, number of replications; e, number of environments; P, probability

Correlation analysis of the GI scores from the trial environments detected significant and positive correlations ($P < 0.001$) among all environments (Table 3.4). The highest phenotypic correlation of 0.77 was observed between MOR19 and BRD18 as well as between MOR19 and BRD20 whereas the lowest but significant correlation of 0.45 was found between BRD19 and BRD18.

Table 3.4 Correlation coefficients of the germination index of the association mapping panel between the trial environments.

Environment[a]	BRD18	BRD19	BRD20
BRD19	0.45***		
BRD20	0.75***	0.52***	
MOR19	0.77***	0.55***	0.77***

[a]Environment as described in Table 3.2
***Significance at $P < 0.001$

3.3.3 Linkage disequilibrium

A total of 26,837 SNP markers mapped in the barley POPSEQ_2017 genetic map (Mascher et al. 2013) was used to estimate pairwise LD (r^2). The intra-chromosomal LD decay over genetic distance was measured by plotting significant ($p<0.001$) syntenic LD (r^2) values against genetic distance (cM) across the whole barley genome. The non-linear regression plot revealed that the LD decayed at a genetic distance of 1.05 cM at a critical threshold of $r^2 = 0.27$, which was derived from unlinked marker pairs (Figure 3.2). In accordance with the calculated LD decay, QTL size was determined by considering all significant markers found in the window of ± 1.05 cM as a part of single QTL.

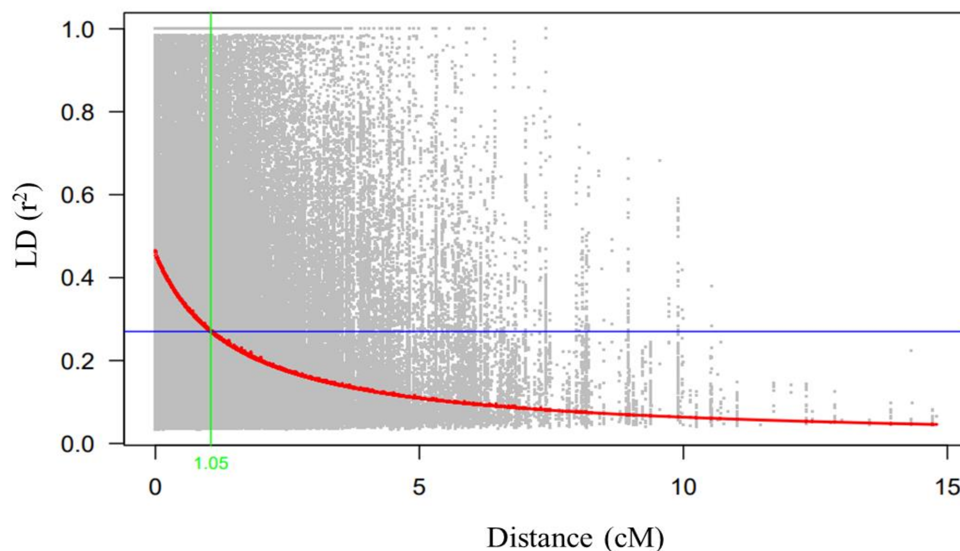


Figure 3.2 Genome wide linkage disequilibrium (LD) decay plot of intra-chromosomal marker pairs as a function of genetic distance (cM) for the association mapping panel. The red curve represents the LD decay based on non-linear regression, the horizontal blue line represents the critical r^2 value based on unlinked markers, and the vertical green line indicates the LD decay value in cM.

3.3.4 Population structure analysis

Three approaches were used for the estimation of number of subpopulations in the association panel consisting of the 255 barley genotypes (Figure 3.3). First, Bayesian cluster analysis using STRUCTURE was used to calculate the optimum number of subpopulations. The highest peak of Evanno's delta k was observed at K = 3, indicating that the genotypes are divided into three subpopulations (Figure 3.3A). The resultant bar plot demonstrated three subpopulations in three different colors, which mainly reflected the geographic origin of the barley genotypes (Figure 3.3B). The Q matrix estimated that subpopulation I (Q1) consists of 105 genotypes, subpopulation II (Q2) consists of 65 genotypes, and sub population III (Q3) consists of 70 genotypes, while the remaining 15 genotypes were considered as admixed. Second, PCA analysis revealed that the first two PC's, PC1 (9.14%) and PC2 (6.97%) explains 16.11% of the total variation (Figure 3.3C). The scatter plot between PC1 and PC2 consisted three subgroups that produced three distinct clusters with few admixed genotypes. Third, the phylogenetic tree analysis using the neighbor-joining clustering method, also inferred three different groups of the population as shown in the archaeopteryx tree (Figure 3.3D). Overall, the consistency between the results of STRUCTURE, PCA and phylogenetic tree validated that the association panel is composed mainly of three subpopulations.

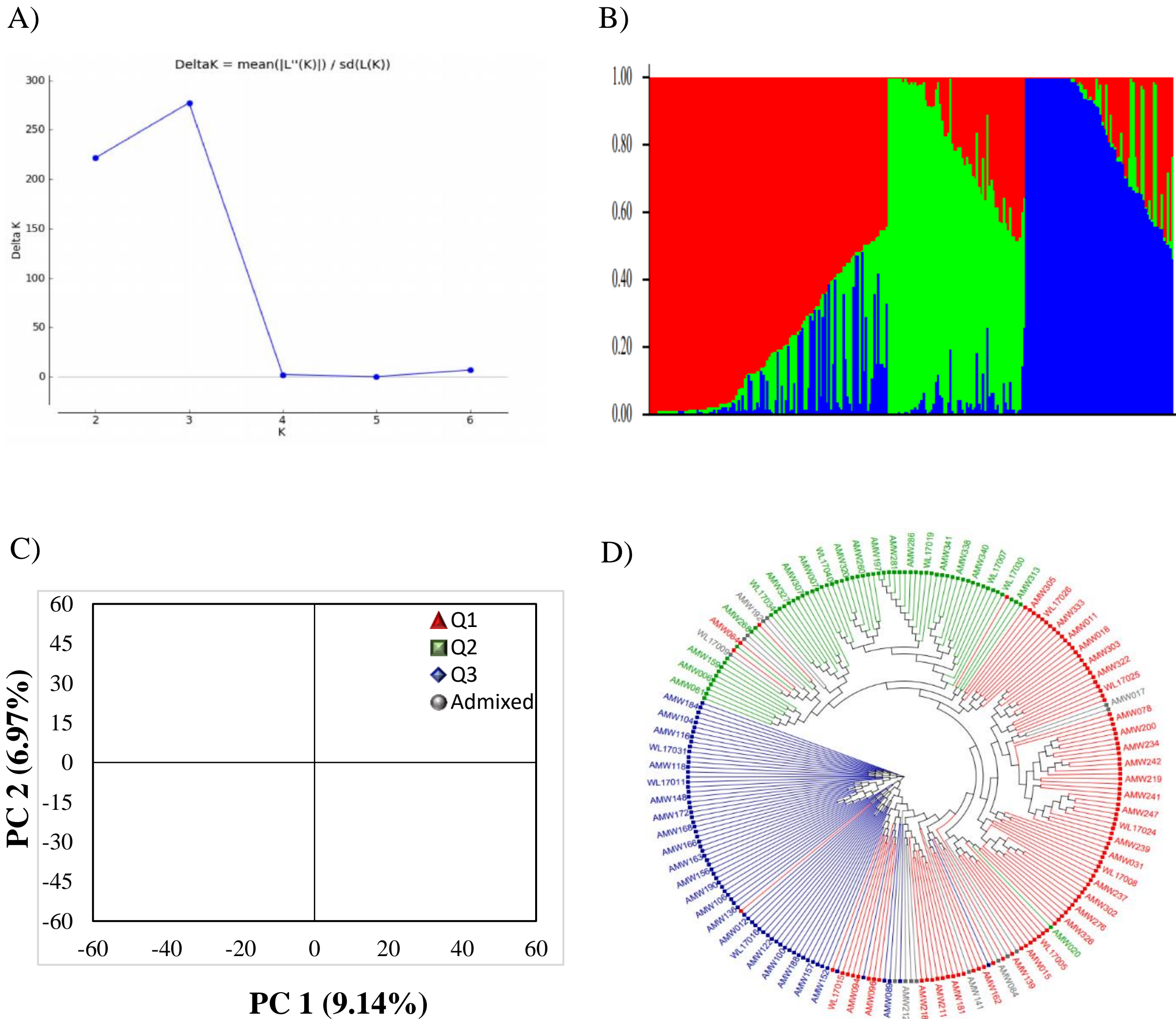


Figure 3.3 Estimation of number of subpopulations. Number of subpopulations are estimated using delta k method (A), the x and y-axis represents k values (number of subgroups) and the delta k values, respectively, and the highest peak point depicts the appropriate number of subpopulations; STRUCTURE (B), which divided the barley genotypes into three subpopulations represented by the three different Q-matrices - Q1 (red), Q2 (green) and Q3 (blue); Principal component analysis (PCA) (C), the scatterplot derived from PCA represents PC1(x-axis) and PC2 (y-axis) and the percentages shown in the parentheses indicate the proportion of variation explained by each component. Archaeopteryx tree of the association panel (D), showing three distinct clusters as Q1 (red), Q2 (green) and Q3 (blue) along with the admixed genotypes (grey).

3.3.5 Marker-trait associations

The cumulative distribution of *P* values across different models enabled identification of best fitting model, which was subsequently used for detection of marker trait associations. The QQ plots from all the models including GLM (PCA) (Figure 3.4A), GLM (Q) (Figure 3.4B), MLM (PCA+K) (Figure 3.4C) and MLM (Q+K) (Figure 3.4D), generated using the BLUEs pooled across all trial environments, were analysed. The observed -log10 *P* values were compared with the expected -log10 *P* values for each model and a highly uniform distribution of these *P* values was observed in MLM (PCA+K) and MLM (Q+K) models. However, the MLM (Q+K) model resulted in slightly better uniform distribution than the MLM (PCA+K) model. Therefore, the marker-trait association analysis was conducted with MLM (Q+K) model using 31,899 SNPs and the association panel of 255 genotypes.

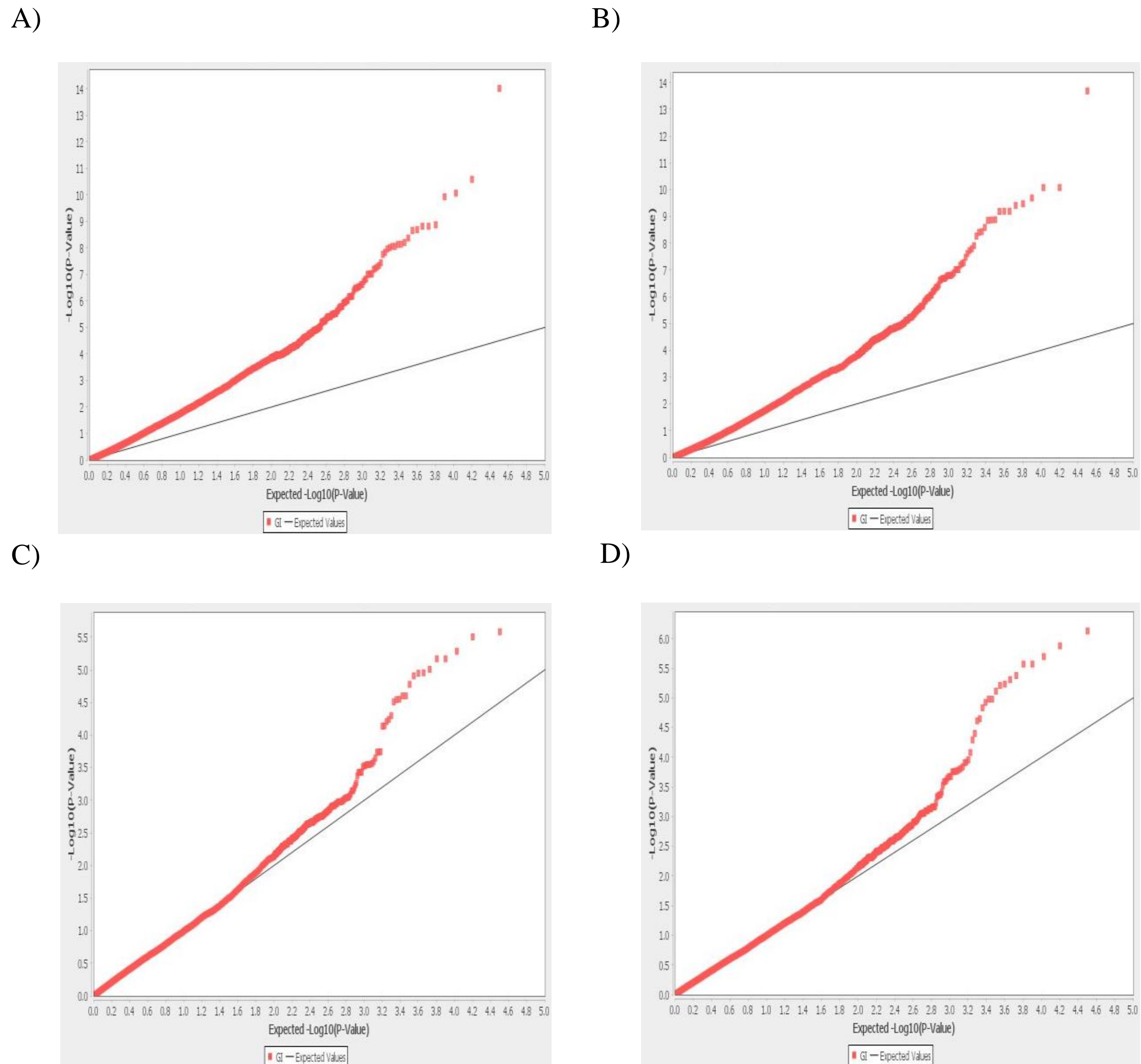


Figure 3.4 Quantile-quantile plots showing the observed -log10 *P* values against the cumulative -log10 (*P* values) derived from marker-trait association analysis of the dataset pooled from all trial environments using 31,899 SNP markers and the general linear model (GLM; PCA) (A); GLM (Q) (B); mixed linear model (MLM; PCA+K) (C) and MLM (Q+K) (D).

3.3.6 Detection of significant markers

The output from the MLM (Q+K) model is presented in Manhattan plot that was generated using the -log10 (*P* value) and marker positions (Figure 3.5). The markers were then examined for their significance using FDR adjusted *p* values < 0.05 and lowest 0.1 percentile of the *P* value distribution. The FDR method resulted in a *P* value threshold of 2.4×10^{-5} [-log10 (*P*) > 4.6] and

identified 16 significant SNPs associated with the PHS trait based on the dataset pooled from all trial environments, which are distributed on chromosomes 1H, 2H, 5H and 7H (Table 3.5). These markers were found to be responsible for phenotypic variation ranging from 7.53% to 11.69%. The most significant marker (-log10 (*P*) = 6.1) was located on chromosome 7H (JHI-Hv50k-2016-484268) and it explained 10.41% of the phenotypic variation, followed by JHI-Hv50k-2016-147824, which is located on chromosome 2H with a -log10 (*P*) of 5.9 and explained 10.01% of the phenotypic variation, and BOPA2_12_30360 on chromosome 5H with a -log10 (*P*) of 5.7 and explained 9.45% of the phenotypic variation. Out of the 16 significant markers, 4 markers namely, JHI-Hv50k-2016-48106, JHI-Hv50k-2016-100680, JHI-Hv50k-2016-481101 and BOPA2_12_30360 were detected as significant markers in two environments, BRD18 and MOR19.

Since the FDR correction method was observed to be stringent, the significant markers associated with dormancy in each and across all the trial environments were also analysed using the less strict 0.1 percentile of *P* value distribution method. The more permissive 0.1 percentile of *P* value method resulted in *P* value threshold of 2.0×10^{-4} [-log10 (*P*) > 3.7] and identified 30 significant SNP markers associated with PHS trait based on the dataset pooled from all trial environments (Appendix 1, 2). Among the 30 markers identified, five markers, namely JHI-Hv50k-2016-48106, JHI-Hv50k-2016-147824, BOPA2_12_30360, JHI-Hv50k-2016-481101 and JHI-Hv50k-2016-484268 were found to be significant in three trial environments (BRD18, BRD20, and MOR19) based on the lowest 0.1 percentile *P* value threshold value calculated for each environment. Comparison between two individual trial environments also identified common significant markers. The comparison between BRD20 and MOR19 identified 14 significant markers in common while eight common significant markers were identified in each comparison

between BRD18 and BRD20, and BR18 and MOR19 (Appendix 1). Comparison of the number of significant markers identified by both methods of determining the significance threshold in the individual and across all the trial environments is shown in Appendix 1.

Although the 0.1 percentile method detected more markers in common among the individual trial environments, the FDR method was preferred to narrow down the overall number of significant markers across all the trial environments.The positions of all the 16 significant SNP markers were anchored to the barley physical map based on the reference genome of Morex v1 and Morex v3 (Mascher et al. 2017, 2021) (Table 3.5, Appendix 3, 4) and the genetic map positions were obtained based on POPSEQ_2017 (Mascher et al. 2013) using Barleymap (Table 3.5).

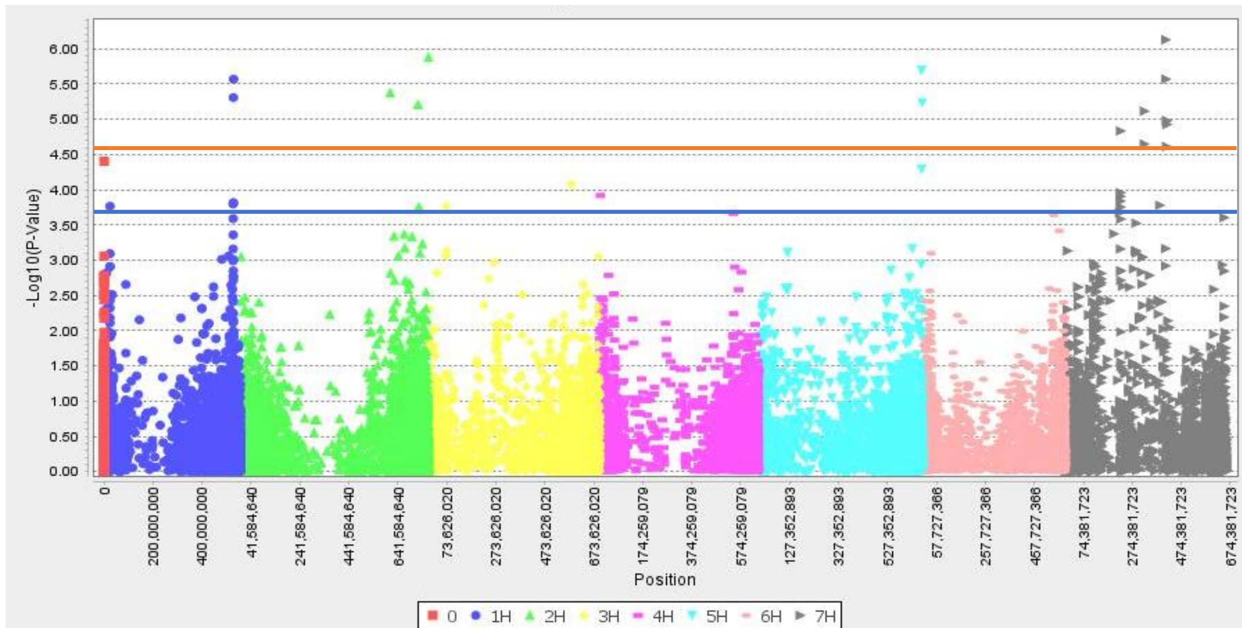


Figure 3.5 Manhattan plot derived from genome wide association analysis of the dataset pooled from all trial environments. Markers located on all the seven chromosomes of barley are shown in different colours. The x-axis represents the position of the markers while the y-axis indicates the -log10 (*P* value). The orange and blue-coloured horizontal lines indicate the FDR derived *P* value threshold of 4.6 and the 0.1 percentile *p* value threshold of 3.7, respectively.

3.3.7 QTL detection

The LD decay analysis resulted in identification of seven QTL regions on chromosome 1H (1 QTL), 2H (3 QTLs), 5H (1 QTL) and 7H (2 QTLs). The significant SNPs located within 1.05 cM between each other were considered to represent a single QTL, as such the JHI-Hv50k-2016-48106 (103.82 cM) and JHI-Hv50k-2016-48443 (104.75 cM) markers located on chromosome 1H represented a QTL defined as QTL.1H-1. The three QTLs on chromosome 2H were each detected by a single SNP (Table 3.5). The QTL.5H-1 is comprised of markers BOPA2_12_30360 (168.54 cM) and BOPA2_12_10322 (168.89 cM) located on chromosome 5H. The QTL.7H-1 contained two SNPs, JHI-Hv50k-2016-481098 and JHI-Hv50k-2016-481101 located on chromosome 7H at 67.92 cM while the QTL.7H-2 contained seven SNPs located at 70.54 cM (with six of the SNPs exhibiting very close physical proximity to each other); these SNPs include JHI-Hv50k-2016-476027, JHI-Hv50k-2016-484268, JHI-Hv50k-2016-484348, JHI-Hv50k-2016-484438, JHI-Hv50k-2016-484446, JHI-Hv50k-2016-484470 and JHI-Hv50k-2016-484514. Although, the position of JHI-Hv50k-2016-484438 was not defined in the POPSEQ map, we considered it as a part of the QTL.7H-2 based on its position in the Morex physical map (Table 3.5).

Table 3.5 Significant SNPs associated with PHS using the dataset pooled from all trial envirnoments based on FDR derived *p* value threshold[a].

QTL[b]	Marker name	Chr[c]	Position (bp)[d]	Position (bp)[e]	Position (cM)[f]	-log10 (*P*)[g]	R^2 (%)[h]
QTL.1H-1	JHI-Hv50k-2016-48106	1H	528156932	490365852	103.82	5.31	8.71
QTL.1H-1	JHI-Hv50k-2016-48443	1H	528878417	491096697	104.75	5.57	9.23
QTL.2H-1	JHI-Hv50k-2016-100680	2H	610410199	536025429	61.98	5.38	8.95
QTL.2H-2	JHI-Hv50k-2016-127728	2H	723373248	629611707	122.86	5.21	8.55
QTL.2H-3	JHI-Hv50k-2016-147824	2H	766316093	664038952	149.15	5.88	10.01
QTL.5H-1	BOPA2_12_30360	5H	666003793	583373875	168.54	5.70	9.45
QTL.5H-1	BOPA2_12_10322	5H	668721862	585992348	168.89	5.24	8.74
QTL.7H-1	JHI-Hv50k-2016-481098	7H	323672368	123258022	67.92	4.65	7.53
QTL.7H-1	JHI-Hv50k-2016-481101	7H	323672640	123257750	67.92	5.12	8.37
QTL.7H-2	JHI-Hv50k-2016-476027	7H	226518197	215025659	70.54	4.83	9.75
QTL.7H-2	JHI-Hv50k-2016-484268	7H	411369776	399422788	70.54	6.13	10.41
QTL.7H-2	JHI-Hv50k-2016-484348	7H	412253334	400377565	70.54	5.57	9.23
QTL.7H-2	JHI-Hv50k-2016-484438	7H	414363605	402010687	N/A	4.61	8.19
QTL.7H-2	JHI-Hv50k-2016-484446	7H	414470218	402026315	70.54	4.98	8.10
QTL.7H-2	JHI-Hv50k-2016-484470	7H	415119064	402635363	70.54	4.98	8.10
QTL.7H-2	JHI-Hv50k-2016-484514	7H	415786016	403282968	70.54	4.93	11.69

[a]Environment as described in Table 3.2
[b] Putative QTL containing one or more than one SNP- described on the basis of LD decay distance
[c]Chromosome
[d]Marker position in bp obtained from the barley reference genome of Morex v1 (Mascher et al. 2017)
[e]Marker position in bp obtained from the barley reference genome of Morex v3 (Mascher et al. 2021)
[f]Marker positions in cM derived from POPSEQ_2017 genome map using Barleymap (Mascher et al. 2013; http://floresta.eead.csic.es/barleymap/)
[g]Markers selected on the basis of FDR method, *P* value threshold of 2.4×10^{-5} [-log10 (*P*) > 4.6]
[h]R^2 (%) denotes the percentage of phenotypic variation explained by each significant SNP

3.3.8 Identification of candidate genes

Genes co-localized with the significant SNPs were identified as the putative candidate genes regulating PHS. Our analysis identified 13 candidate genes on chromosome 1H, 2H, 5H and 7H based on the physical map of the reference genome of Morex v1 (Table 3.6), and all the candidate genes identified contained significant markers. The genes associated with the SNPs were verified by BLAST searching the sequences of markers using GrainGenes against the reference genome sequence assembly of cv. Morex (Morex v2) by TRITEX (Monat et al. 2019). Since the association mapping panel used in this study mainly consists of two row barley genotypes, the candidate genes were also BLAST searched against reference genome sequence assembly of two row cv. GP v1 using EnsemblPlants. The contig regions in GP v1 were found overlapping with the target genes in the Morex v2 genome assembly (Appendix 5). The candidate genes exhibited high sequence identity with their respective homologs in other species including *Triticum dicoccoides, Aegilops tauschii, Brachypodium distachyon* and *Triticum aestivum* (Appendix 6).

The candidate genes, which represented several functional categories, were annotated using Barleymap and PANTHER v16 software (Table 3.6). Out of the 13 candidate genes, the *LDL1* and *PRORP1* genes were found to be associated with the two most significant markers, namely JHI-Hv50k-2016-484268 (-log 10(P) = 6.13) located in QTL.7H-2 and JHI-Hv50k-2016-147824 (-log 10(P) = 5.88) located in QTL.2H-3, respectively. Notably, the *GlyRs, LDL1, PUB8*, *CRLK1, GT* and *FIE* genes were associated with a single QTL on chromosome 7H (QTL.7H-2), which consists of seven significant SNPs (Table 3.5, 3.6).

3.3.9 Expression analysis of candidate genes

Two genotypes, Himalaya and Steptoe which represented the most non-dormant and dormant genotypes, respectively, were selected for expression analysis of the candidate genes. Analysis of the germination phenotype of Himalaya and Steptoe seeds used for expression studies revealed that seeds of Himalaya exhibit very high GI value while those of Steptoe show very low GI value (Figure 3.6).

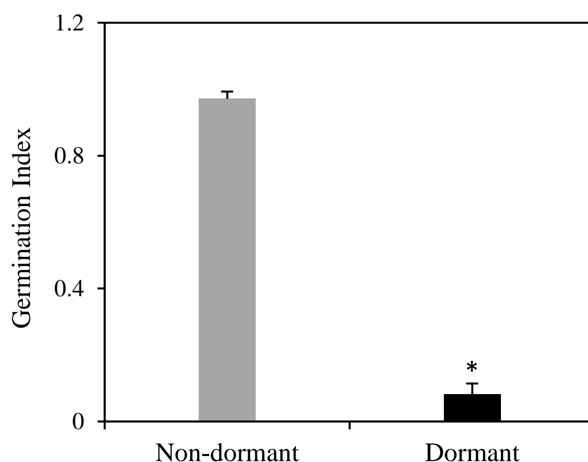


Figure 3.6 Germination index of non-dormant (Himalaya) and dormant (Steptoe) genotypes over a period of seven days. Data shown are means of three biological replicates ± SE. Statistically significant differences at $p<0.05$ (student's t test) are shown by asterisks.

Analysis of the expression patterns of the candidate genes in the embryos of 24 hours imbibed seeds of the two genotypes showed higher expression levels of *HvFIE* (~3-fold), *HvPM19* (over 8-fold), *HvGT* (over 18-fold) and *HvGlyRS* (1.4-fold) in the dormant genotype as compared to that observed in the non-dormant genotype (Figure 3.7A-D).

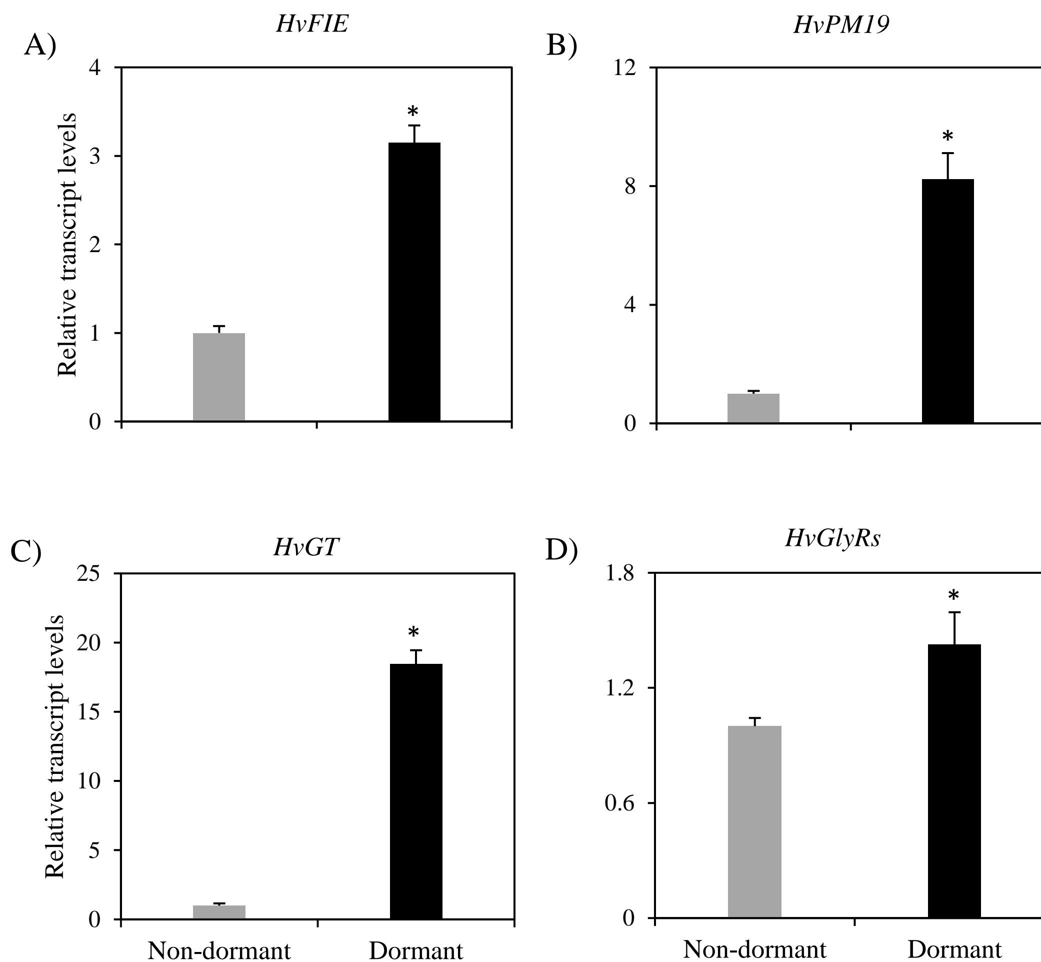


Figure 3.7 Relative transcript levels of *HvFIE* (A), *HvPM19* (B), *HvGT* (C) and *HvGlyRS (D)* in the embryo of non-dormant (Himalaya) and dormant (Steptoe) seeds imbibed for 24 h. The transcript level for each gene was determined using *Hvβ-actin* as a reference gene and calculated relative to that in the non-dormant seeds, which were set to 1. Data shown are means of three biological replicates ± SE. Statistically significant differences at $p<0.05$ (student's t test) are shown by asterisks.

As compared to that detected in the non-dormant genotype, lower expression levels of *HvLDL1* (1.7-fold), *HvMTHD* (2.9-fold), *HvPRORP1* (2.9-fold), *HvHUP6* (3.7-fold), *HvCRLK1* (3.7-fold) and *HvPRX16* (over 28-fold) were evident in the dormant genotype (Figure 3.8A-F). No significant difference in the expression level of *HvTOP1* and *HvPUB8* was observed between the two genotypes (Figure 3.8G, H). No expression of *HvJMT* was detected in both genotypes.

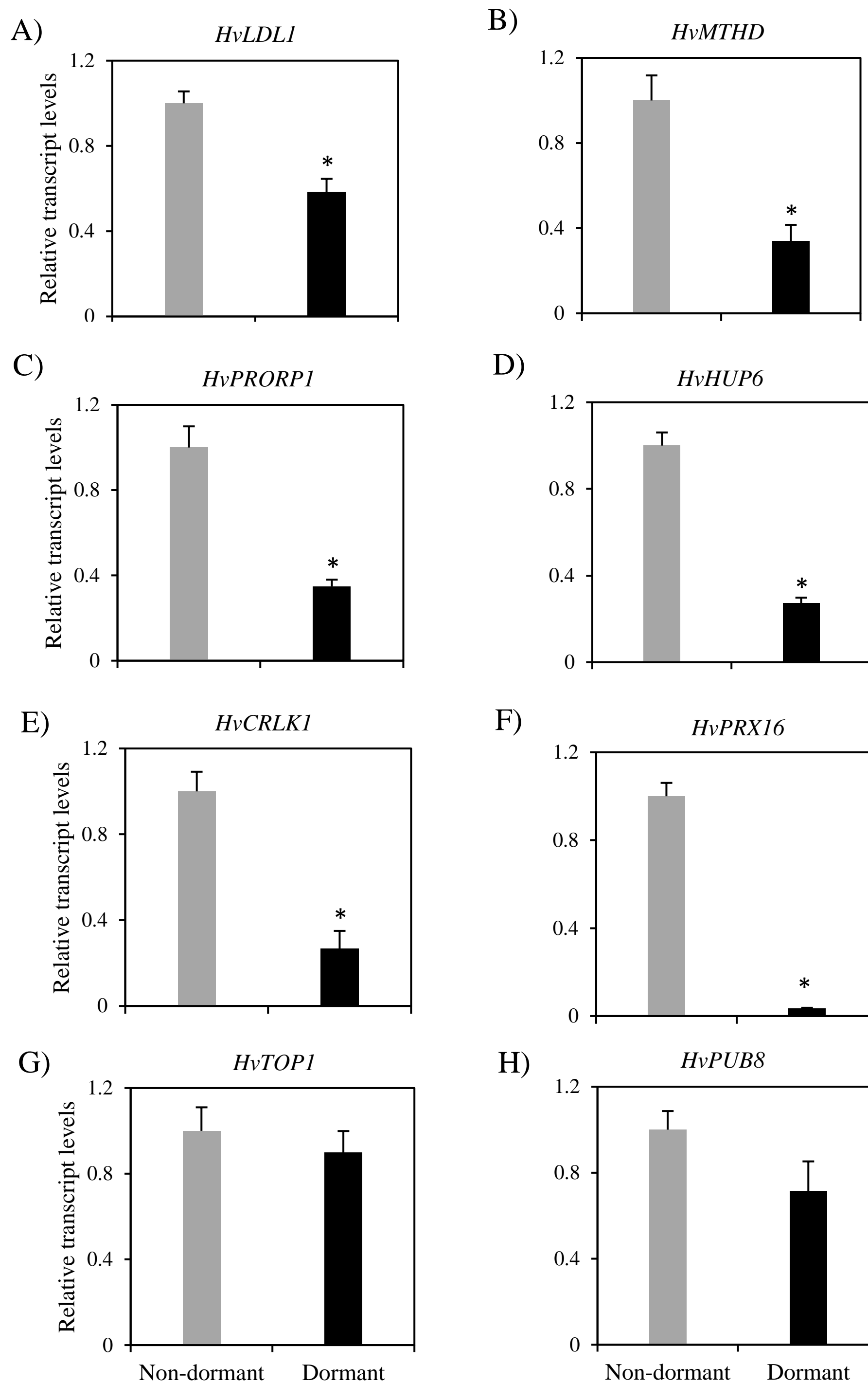


Figure 3.8 Relative transcript levels of *HvLDL1* (A), *HvMTHD* (B), *HvPRORP1* (C), *HvHUP6* (D), *HvCRLK1* (E), *HvPRX16* (F) *HvTOP1* (G) and *HvPUB8* (H) in the embryo of non-dormant (Himalaya) and dormant (Steptoe) seeds imbibed for 24 h. The transcript level for each gene was determined using *Hvβ-actin* as a reference gene and calculated relative to that in the non-dormant seeds, which were set to 1. Data shown are means of three biological replicates ± SE. Statistically significant differences at $p<0.05$ (student's t test) are shown by asterisks.

Table 3.6 Putative candidate genes co-localized with significant markers associated with PHS.

Short name	Gene ID	Chr	Start	End	Associated SNP	Description in Barleymap	Description in PANTHER Family/Subfamily database
MTHD	HORVU1Hr1G082150	1H	528873108	528879169	JHI-Hv50k-2016-48443	MaoC family dehydratase	Hydroxyacyl-thioester dehydratase type 2, mitochondrial-related (pthr43437:sf3)
HUP6	HORVU2Hr1G084190	2H	610406687	610410941	JHI-Hv50k-2016-100680	Kelch-like protein 20	Os04g0487100 protein (pthr46773:sf3)
PRX16	HORVU2Hr1G111610	2H	723371599	723375534	JHI-Hv50k-2016-127728	Peroxidase superfamily protein	Peroxidase 16-related (pthr31517:sf48)
PRORP1	HORVU2Hr1G127250	2H	766313033	766318728	JHI-Hv50k-2016-147824	proteinaceous RNase P 1	Mitochondrial ribonuclease p catalytic subunit (pthr13547:sf1)
JMT	HORVU5Hr1G124360	5H	666003200	666017332	BOPA2_12_30360	Jasmonate O-methyltransferase	Subfamily not named (pthr31009:sf57)
PM19	HORVU5Hr1G125460	5H	668718822	668723315	BOPA2_12_10322	AWPM-19-like family protein	AWPM-19-like family protein (pthr33294:sf5)
GlyRs	HORVU7Hr1G055340	7H	226514183	226520878	JHI-Hv50k-2016-476027	Glycine--tRNA ligase	Glycine--trna ligase (pthr10745:sf0
TOP1	HORVU7Hr1G065200	7H	323671408	323687387	JHI-Hv50k-2016-481098 JHI-Hv50k-2016-481101	DNA topoisomerase 1	Omega-protein (pthr42785:sf1)
LDL1	HORVU7Hr1G073020	7H	411366895	411370575	JHI-Hv50k-2016-484268	Lysine-specific histone demethylase 1 homolog 2	Lysine-specific histone demethylase 1a (pthr10742:sf373)
PUB8	HORVU7Hr1G073100	7H	412252306	412361703	JHI-Hv50k-2016-484348	U-box domain-containing protein	U-box domain-containing protein 8 (pthr23315:sf112)
CRLK1	HORVU7Hr1G073300	7H	414469385	414474517	JHI-Hv50k-2016-484438 JHI-Hv50k-2016-484446	Calcium/calmodulin-regulated receptor-like kinase 1 (Protein kinase superfamily protein)	Calcium/calmodulin-regulated receptor-like kinase 1 (pthr27001:sf542)
GT	HORVU7Hr1G073370	7H	415116742	415120193	JHI-Hv50k-2016-484470	Protein of unknown function (DUF604)	Glycosyltransferase (pthr10811:sf25)
FIE	HORVU7Hr1G073410	7H	415781576	415787023	JHI-Hv50k-2016-484514	Polycomb group protein FERTILIZATION-INDEPENDENT ENDOSPERM	Polycomb protein EED (pthr10253:sf5)

3.4 Discussion

Pre-harvest sprouting in barley is a complex trait controlled by multiple genes. As a result, identifying the genomic regions/genes associated with it is a crucial step for the development of cultivars with PHS resistance. This study was aimed at identifying QTLs and genes controlling PHS resistance using GWAS through examining the genotypic and GI data of an association panel consisting of 255 barley genotypes over four environments. Our analysis of the GI data revealed variation in dormancy level/PHS resistance among the genotypes across the four trail environments (Table 3.2, Figure 3.1). The BLUEs were computed by considering trial years as fixed effects and replications as random effects using Meta-R software (Hess et al. 2006) to eliminate any discrepancies in the GI data that might be influenced by environmental factors. As such the wide variation in the dormancy level among the genotypes is the result of genetic variation, which confirms the suitability of the association panel used for uncovering genomic regions/genes for PHS resistance. Our analysis also revealed a heritability of 0.87, based on the dataset pooled from all environments, indicating strong influence of heritable genetic effects with respect to dormancy (Table 3.2).

Evaluating subpopulations in marker-trait association analysis is imperative as it overcomes the problems that arise from false associations due to population structure. It has been reported that the barley population structure is affected mainly by geographical origin, growth habit and row type (Hamblin et al. 2010; Malysheva-Otto et al. 2006). Three methods namely, PCA, delta K and phylogenetic tree, are often used to estimate the number of subpopulations. The PCA and delta K method have been widely used to validate population structure results in several association studies (Milner et al. 2019; Mwando et al. 2020; Thabet et al. 2018). A recent association analysis for waterlogging tolerance has estimated three subpopulations in their

association panel consisting of 247 spring barley genotypes collected from different parts of the world by employing all three methods (Borrego-Benjumea et al. 2021). Consistently, analysis of the structure of the association panel, which consists of a similar association panel of 255 spring barley genotypes originated from around the world, considered in this study using PCA, delta K and phylogenetic tree methods, identified three subpopulations. Moreover, the analysis showed that 15 genotypes out of the 255 genotypes are admixed since the estimated proportion of these genotypes was less than 0.5 in any of the subgroups, indicating their admixed ancestry (Figure 3.3).

Linkage disequilibrium decay, which is generally calculated in the form of correlation coefficient (r^2) between the marker pairs on each chromosome (Zhu et al. 2008), is a crucial factor in association mapping. A range of factors influences how rapidly LD decays including effective population size, demographic histories, admixture, mating system, recombination rate as well as selection effects (Gaut and Long 2003; Nordborg et al. 2002). Previous studies have indicated that the extent of LD in barley ranges from 1 cM to over 10 cM depending on the type of population used (Comadran et al. 2009; Kraakman et al. 2004; Malysheva-Otto et al. 2006; Nagel et al. 2019; Zhang et al. 2009). Moreover, a rapid LD decay has been reported at a genetic distance ranging between 0 cM to 4 cM at $r^2 = 0.2$ in diverse barley genotypes (Bengtsson et al. 2017), and LD decays more quickly in genetically and geographically diverse germplasms (Mohammadi et al. 2015). In this study, the average genome-wide intra-chromosomal LD decay was estimated to be 1.05 cM ($r^2 = 0.27$) (Figure 3.2). Such a high LD decay rate enhances the mapping resolution as it narrows down the QTL interval, however, a higher number and marker density is required (Zhu et al. 2008). Therefore, in the current study, the markers were derived from 50k Barley iSelect SNP

Chip (Bayer et al. 2017) that provided a total of 31,899 SNP markers, which appeared to be sufficient to cover the whole barley genome and perform GWAS.

Furthermore, LD decay distance is useful in determining the size of a QTL and the associated genes. A recent GWAS has shown that significant SNPs associated with the traits of interest can be grouped into a QTL region based on the LD decay distance (Abdel-Ghani et al. 2019). Likewise, in this study, the identified SNP markers located within ±1.05 cM were considered as a part of a single QTL. Reproducibility of association mapping results depend on robustness of the phenotypic data, choice of model and method of handling spurious associations (Alqudah et al. 2020). Thus, the current study calculated BLUEs to obtain unbiased mean estimates of the phenotypic data, which is required to perform association analysis. Out of the four statistical models, the MLM (Q+K) model was selected for association analysis (Figure 3.4) mainly because of its high statistical power to control false positives; the model can correct false marker-trait associations arising due to both population structure and familial relatedness (Jabbari et al. 2018; Yu et al. 2006). Using the FDR adjusted P values < 0.05, we identified 16 significant marker trait associations leading to seven QTL regions on chromosome 1H, 2H, 5H and 7H that are associated with dormancy/PHS using the dataset pooled from all trial environments.

Previous QTL analyses for seed dormancy in barley identified major QTLs, SD1 and SD2 near the centromeric and telomeric regions of the long arm of chromosome 5H, respectively (Han et al. 1996; Oberthur et al. 1995). These QTLs are considered as the most important QTLs for controlling dormancy (Lohwasser et al. 2013; Romagosa et al. 1999; Sato et al. 2009). In this study, one QTL was detected on chromosome 5H, designated as QTL.5H-1, around the telomeric region (168.54 cM to 168.89 cM) that appeared to be mapped to the SD2 region, indicating its potential in regulating seed dormancy. This QTL contained two significant SNPs named

BOPA2_12_10322 and BOPA_12_30360 (Table 3.5). A recent study has reported a significant marker trait association in the SD2 region, which is linked to one of the key GA biosynthesis genes, *HvGA20ox1* (Yamaguchi et al. 2008), at 168.9 cM (Nagel et al. 2019). Consistently, the significant SNP (BOPA2_12_10322) that explains 8.74% of the phenotypic variation was found to be associated with *HvGA20ox* at 168.89 cM on chromosome 5H, indicating that this SNP might influence seed dormancy.

This study detected many genes that were found to be associated with the significant SNP markers and located within the QTL interval, and as such some of these genes could play important roles in regulating seed dormancy/PHS. Of these genes, 13 genes were identified to harbour the significant SNP markers, and thus considered as causal candidate genes. The study analyzed the expression patterns of these candidate genes in dormant and non-dormant seeds to gain further insights into their roles in regulating seed dormancy. The SNP, BOPA2_12_10322 at 168.89 cM (668.7 Mbp) on telomeric region of chromosome 5HL was found to be located within the gene designated as *Plasma Membrane Protein 19* (*PM19*) (Table 3.6, Appendix 3). The PM19 belongs to AWPM-19 (ABA-induced wheat plasma membrane polypeptide-19) protein family and its corresponding gene *PM19* has been previously reported to be associated with seed dormancy in wheat and barley (Barrero et al. 2015; Koike et al. 1997; Ranford et al. 2002). In agreement with this, the *PM19* gene, located in the QTL region detected on 5HL (Zhou et al. 2016), is expressed in developing seeds of barley (Ranford et al. 2002). In this study, expression analysis of *HvPM19* revealed its higher level of expression in the embryo of dormant seeds than in non-dormant seeds. Similarly, high expression level of *HvPM19* was detected in embryos of dormant seeds of barley for up to 72 h after imbibition (Ranford et al. 2002).

The BOPA_12_30360 SNP, which is located at 168.54 cM (666.0 Mbp) of the QTL.5H-1 with the third highest -log10(*P*) value of 5.7 and explained 9.45% of the phenotypic variation, was identified as significant SNP in two environments (Table 3.5, Appendix 1,3). This SNP was located within the gene encoding jasmonate O-methyltransferase (*HvJMT*) (Table 3.6), which catalyzes the conversion of jasmonic acid into methyl jasmonate. A study in wheat has been shown that jasmonates such as methyl jasmonate control seed dormancy by regulating ABA level through altering the expression level of ABA biosynthesis and catabolic genes (Jacobsen et al. 2013). However, the expression of *HvJMT* was not detected in both dormant and non-dormant genotypes.

Dormancy related minor QTLs including SD3 on chromosome 7H and SD4 on chromosome 4H have been identified in barley previously (Han et al.1996; Oberthur et al. 1995). Furthermore, a linkage mapping study has detected a dormancy QTLs in the centromeric region of chromosome 7H, which is equivalent to the SD3 QTL region in the Steptoe/Morex population (Ullrich et al. 2009). Three QTLs that are responsible for dormancy and PHS in barley have also been detected near the centromeric region of chromosome 7H, using segregation and association mapping (Lohwasser et al. 2013). Consistently, this study found two QTLs around the centromere region of 7H (QTL.7H-1 and QTL.7H-2). The QTL.7H-1 consisted of two SNPs, namely JHI-Hv50k-2016-481098 and JHI-Hv50k-2016-481101 at 67.92 cM (323.6 Mbp), that explained 7.53% and 8.37% of the phenotypic variation, respectively (Table 3.5, Appendix 3). Both of these SNPs are found to be located within the gene encoding DNA topoisomerase 1 (*HvTOP1*) (Table 3.6). The *TOP1* gene has been shown to have a role in DNA replication, transcription and repair activities (Wang 1985), and a previous study in rice has reported TOP1 activity during early germination stages (Yoshida et al. 1991). However, no significant difference in the expression level of *HvTOP1* was detected between the dormant and non-dormant genotypes (Figure 3.8G).

Given that this gene has an active role in cell division and growth (Wang 1985; Yoshida et al. 1991), it might be subjected to regulation at the post-transcriptional level.

The other significant QTL.7H-2 was positioned near the centromere of 7HL at 70.54 cM and it consists of seven significant SNPs including JHI-Hv50k-2016-476027, JHI-Hv50k-2016-484268, JHI-Hv50k-2016-484348, JHI-Hv50k-2016-484438, JHI-Hv50k-2016-484446, JHI-Hv50k-2016-484470 and JHI-Hv50k-2016-484514, accounting for phenotypic variation ranging from 8.10% to 11.69% (Table 3.5). The JHI-Hv50k-2016-476027 SNP was genetically mapped at 70.54 cM but physically mapped at 226.5 Mbp while the SNPs (JHI-Hv50k-2016-481101 and JHI-Hv50k-2016-481098) mapped at 67.92 cM were located at 323 Mbp, indicating that the SNP's predicted physical position might be incorrect (Table 3.5, Appendix 3). To verify this, the positions of these three SNPs on the physical map of Morex v1 were compared with their positions on the physical map of Morex v3 (Appendix 4). Our comparison showed that JHI-Hv50k-2016-481101 and JHI-Hv50k-2016-481098 SNPs are mapped at 123 Mbp, justifying that these two SNPs represent a separate QTL (QTL.7H-1) at 67.92 cM (Table 3.5). Whereas JHI-Hv50k-2016-476027 was mapped at 215 Mbp, thus, this study considered it as a part of the QTL.7H-2 based on its genetic position (70.54 cM) (Table 3.5). This SNP marker was found to be located within the gene encoding enzyme glycine-tRNA ligase (*HvGlyRS*) (Table 3.6). The GlyRs belongs to the protein family of aminoacyl-tRNA synthetases (Eriani et al. 1990) that have been reported to have role in embryo development in Arabidopsis (Berg et al. 2005). Although the role of this gene in seed dormancy has not been characterized to date, it showed differential expression between dormant and non-dormant genotypes, that is, higher level of expression in the dormant than the non-dormant genotype (Figure 3.7D).

JHI-Hv50k-2016-484268, a SNP mapped at 70.54 cM (411.3 Mbp), is the topmost significant marker with a -log10(*P*) value of 6.13 and explaining the second highest phenotypic variation of 10.41%. (Table 3.5). The current study detected that this SNP was located within the the gene encoding lysine-specific histone demethylase 1 homolog 2 (*HvLDL1*) (Table 3.6). A study in Arabidopsis has been shown that LDL1 plays an important role in seed dormancy through negative regulation of the expression of genes that are known to control seed dormancy including *DOG1*, *Absicisic Acid 2* (*ABA2*) and *Abscisic Acid Insensitive 3* (*ABI3*) (Zhao et al. 2015). Consistent with these results, lower expression level of *HvLDL1* was evident in the dormant than the non-dormant genotype (Figure 3.8A), suggesting that it acts as a negative regulator of seed dormancy in barley seeds.

The other SNP in the QTL-7H.2, designated as JHI-Hv50k-2016-484348 and mapped at 412.2 Mbp (Table 3.5, Appendix 3) was found to be located within the gene encoding U-box domain-containing protein 8 (*HvPUB8*), which is a member of E3 ubiquitin ligases (Azevedo et al. 2001) (Table 3.6). It has been reported previously that PLANT U-box (PUB) E3 ligases regulate ABA signaling both positively and negatively (Seo et al. 2019; Yee and Goring 2009). For instance, *PUBs* such as *PUB18* and *PUB19* of Arabidopsis are reported to be upregulated in response to ABA (Seo et al. 2012). Despite these reports, no difference in the expression level of *HvPUB8* was apparent between the dormant and non-dormant genotypes (Figure 3.8H).

The other SNPs in QTL.7H.2, namely JHI-Hv50k-2016-484438 (414.3 Mbp) and JHI-Hv50k-2016-484446 (414.4 Mbp) (Table 3.5, Appendix 3) were found to be located within the gene encoding calcium/calmodulin-regulated receptor-like kinase 1 (*HvCRLK1*), which belongs to the protein kinase superfamily protein (Table 3.6). A previous study has reported the function of *CRLK1* in controlling stress response via its interaction with MAPK (Yang et al. 2010). The

MAPK signal relay system consists of MAPK, MAPK kinase (MKK) and MAPK kinase kinase (MKKK) proteins (Mao et al. 2019), of which MKK3 has been reported to negatively regulate seed dormancy in wheat and barley, thereby promoting germination (Nakamura et al. 2016; Torada et al. 2016). A study in rice has also demonstrated that the MKKK62-MKK3-MAPK7/14 module has a negative effect on seed dormancy (Mao et al. 2019). In agreement with these reports, *HvCRLK1* exhibited lower expression level in the dormant than non-dormant genotype (Figure 3.8E). Further studies are needed to understand whether interaction of *HvCRLK1* with the MAPK cascade pathway has a role in regulating seed dormancy.

JHI-Hv50k-2016-484470 of QTL.7H.2, a SNP located at 415.1 Mbp (Table 3.5, Appendix 3), was found to be located within the gene encoding glycosyltranferase (*HvGT*) (Table 3.6), an enzyme involved in the formation of glycosidic linkages. Although the specific role of this gene in seed dormancy is not known, our analysis indicated higher expression level of *HvGT* in dormant genotype as compared to that found in the non-dormant genotype (Figure 3.7C). The other SNP in QTL.7H-2, JHI-Hv50k-2016-484514, which is located at 415.7 Mbp and explained the highest phenotypic variation of 11.69% (Table 3.5, Appendix 3), was found to be located within the gene encoding fertilization-independent endosperm (HvFIE) protein, which belongs to the polycomb group proteins (Table 3.6). In barley, *HvFIE* has been mapped to the centromeric region of chromosome 7H and its expression in seeds is positively regulated by ABA (Kapazoglou et al. 2010), which is known to induce dormancy. Similarly, a higher expression level of *HvFIE* was observed in the dormant than the non-dormant genotype (Figure 3.7A).

A previous study that used biparental mapping population of barley has identified two QTLs controlling dormancy and PHS, which are located on the short and long arms of chromosome 2H, respectively (Lohwasser et al. 2013). The dormancy QTL was found to be associated with few

proteins in rice and wheat, of which the peroxidase protein was considered as a predominant player in the regulation of seed dormancy. The current GWAS identified 3 QTLs (QTL.2H-1,2,3), each containing one SNP, on the long arm of chromosome 2H (Table 3.5, Appendix 3). One of these SNPs, SNP JHI-Hv50k-2016-127728, located at 122.86 cM (723.3 Mbp) accounted for 8.55% of the phenotypic variation and was located within *HvPRX16*, which encodes peroxidase superfamily protein (Table 3.6). A study in Arabidopsis has demonstrated that *PRX16,* which encodes class III peroxidases, has a role in seed germination (Jemmat et al. 2020; Linkies et al. 2010). Consistently, *HvPRX16* showed very low expression level in the dormant genotype as compared to that detected in the non-dormant genotype (Figure 3.8F), suggesting that it acts as a negative regulator of seed dormancy.

The second most significant SNP (JHI-Hv50k-2016-147824), which represented QTL.2H-2, and mapped at 149.15 cM (766.3 Mbp) and explained 10.01% of the phenotypic variation (Table 3.5, Appendix 3), was found to be located within *mitochondrial proteinaceous RNase P 1*/(*HvPRORP1*), which encodes a protein that consists of pentatricopeptide repeat (PPR) motifs (Table 3.6). Previous studies have demonstrated that genes encoding PPR domain containing proteins are transiently expressed during seed germination (Narsai et al. 2011; Schmitz-Linneweber et al. 2008; Xue et al. 2021). Likewise, the non-dormant genotype showed higher expression level of *HvPRORP1* than the dormant genotype (Figure 3.8C), indicating its possible role in seed germination.

Another significant SNP in QTL.2H-3, which is designated as JHI-Hv50k-2016-100680 and located at 61.98 cM (610.4 Mbp) on chromosome 2H, was detected in 50% of the trial environments (Table 3.5, Appendix 1,3). This SNP, which is responsible for 8.95% of the phenotypic variation, is located within the *hypoxia response unknown protein 6* (*HvHUP6)* gene

that encodes the Kelch-like protein 20 (Table 3.6). It has been shown previously that wheat and barley seeds are low oxygen (hypoxia) sensitive, as such they do not exhibit amylase activity and their germination is inhibited in the absence of oxygen (Guglielminetti et al. 1995). A hypoxic environment created by the glumellae could interfere with ABA metabolism in the embryo of barley seeds and thereby leading to induction of dormancy (Benech-Arnold et al. 2006). Furthermore, induction in the expression of *HUP* genes has been reported in Arabidopsis seeds exposed to low oxygen level (Mustroph et al. 2010). Contrary to these reports, *HvHUP6* showed lower expression level in the dormant genotype as compared to the non-dormant genotype (Figure 3.8D), suggesting that this gene could be regulated at post-transcriptional level.

A previous study that used a Harrington/TR306 barley mapping population also detected dormancy QTL on chromosome 1HL (Ullrich et al. 2009) and the presence of dormancy QTL on chromosome 1H was also confirmed by another study (Bonnardeaux et al. 2008). Our analysis identified QTL.1H-1 (103.82 cM to 104.75 cM) containing two SNPs, JHI-Hv50k-2016-48106 positioned at 103.82 cM (528.1 Mbp) and JHI-Hv50k-2016-48443 at 104.75 cM (528.8 Mbp). Although, SNP JHI-Hv50k-2016-48106 was detected in 50% of the trial environments, no associated candidate gene was identified (Appendix 1). The other significant SNP, which was located on the long arm of chromosome 1H and explained 9.23% of the phenotypic variation (Table 3.5, Appendix 3), is found to be located within the gene encoding mitochondrial hydroxyacyl-thioester dehydratase type 2 protein containing the MaoC family dehydratase domain (*HvMTHD*) (Table 3.6). Although its role in regulating dormancy has not been reported to date, *HvMTHD* showed lower level of expression in dormant as compared to the non-dormant genotype (Figure 3.8B), providing insights into its role in dormancy and germination of barley seeds.

4.0 GENERAL DISCUSSION AND CONCLUSION

Barley is an important crop used for human consumption, brewing purposes and animal feed. Pre-harvest sprouting, the germination of seeds while on the mother plant, is a major recurring problem in the production of barley as it causes drastic reduction in yield and end use quality. It occurs due to exposure of mature seeds to wet and humid weather conditions before harvest. Seeds damaged by PHS loose their viability, therefore they are no longer considered desirable for malting purposes. The phenomenon of PHS is linked with the degree of dormancy in which seeds with low level of dormancy are more prone to PHS whereas seeds with high level of dormancy are less prone to PHS. However, strong dormancy is also not favourable as it may lead to non-uniform germination, delay in the malting process, and an increase in storage costs as after-ripening of strongly dormant seeds require storage for longer duration. Therefore, there is a need to develop barley cultivars with an intermediate level of seed dormancy. Seed dormancy is a quantitative trait influenced by genetic and environmental factors. Several genetic studies have been performed in order to identify genes/genomic regions associated with dormancy and PHS. Over the past decade, GWAS has been widely used to study quantitative traits in barley such as grain quality, abiotic stress tolerance and disease resistance. This is because of its high mapping resolution as it involves the use of genetically diverse population to identify marker trait associations.

In the present study, a diverse set of 255 barley genotypes and 31,899 SNPs obtained from iSelect 50K Illumina Infinium iSelect SNP chip were used to identify QTLs and genes related to dormancy/PHS. The association mapping panel was divided into three subpopulations and the average genome wide intrachromosomal LD decayed rapidly at 1.05 cM ($r^2 = 0.27$), suggesting high genetic diversity. The dormancy phenotypic data obtained from the mapping panels grown over four environments was determined based on GI values, and our results revealed variation in

the level of dormancy consisting of highly dormant, moderately dormant and low dormant genotypes. The variation as well as high heritability of the GI across the environments depicted the robustness of the dormancy phenotype. For analyzing the association between the phenotypic and genotypic data, the best fitting model MLM (Q+K) was used as it minimizes the false marker-trait associations due to population structure and familial relatedness. Based on this model and FDR derived significant threshold of -log10 (*P*) = 4.6, a total of 16 significant marker trait associations were identified, and this led to the detection of seven putative QTL regions containing one or more SNPs on chromosomes 1H, 2H, 5H and 7H. The QTL positions were determined in cM and bp using Barleymap based on POPSEQ 2017 and Morex genome map, respectively. The QTLs detected in the telomeric region of 5H (QTL.5H-1 at 168.54 to 168.89 cM) and centromeric region of 7H (QTL.7H-1 at 67.92 cM and QTL.7H.2 at 70.54 cM) appeared to be comparable with the previously identified SD2 and SD3 QTLs, respectively. The present findings also identified a dormancy QTL on 1HL (QTL.1H-1) and three QTLs on 2HL (QTL.2H-1, 2, 3). Although other studies have also identified QTLs on 1H and 2H, we can not directly compare the positions of our QTLs with those detected by the other studies as the respective mapping experiments involved different markers and crosses. Moreover, the order of markers and the map distance vary among crosses depending on map saturation. The use of different statistical methods and marker significance thresholds for mapping is also another factor that contributes to the difficulty of comparing QTL positions derived from different genetic mapping studies.

Given that dormancy and PHS are quantitatively inherited traits, this study identified several genes as candidate genes associated with the significant markers. To narrow down the list of genes, we considered only the genes containing the significant SNP markers within itself as potential candidates responsible for regulating dormancy/PHS. Consequently, 13 candidate genes

were identified. Seven of these genes, namely *HvFIE*, *HvLDL1*, *HvPROPRP1*, *HvPRX16*, *HvPM19*, *HvTOP1* and *HvJMT*, have been earlier reported to have roles in seed dormancy and germination. Whereas no information about the role in dormancy and germination is available for the remaining six genes, namely *HvHUP6*, *HvMTHD*, *HvGlyRs*, *HvPUB8*, *HvCRLK1* and *HvGT*. Three SNPs that have been shown to be responsible for highest phenotypic variation (> 10%) in the present study are located within *HvPRORP1* on chromosome 2H, and *HvFIE* and *HvLDL1* on chromosome 7H, indicating the importance the three genes and their SNPs in controlling dormancy/PHS trait.

Our gene expression data revealed that the *HvFIE*, *HvPM19*, *HvGT* and *HvGlyRs* genes exhibit higher expression levels in the dormant as compared to the non-dormant genotype, indicating their positive role in the regulation of dormancy. In contrast, the expression levels of *HvLDL1, HvMTHD*, *HvPRORP1*, *HvHUP6*, *HvCRLK1* and *HvPRX16* were relatively lower in the dormant genotype but higher in the non-dormant genotype, suggesting their negative roles in the regulation of seed dormancy. However, the expression levels of *HvPUB8* and *HvTOP1* did not show significant difference between the two tested genotypes while the expression of *HvJMT* was not detected in any of the genotypes. Expression results of the remaining genes that are known to negatively and positively regulate seed dormancy were in accordance with the previous reports. Determining the roles of previously uncharacterized genes, which showed differential expression between the dormant and non-dormant genotypes, in regulating seed dormancy and PHS requires further investigation.

Overall, this study identified novel loci and candidate genes as well as validated previously identified QTLs associated with seed dormancy/PHS. The use of a diverse set of germplasm along with the presence of high-density DNA marker platform and high-quality reference genome

assembly enabled precise detection of genomic regions and genes linked to seed dormancy and therefore PHS. Through genome wide association analysis, we were able to detect genomic regions associated with dormancy QTLs in the form of SNP markers and identify functional SNPs which could directly be used in marker-assisted selection of PHS resistant genotypes. Further studies on the physiological role of the casual genes of the genomic regions identified by this study could be of great importance to improve our understanding of the genetics of seed dormancy and assist the development of barley cultivars that possess an optimum level of dormancy to prevent PHS.

LITERATURE CITED

Abdel-Ghani AH, Sharma R, Wabila C, Dhanagond S, Owais SJ, Duwayri MA, Neumann K (2019) Genome-wide association mapping in a diverse spring barley collection reveals the presence of QTL hotspots and candidate genes for root and shoot architecture traits at seedling stage. BMC Plant Biol 19:1-19

Abe A, Takagi H, Fujibe T, Aya K, Kojima M, Sakakibara H, Terauchi R (2012) *OsGA20ox1*, a candidate gene for a major QTL controlling seedling vigor in rice. Theor Appl Genet 125:647-657

Abou-Elwafa SF, Amein KA (2016) Genetic diversity and potential high temperature tolerance in barley (*Hordeum vulgare*). World J Agric Res 4:1-8

Abou-Elwafa SF, Shehzad T (2021) Genetic diversity, GWAS and prediction for drought and terminal heat stress tolerance in bread wheat (*Triticum aestivum* L.). Genet Resour Crop Evol 68:711-728

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

Al-Abdallat AM, Karadsheh A, Hadadd NI, Akash MW, Ceccarelli S, Baum M, Elenein JA (2017) Assessment of genetic diversity and yield performance in Jordanian barley (*Hordeum vulgare* L.) landraces grown under Rainfed conditions. BMC Plant Biol 17:1-13

Albrecht T, Oberforster M, Kempf H, Ramgraber L, Schacht J, Kazman E, Mohler V (2015) Genome-wide association mapping of preharvest sprouting resistance in a diversity panel of European winter wheats. J Appl Genet 56:277-285

Ali A, Cao J, Jiang H, Chang C, Zhang HP, Sheikh SW, Ma C (2019) Unraveling molecular and genetic studies of wheat (*Triticum aestivum* L.) resistance against factors causing pre-harvest sprouting. Agronomy 9:117

Alonso-Blanco C, Aarts MG, Bentsink L, Keurentjes JJ, Reymond M, Vreugdenhil D, Koornneef M (2009) What has natural variation taught us about plant development, physiology, and adaptation? Plant Cell 21:1877-1896

Alotaibi F, Al-Qthanin RN, Aljabri M, Albaqami M, Abou-Elwafa S (2021) Identification of genomic regions associated with agronomical traits of bread wheat under two levels of salinity using GWAS

Alqudah AM, Koppolu R, Wolde GM, Graner A, Schnurbusch T (2016) The genetic architecture of barley plant stature. Front Genet 7:117

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

Alqudah AM, Sharma R, Pasam RK, Graner A, Kilian B, Schnurbusch T (2014) Genetic dissection of photoperiod response based on GWAS of pre-anthesis phase duration in spring barley. PLoS One 9:e113120

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

Angessa TT, Li C (2016) Exploration and Utilization of Genetic Diversity Exotic Germplasm for Barley Improvement. In Exploration, Identification and Utilization of Barley Germplasm Academic Press, pp, 223-240

Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, Nordborg M (2005) Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. PLoS Genet 1:e60

Arif MR, Neumann K, Nagel M, Kobiljski B, Lohwasser U, Börner A (2012) An association mapping analysis of dormancy and pre-harvest sprouting in wheat. Euphytica 188:409-417

Aspinall D (1965) The effects of soil moisture stress on the growth of barley. II. Grain growth. Aust J Agric Res 16:265-275

Autio K, Simoinen T, Suortti T, Salmenkallio-Marttila M, Lassila K, Wilhelmson A (2001) Structural and enzymic changes in germinated barley and rye. J Inst Brew 107:19-25

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

Azevedo C, Santos-Rosa, MJ, Shirasu K (2001) The U-box protein family in plants. Trends Plant Sci 6:354-358

Badr A, Rabey HE, Effgen S, Ibrahim HH, Pozzi C, Rohde W, Salamini F (2000) On the origin and domestication history of barley (*Hordeum vulgare*). Mol Biol Evol 17:499-510

Baik BK, Ullrich SE (2008) Barley for food: Characteristics, improvement, and renewed interest. J Cereal Sci 48:233-242

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

Bayer MM, Rapazote-Flores P, Ganal M, Hedley PE, Macaulay M, Plieske J, Waugh R (2017) Development and evaluation of a barley 50k iSelect SNP array. Front Plant Sci 8:1792

Benech-Arnold RL (2001) Bases of pre-harvest sprouting resistance in barley: physiology, molecular biology and environmental control of dormancy in the barley grain. Barley Sci 481-502

Benech-Arnold RL, Gualano N, Leymarie J, Coˆme 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

Bengtsson T, Manninen O, Jahoor A, Orabi J (2017) Genetic diversity, population structure and linkage disequilibrium in Nordic spring barley (*Hordeum vulgare* L. subsp. *vulgare*). Genet Resour Crop Evol 64:2021-2033

Bentsink L, Jowett J, Hanhart CJ, Koornneef M (2006) Cloning of *DOG1*, a quantitative trait locus controlling seed dormancy in Arabidopsis. Proc Natl Acad Sci USA 103:17042-17047

Berg M, Rogers R, Muralla R, Meinke D (2005) Requirement of aminoacyl-tRNA synthetases for gametogenesis and embryo development in Arabidopsis. Plant J 44:866-878

Bergler J and Hoth S (2011) Plant U-box armadillo repeat proteins AtPUB18 and AtPUB19 are involved in salt inhibition of germination in Arabidopsis. Plant Biol 13:725-730

Bethke PC, Libourel IG, Jones RL (2006) Nitric oxide reduces seed dormancy in Arabidopsis. J Exp Bot 57: 517–26

Bewley JD (1997) Seed germination and dormancy. Plant cell 9:1055

Bewley JD, Black M (1994) Seeds: physiology of development and germination. 2nd edn New York: Plenum Press

Bewley JD, Black M (2013) Seeds: physiology of development and germination. Springer Science & Business Media

Biddulph TB, Plummer JA, Setter TL, Mares DJ (2008) Seasonal conditions influence dormancy and preharvest sprouting tolerance of wheat (*Triticum aestivum* L.) in the field. Field Crops Res 107:116-128

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

Bonnardeaux Y, Li C, Lance R, Zhang XQ, Sivasithamparam K, Appels R (2008) Seed dormancy in barley: identifying superior genotypes through incorporating epistatic interactions. Aust J Agric Res 59:517-526

Borrego-Benjumea A, Carter A, Zhu M, Tucker JR, Zhou M, Badea A (2021) Genome-Wide Association Study of Waterlogging Tolerance in Barley (*Hordeum vulgare* L.) Under Controlled Field Conditions. Front Plant Sci 12

Bothmer RV, Jacobsen N, Baden C, Jorgensen RB, Linde Laursen I (1991) An ecogeographical study of the genus Hordeum. International Board for Plant Genetic Resources

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

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

Busconi MARCO, Bosco CD, Crosatti CRISTINA, Baldi PAOLO, Marè CATERINA, Grossi M, Stanca A M (2001) The cold-regulated genes are involved in the physiological response of barley to cold environment. ICL Agri Sci 14:17-27

Canadian Grain Commission (CGC) https://www.grainscanada.gc.ca/en/

Cantalapiedra CP, Boudiar R, Casas AM, Igartua E, Contreras-Moreira B (2015) BARLEYMAP: physical and genetic mapping of nucleotide sequences and annotation of surrounding loci in barley. Mol Plant Breed 35:1-11

Cantoro R, Fernández LG, Cervigni GD, Rodriguez MV, Gieco JO, Paniego N, Benech-Arnold RL (2016) Seed dormancy QTL identification across a Sorghum bicolor segregating population. Euphytica 211:41-56

Carrari F, Benech-Arnold R, Osuna-Fernandez R, Hopp E, Sanchez R, Iusem N, Lijavetzky D (2003) Genetic mapping of the Sorghum bicolor vp1 gene and its relationship with preharvest sprouting resistance. Genome 46:253-258

Carvalho PDT, Beléia ADP (2019) Malting of preharvest sprouted wheat1. Rev Ciênc Agron 50:571-577

Cattivelli L, Ceccarelli S, Romagosa I, Stanca M (2010) Abiotic stresses in barley: problems and solutions. Barley, 282-306

Cavanagh C, Morell M, Mackay I, Powell W (2008) From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. Curr Opin Plant Biol 11:215-221

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

Clarke FR, Clarke JM, DePauw RM, Fernandez MR, Fox S, Gilbert J, Somers D (2005) Strategic approach to mitigating weather induced defects of wheat quality. Euphytica 143:285-290

Clarke JM, Pauw RMD, Christensen JV (1984) Effect of weathering on falling numbers of standing and windrowed wheat. Can J Plant Sci 64:457-463

Clarke JM, Pauw RMD, Christensen JV (1984) Effect of weathering on falling numbers of standing and windrowed wheat. Can J Plant Sci 64:457-463

Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 74:829-836

Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, Ramsay L, Waugh R (2009) Development and implementation of high-throughput SNP genotyping in barley. BMC Genomics 10:1-13

Cockram J, Mackay I (2018) Genetic mapping populations for conducting high-resolution trait mapping in plants. Plant Genet Mol Biol 109-138

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

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

Comadran J, Kilian B, Russell J, Ramsay L, Stein N, Ganal M, Waugh R (2012) Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. Nat Genet 44:1388-1392

Comadran J, Thomas WTB, Van Eeuwijk FA, Ceccarelli S, Grando S, Stanca AM, Russell JR (2009) Patterns of genetic diversity and linkage disequilibrium in a highly structured Hordeum vulgare association-mapping population for the Mediterranean basin. Theor Appl Genet 119:175-187

Copeland LO, McDonald MB (1999) Seed dormancy. In Principles of seed science and technology Springer, Boston, MA, pp, 127-152

Corbineau F, Côme D (1996) Barley seed dormancy. Boissons Bios Conditionnement 261:113-119

Dai F, Nevo E, Wu D, Comadran J, Zhou M, Qiu L, Zhang G (2012) Tibet is one of the centers of domestication of cultivated barley. Proc Natl Acad Sci USA 109:16969-16973

Dai F, Wang X, Zhang XQ, Chen Z, Nevo E, Jin G, Zhang G (2018) Assembly and analysis of a qingke reference genome demonstrate its close genetic relation to modern cultivated barley. Plant Biotechnol J 16:760-770

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499-510

Debeaujon I, Lepiniec L, Pourcel L, Routaboul JM (2007) Seed coat development and dormancy. Seed development, dormancy and germination 27:392

DePauw RM, Knox RE, Singh AK, Fox SL, Humphreys DG, Hucl P (2012) Developing standardized methods for breeding preharvest sprouting resistant wheat, challenges and successes in Canadian wheat. Euphytica 188:7-14

Dhingani RM, Umrania VV, Tomar RS, Parakhia MV, Golakiya B (2015) Introduction to QTL mapping in plants. Ann Plant Sci 4:1072-1079

Dóczi R, Brader G, Pettkó-Szandtner A, Rajh I, Djamei A, Pitzschke A, Hirt H (2007) The Arabidopsis mitogen-activated protein kinase kinase MKK3 is upstream of group C mitogen-activated protein kinases and participates in pathogen signaling. Plant Cell 19:3266-3279

Druka A, Kudrna D, Han F, Kilian A, Steffenson B, Frisch D, Kleinhofs A (2000) Physical mapping of the barley stem rust resistance gene rpg4. Mol Gen Genet 264:283-290

Duff SM, Rydel TJ, McClerren AL, Zhang W, Li JY, Sturman EJ, Evdokimov A (2012) The enzymology of alanine aminotransferase (AlaAT) isoforms from Hordeum vulgare and other organisms, and the HvAlaAT crystal structure. Arch Biochem Biophys 528:90-101

Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359-361

Eriani G, Delarue M, Poch O, Gangloff J, Moras D (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature 347:203-206

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

Fan Y, Zhou G, Shabala S, Chen ZH, Cai S, Li C, Zhou M (2016) Genome-wide association study reveals a new QTL for salinity tolerance in barley (*Hordeum vulgare* L.). Front Plant Sci 7:946

Fang J, Chu C (2008) Abscisic acid and the pre-harvest sprouting in cereals. Plant Signal Behav 3:1046-1048

Fastnaught CE (2001) Barley fiber. In Handbook of dietary fiber CRC Press, pp, 519-542

Feldman M (2000) Origin of cultivated wheat. In: Alin, P.B. and William J.A. (eds.) The World Wheat Book: A Hstory of Wheat Breeding, TEC and DOC, London, pp, 3-56

Fenner M (1991) The effects of the parent environment on seed germinability. Seed Sci Res 1:75-84

Finch-Savage WE, Footitt S (2017) Seed dormancy cycling and the regulation of dormancy mechanisms to time germination in variable field environments. J Exp Bot 68:843-856

Finch-Savage WE, Leubner-Metzger G (2006) Seed dormancy and the control of germination. New Phytol 171:501-523

Finkelstein R, Gampala S, Rock C (2002) Abscisic acid signaling in seeds and seedlings. Plant Cell 14:S15-S45

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

Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, Mitchell SE, Buckler ES (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 44:1054-1064

Food and Agriculture Organization (FAO) https://www.fao.org/statistics/en/

Forster BP, Ellis RP, Moir J, Talame V, Sanguineti MC, Tuberosa R, Ben Salem M (2004) Genotype and phenotype associations with drought tolerance in barley tested in North Africa. Ann Appl Biol 144:157-168

Gao F, Ayele BT (2014) Functional genomics of seed dormancy 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 W, Clancy JA, Han F, Prada D, Kleinhofs A, Ullrich SE (2003) Molecular dissection of a dormancy QTL region near the chromosome 7 (5H) L telomere in barley. Theor Appl Genet 107:552-559

Gao X, Hu C, Li H, Yao Y, Meng M, Dong J, Zhao W, Chen Q, Li X (2013) Factors affecting pre-harvest sprouting resistance in wheat (*Triticum aestivum*). J Anim Plant Sci 23:556-565

Gaut BS, Long AD (2003) The lowdown on linkage disequilibrium. Plant Cell 15:1502-1506

Giese H, Holm-Jensen AG, Mathiassen H, Kjaer B, Rasmussen SK, Bay H, Jensen J (1994) Distribution of RAPD markers on a linkage map of barley. Hereditas 120:267-273

Gireesh C, Sundaram RM, Anantha SM, Pandey MK, Madhav MS, Rathod S, Rakshit S (2021) Nested Association Mapping (NAM) Populations: Present Status and Future Prospects in the Genomics Era. Crit Rev Plant Sci 40:49-67

Gordon IL (1980) Germinability, dormancy and grain development. Cereal Res Commun 115-129

Gualano N, Carrari F, Rodríguez MV, Pérez-Flores L, Sánchez R, Iusem N, Benech-Arnold R (2007) Reduced embryo sensitivity to abscisic acid in a sprouting-susceptible sorghum (*Sorghum bicolor*) variety is associated with altered ABA signalling. Seed Sci Res 17:81-90

Gualano NA, Benech-Arnold RL (2009) The effect of water and nitrogen availability during grain filling on the timing of dormancy release in malting barley crops. Euphytica 168:291-301

Guan X, Okazaki Y, Lithio A, Li L, Zhao X, Jin H, Nikolau BJ (2017) Discovery and characterization of the 3-hydroxyacyl-ACP dehydratase component of the plant mitochondrial fatty acid synthase system. Plant Physiol 173:2010-2028

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: 886-896

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

Gudys K, Guzy-Wrobelska J, Janiak A, Dziurka MA, Ostrowska A, Hura K, Szarejko I (2018) Prioritization of candidate genes in QTL regions for physiological and biochemical traits underlying drought response in barley (*Hordeum vulgare* L.). Front Plant Sci 9:769

Guglielminetti L, Yamaguchi J, Perata P, Alpi A (1995) Amylolytic activities in cereal seeds under aerobic and anaerobic conditions. Plant Physiol 109:1069-1076

Gupta M, Abu-Ghannam N, Gallaghar E (2010) Barley for brewing: Characteristic changes during malting, brewing and applications of its by-products. Compr Rev Food Sci Food Saf 9:318-328

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

Gutiérrez L, Berberian N, Capettini F, Falcioni E, Fros D, Germán S, Castro A (2013) Genome-wide association mapping identifies disease-resistance QTLs in barley germplasm from Latin America. In Advance in Barley Sciences Springer, Dordrecht, pp, 209-215

Haas M, Schreiber M, Mascher M (2019) Domestication and crop evolution of wheat and barley: Genes, genomics, and future directions. J Integr Plant Biol 61:204-225

Hamblin MT, Close T J, Bhat PR, Chao S, Kling JG, Abraham KJ, Jannink JL (2010) Population structure and linkage disequilibrium in US barley germplasm: implications for association mapping. Crop Sci 50:556-566

Han F, Kleinhofs A, Ullrich SE, Kilian A, Yano M, Sasaki T (1998) Synteny with rice: analysis of barley malting quality QTLs and rpg4 chromosome regions. Genome 41:373-380

Han F, Ullrich SE, Clancy JA, Jitkov V, Kilian A, Romagosa I (1996) Verification of barley seed dormancy loci via linked molecular markers. Theor Appl Genet 92:87-91

Harwood WA (2019) An introduction to barley: the crop and the model. In Barley Humana Press, New York, NY, pp, 1-5

Hassan AS, Houston K, Lahnstein J, Shirley N, Schwerdt JG, Gidley MJ, Burton RA (2017) A genome wide association study of arabinoxylan content in 2-row spring barley grain. PloS One 12: e0182537

Hayes B (2013) Overview of statistical methods for genome-wide association studies (GWAS). Genome-wide association studies and genomic prediction, pp, 149-169

Hayes PM, Costro A, Marquez-Cedillo L, Corey A, Henson CA, Jones BL, Kling J (2003) Genetic diversity for quantitatively inherited agronomic and malting quality traits. In: Von Bothmer R, Van Hintum T, Knupffer H, Sato K (eds) Diversity in barley (*Hordeum vulgare*), Elsevier Press, New York, pp 147–169

Hedrick PW (1987) Gametic disequilibrium measures: proceed with caution. Genetics 117:331-341

Hedrick PW (1987) Genetics 117:331-341

Hess KT, Schmidt KD, Zocher M (2006) Multivariate loss prediction in the multivariate additive model. Insur Math Econ 39:185-191

Hickey LT, Lawson W, Arief VN, Fox G, Franckowiak J, Dieters MJ (2012) Grain dormancy QTL identified in a doubled haploid barley population derived from two non-dormant parents. Euphytica 188:113-122

Hilhorst HW (1995) A critical update on seed dormancy. I. Primary dormancy1. Seed Sci Res 5:61-73

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

Hoang HH, Bailly C, Corbineau F, Leymarie J (2013) Induction of secondary dormancy by hypoxia in barley grains and its hormonal regulation. J Exp Bot 64:2017-2025

Hoang HH, Sechet J, Bailly C, Leymarie J, Corbineau F (2014) Inhibition of germination of dormant barley (*Hordeum vulgare* L.) grains by blue light as related to oxygen and hormonal regulation. Plant Cell Environ 37:1393-1403

Holdsworth M, Kurup S, McKibbin R (1999) Molecular and genetic mechanisms regulating the transition from embryo development to germination. Trends Plant Sci 4:275-280

Holland JB (2007) Genetic architecture of complex traits in plants. Curr Opin Plant Biol 10:156-161

Hu J, Mitchum MG, Barnaby N, Ayele BT, Ogawa M, Nam E, Swain SM (2008) Potential sites of bioactive gibberellin production during reproductive growth in Arabidopsis. Plant Cell 20: 320-336

Hübner S, Höffken M, Oren E, Haseneyer G, Stein N, Graner A, Fridman E (2009) Strong correlation of wild barley (*Hordeum spontaneum*) population structure with temperature and precipitation variation. Mol Ecol 18:1523-1536

Idehen E, Tang Y, Sang S (2017) Bioactive phytochemicals in barley. J Food Drug Anal 25:148-161

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

Ishikawa S, Barrero JM, Takahashi F, Nakagami H, Peck SC, Gubler F, Umezawa T (2019) Comparative phosphoproteomic analysis reveals a decay of ABA signaling in barley embryos during after-ripening. Plant Cell Physio 60:2758-2768

Izydorczyk C, Nguyen T-N, Jo S, Son S, Tuan PA, Ayele BT (2018) Spatiotemporal modulation of abscisic acid and gibberellin metabolism and signalling mediates the effects of suboptimal and supraoptimal temperatures on seed germination in wheat (*Triticum aestivum* L.). Plant Cell Environ 41:1022-1037

Jabbari M, Fakheri BA, Aghnoum R, Mahdi Nezhad N, Ataei R (2018) GWAS analysis in spring barley (*Hordeum vulgare* L.) for morphological traits exposed to drought. PLoS One 13:e0204952

Jacobsen JV, Barrero JM, Hughes T, Julkowska M, Taylor JM, Xu Q, Gubler F (2013) Roles for blue light, jasmonate and nitric oxide in the regulation of dormancy and germination in wheat grain (*Triticum aestivum* L.). Planta 238:121-138

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

Jamil M, Ali A, Akbar KF, Napar AA, Gul A, Mujeeb-Kazi A (2016) QTL analysis in plants: ancient and modern perspectives. In Plant Omics: Trends and Applications, Springer, Cham, pp, 59-82

Jehan T, Lakhanpaul S (2006) Single nucleotide polymorphism (SNP)–methods and applications in plant genetics: a review

Jemmat AM, Ranocha P, Le Ru A, Neel M, Jauneau A, Raggi S, Dunand C (2020) Coordination of five class III peroxidase-encoding genes for early germination events of *Arabidopsis thaliana*. Plant Sci 298:110565

Kader MA (2005) A comparison of seed germination calculation formulae and the associated interpretation of resulting data. J Proc R Soc N S W 138:65-75

Kaneko M, Itoh H, Ueguchi-Tanaka M, Ashikari M, Matsuoka M (2002) The α-amylase induction in endosperm during rice seed germination is caused by gibberellin synthesized in epithelium. Plant Physiol 128:1264-1270

Kapazoglou A, Tondelli A, Papaefthimiou D, Ampatzidou H, Francia E, Stanca MA, Tsaftaris AS (2010) Epigenetic chromatin modifiers in barley: IV. The study of barley polycomb group (PcG) genes during seed development and in response to external ABA. BMC Plant Biol 10:1-19

Kashiwakura Y-i, Kobayashi D, Jikumaru Y, Takebayashi Y, Nambara E, Seo M, 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

Kilian B, Martin W, Salamini F (2010) Genetic diversity, evolution and domestication of wheat and barley in the Fertile Crescent. Evolution in action, pp, 137-166

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Nordborg M (2007) Recombination and linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet 39:1151-1155

King RW (1984) Water uptake in relation to pre-harvest sprouting damage in wheat: grain characteristics. Aust J Agric Res 35:337-345

King RW, Richards RA (1984) Water uptake in relation to pre-harvest sprouting damage in wheat: ear characteristics. Aust J Agric Res 35:327-336

King RW, von Wettstein-Knowles P (2000) Epicuticular waxes and regulation of ear wetting and pre-harvest sprouting in barley and wheat. Euphytica 112:157-166

Klein RJ (2007) Power analysis for genome-wide association studies. BMC Genet 8:58

Kleinhofs A, Kilian A, Maroof MS, Biyashev RM, Hayes P, Chen FQ, Steffenson BJ (1993) A molecular, isozyme and morphological map of the barley (Hordeum vulgare) genome. Theor Appl Genet 86:705-712

Kling JG, Hayes PM, Ullrich SE (2004) BARLEY. Genetics and Breeding

Koike M, Takezawa D, Arakawa K, Yoshida S (1997) Accumulation of 19-kDa plasma membrane polypeptide during induction of freezing tolerance in wheat suspension-cultured cells by abscisic acid. Plant Cell Physiol 38:707-716

Komatsuda T, Pourkheirandish M, He C, Azhaguvel P, Kanamori H, Perovic D, Yano M (2007) Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. Proc Natl Acad Sci USA 104:1424-1429

Koppolu R, Anwar N, Sakuma S, Tagiri A, Lundqvist U, Pourkheirandish M, Schnurbusch T (2013) *Six-rowed spike4* (*Vrs4*) controls spikelet determinacy and row-type in barley Proc Natl Acad Sci USA 110:13198-13203

Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9:1-9

Kraakman AT, Niks RE, Van den Berg PM, Stam P, Van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435-446

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. Theor 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. Theor Appl Genet 111:1052-1059

Kumar J, Pratap A, Solanki RK, Gupta DS, Goyal A, Chaturvedi SK, Kumar S (2012) Genomic resources for improving food legume crops. J Agric Sci 150:289-318

Kuntz RJ, Bamforth CW (2007) Time Course for the Development of Enzymes in Barley 1. J Inst Brew 113:196-205

Kushiro T, Okamoto M, Nakabayashi K, Yamagishi K, Kitamura S, Asami T, Nambara E (2004) The Arabidopsis cytochrome P450 CYP707A encodes ABA 8′-hydroxylases: key enzymes in ABA catabolism. EMBO J 23:1647-1656

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185-199

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181

Langridge P (2018) Economic and academic importance of barley. In The barley genome Springer, Cham, pp, 1-10

Lee JS, Chebotarov D, McNally KL, Pede V, Setiyono TD, Raquid R, Mo Y (2021) Novel Sources of Pre-Harvest Sprouting Resistance for Japonica Rice Improvement. Plants 10:1709

Lefebvre V, North H, Frey A, Sotta B, Seo M, Okamoto M, 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

Leymarie J, Robayo-Romero ME, Gendreau E, Benech-Arnold RL, Corbineau F (2008) Involvement of ABA in induction of secondary dormancy in barley (*Hordeum vulgare* L.) seeds. Plant Cell Physiol 49:1830-1838

Li C, Ni P, Francki M, Hunter A, Zhang Y, Schibeci D, Appels R (2004) Genes controlling seed dormancy and pre-harvest sprouting in a rice-wheat-barley comparison. Funct Integr Genomics 4:84-93

Li CD, Tarr A, Lance RCM, Harasymow S, Uhlmann J, Westcot S, Appels R (2003) A major QTL controlling seed dormancy and pre-harvest sprouting/grain α-amylase in two-rowed barley (*Hordeum vulgare L.*). Aust J Agric Res 54:1303-1313

Li CD, Tarr A, Lance RCM, Harasymow S, Uhlmann J, Westcot S, Appels R (2003) A major QTL controlling seed dormancy and pre-harvest sprouting/grain α-amylase in two-rowed barley (*Hordeum vulgare* L). Aust J Agric Res 54:1303-1313

Lin M, Zhang D, Liu S, Zhang G, Yu J, Fritz AK, Bai G (2016) Genome-wide association analysis on pre-harvest sprouting resistance and grain color in US winter wheat. BMC Genomics 17:1-16

Lin PC, Hwang SG, Endo A, Okamoto M, Koshiba T, Cheng WH (2007) Ectopic expression of *ABSCISIC ACID 2/GLUCOSE INSENSITIVE 1* in Arabidopsis promotes seed dormancy and stress tolerance. Plant Physiol 143:745-758

Lin R, Horsley RD, Schwarz PB (2008) Associations between caryopsis dormancy, α-amylase activity, and pre-harvest sprouting in barley. J Cereal Sci 48:446-456

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

Linkies A, Schuster-Sherpa U, Tintelnot S, Leubner-Metzger G, Müller K (2010) Peroxidases identified in a subtractive cDNA library approach show tissue-specific transcript abundance and enzyme activity during seed germination of *Lepidium sativum*. J Exp Bot 61:491-502

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Zhang Z (2012) GAPIT: genome association and prediction integrated tool. J Bioinform 28:2397-2399

Liu A, Gao F, Kanno Y, Jordan MC, Kamiya Y, Seo M, Ayele BT (2013) Regulation of wheat seed dormancy by after-ripening is mediated by specific transcriptional switches that induce changes in seed hormone metabolism and signaling. PLoS One 8:e56570

Liu C, Ding F, Hao F, Yu M, Lei H, Wu X, Tang H (2016) Reprogramming of seed metabolism facilitates pre-harvest sprouting resistance of wheat. Sci rep 6:1-11

Liu M, Li Y, Ma Y, Zhao Q, Stiller J, Feng Q, Liu C (2020) The draft genome of a wild barley genotype reveals its enrichment in genes related to biotic and abiotic stresses compared to cultivated barley. Plant Biotechnol J 18:443-456

Liu M, Li Y, Ma Y, Zhao Q, Stiller J, Feng Q, Liu C (2020) The draft genome of a wild barley genotype reveals its enrichment in genes related to biotic and abiotic stresses compared to cultivated barley. Plant Biotechnol J 18:443-456

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2 (_Delta Delta C (T)) method. Methods 25:402-408

Lohwasser U, Arif MR, Börner A (2013) Discovery of loci determining pre-harvest sprouting and dormancy in wheat and barley applying segregation and association mapping. Biol Plant 57:663-674

Lorieux M (2012) MapDisto: fast and efficient computation of genetic linkage maps. Mol Plant Breed 30:1231-1235

Lu Q, Niu X, Zhang M, Wang C, Xu Q, Feng Y, Liang X (2018) Genome-wide association study of seed dormancy and the genomic consequences of improvement footprints in rice (*Oryza sativa* L.). Front Plant Sci 8:2213

Magwa RA, Zhao H, Xing Y (2016) Genome-wide association mapping revealed a diverse genetic basis of seed dormancy across subpopulations in rice (*Oryza sativa* L.). BMC Genet 17:1-13

Mahmood K (2011) Salinity tolerance in barley (*Hordeum vulgare* L.): effects of varying NaCl, K?/Na? and NaHCO3 levels on cultivars differing in tolerance. Pak J Bot 43:1651-1654

Malysheva-Otto LV, Ganal MW, Röder MS (2006) Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare L.*). BMC Genet 7:1-14

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

Mao X, Zhang J, Liu W, Yan S, Liu Q, Fu H, Wang F (2019) The MKKK62-MKK3-MAPK7/14 module negatively regulates seed dormancy in rice. Rice 12:1-14

Mares D (2019) Fourth International symposium on pre-harvest sprouting in cereals. CRC Press

Martinez SA, Godoy J, Huang M, Zhang Z, Carter AH, Garland Campbell KA, Steber CM (2018) Genome-wide association mapping for tolerance to preharvest sprouting and low falling numbers in wheat. Front Plant Sci 9:141

Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, Stein N (2017) A chromosome conformation capture ordered sequence of the barley genome. Nature 544:427-433

Mascher M, Muehlbauer GJ, Rokhsar DS, Chapman J, Schmutz J, Barry K, Waugh R (2013) Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). Plant J 76:718-727

Mascher M, Wicker T, Jenkins J, Plott C, Lux T, Koh CS, Stein N (2021) Long-read sequence assembly: a technical evaluation in barley. Plant cell koab077

Masojć P, Milczarski P (2009) Relationship between QTLs for preharvest sprouting and alpha-amylase activity in rye grain. Molecular Breeding 23:75-84

Masoud S, Saoub H, Migdadi H, Al-Nashash A (2005) Evaluation of Jordanian barley (*Hordeum vulgare* L.) landraces collected from diverse environments. Dirasat Agric Sci 32:163-170

Matthies IE, Van Hintum T, Weise S, Röder MS (2012) Population structure revealed by different marker types (SSR or DArT) has an impact on the results of genome-wide association mapping in European barley cultivars. Mol Plant Breed 30:951-966

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

Maurer A, Draba V, Pillen K (2016) Genomic dissection of plant development and its impact on thousand grain weight in barley through nested association mapping. J Exp Bot 67:2507-2518

Mayer KF, Waugh R, Langridge P, Close TJ, Wise RP, Graner A, Stein N (2012) A physical, genetic and functional sequence assembly of the barley genome, pp, 711-716

Melzer JM, Kleinhofs A, Kudrna DA, Warner RL, Blake TK (1988) Genetic mapping of the barley nitrate reductase-deficient nar1 and nar2 loci. Theor Appl Genet 75:767-771

Mester D, Ronin Y, Minkov D, Nevo E, Korol A (2003) Constructing large-scale genetic maps using an evolutionary strategy algorithm. Genetics 165:2269-2282

Mikołajczak K, Ogrodowicz P, Gudyś K, Krystkowiak K, Sawikowska A, Frohmberg W, Kuczyńska A (2016) Quantitative trait loci for yield and yield-related traits in spring barley populations derived from crosses between European and Syrian cultivars. PLoS One 11: e0155938

Milner SG, Jost M, Taketa S, Mazón ER, Himmelbach A, Oppermann M, Stein N (2019) Genebank genomics highlights the diversity of a global barley collection. Nat Genet 51:319-326

Minaiyan M, Ghannadi A, Movahedian A, Hakim-Elahi I (2014) Effect of *Hordeum vulgare* L.(Barley) on blood glucose levels of normal and STZ-induced diabetic rats. Res Pharm Sci 9:173

Mizuno Y, Yamanouchi U, Hoshino T, Nonoue Y, Nagata K, Fukuoka S, Sugimoto K (2018) Genetic dissection of pre-harvest sprouting resistance in an upland rice cultivar. Breed Sci 68:200-209

Mohammadi M, Blake TK, Budde AD, Chao S, Hayes PM, Horsley RD, Smith KP (2015) A genome-wide association study of malting quality across eight US barley breeding programs. Theor Appl Genet 128:705-721

Molina-Cano JL, Igartua E, Casas AM, Moralejo M (2002) New views on the origin of cultivated barley. Barley science: Recent advances from molecular biology to agronomy of yield and quality, pp, 15-29

Monat C, Padmarasu S, Lux T, Wicker T, Gundlach H, Himmelbach A, Mascher M (2019b) TRITEX: chromosome-scale sequence assembly of Triticeae genomes with open-source tools. Genome Biol 20:1-18

Monat C, Schreiber M, Stein N, Mascher M (2019a) Prospects of pan-genomics in barley. Theor Appl Genet 132:785-796

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

Mukherjee S, Stasolla C, Ayele BT (2015). RNA In situ Hybridization. In Plant Microtechniques and Protocols, Springer, Cham pp, 351-370

Mustroph A, Lee SC, Oosumi T, Zanetti ME, Yang H, Ma K, Bailey-Serres J (2010) Cross-kingdom comparison of transcriptomic adjustments to low-oxygen stress highlights conserved and plant-specific responses. Plant Physiol 152:1484-1500

Mwando E, Han Y, Angessa TT, Zhou G, Hill CB, Zhang XQ, LiC (2020) Genome-wide association study of salinity tolerance during germination in barley (*Hordeum vulgare* L.). Front Plant Sci 11:118

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

Nagel M, Alqudah AM, Bailly M, Rajjou L, Pistrick S, Matzig, G, Kranner I (2019) Novel loci and a role for nitric oxide for seed dormancy and preharvest sprouting in barley. Plant Cell Environ 42:1318-1327

Nakamura S (2018) Grain dormancy genes responsible for preventing pre-harvest sprouting in barley and wheat. Breed Sci 17138

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

Nakamura S, Pourkheirandish M, Morishige H, Kubo Y, Nakamura M, Ichimura K, Komatsuda T (2016) *Mitogen-activated protein kinase kinase 3* regulates seed dormancy in barley. Curr Biol 26:775-781

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

Naredo MEB, Juliano AB, Lu BR, De Guzman F, Jackson MT (1998) Responses to seed dormancy-breaking treatments in rice species (*Oryza* L.). Seed Sci Technol 26:675-690

Narsai R, Law SR, Carrie C, Xu L, Whelan J (2011) In-depth temporal transcriptome profiling reveals a crucial developmental switch with roles for RNA processing and organelle metabolism that are essential for germination in Arabidopsis. Plant Physiol 157:1342-1362

Newton AC, Flavell AJ, George TS, Leat P, Mullholland B, Ramsay L, Bingham IJ (2011) Crops that feed the world 4. Barley: a resilient crop? Strengths and weaknesses in the context of food security. Food Secur 3:141-178

Nguyen TN, Son S, Jordan MC, Levin DB, Ayele BT (2016) Lignin biosynthesis in wheat (*Triticum aestivum* L.): its response to waterlogging and association with hormonal levels. BMC Plant Biol 16:1-16

Nonogaki H, Barrero JM, 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, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Weigel D (2002) The extent of linkage disequilibrium in Arabidopsis thaliana. Nat Genet 30:190-193

Oberthur L, Blake TK, Dyer WE, Ullrich SE (1995) Genetic analysis of seed dormancy in barley (*Hordeum vulgare* L.). J Quant Trait Loci 1

Okamoto M, Kuwahara A, Seo M, Kushiro T, Asami T, Hirai N, Kamiya Y, Koshiba T, Nambara E (2006) *CYP707A1* and *CYP707A2*, which encode ABA 8' hydroxylases, are indispensable for a proper control of seed dormancy and germination in Arabidopsis. Plant Physiol 141:97-107

Olaerts H, Roye C, Derde LJ, Sinnaeve G, Meza WR, Bodson B, Courtin CM (2016) Evolution and distribution of hydrolytic enzyme activities during preharvest sprouting of wheat (*Triticum aestivum*) in the field. J Agric Food Chem 64:5644-5652

Olafson K (2018) http://gobarley.com/what-does-malt-taste-like/

Orabi J, Backes G, Wolday A, Yahyaoui A, Jahoor A (2007) The Horn of Africa as a centre of barley diversification and a potential domestication site. Theor Appl Genet 114:1117-1127

Park J-H, Wacholder S, Gail MH, Peters U, Jacobs KB, Chanock SJ, Chatterjee N (2010) Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nat Genet 42:570-575

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 (1990) Inheritance of Grain Dormacy in White-Kernelled Wheat. Crop sci 30:25-30

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

Patwa N, Penning BW (2020) Environmental Impact on Cereal Crop Grain Damage from Pre-harvest Sprouting and Late Maturity Alpha-Amylase. In Sustainable Agriculture in the Era of Climate Change Springer, Cham, pp, 23-41

Paulitz TC, Steffenson BJ (2011) Biotic stress in barley: disease problems and solutions. Barley production, improvement, and uses, pp, 307-354

Payne RW (2009) GenStat. Wiley Interdisciplinary Reviews: Comput Stat 1:255-258

Pfeiffer RM, Gail MH (2003) Sample size calculations for population- and family-based casecontrol association studies on marker genotypes. Genet Epidemiol 25:136

Pham AT, Maurer A, Pillen K, Brien C, Dowling K, Berger B, March TJ (2019) Genome-wide association of barley plant growth under drought stress using a nested association mapping population. BMC Plant Biol 19:1-16

Pourkheirandish M, Hensel G, Kilian B, Senthil N, Chen G, Sameri M, Komatsuda T (2015) Evolution of the grain dispersal system in barley. Cell 162:527-539

Prada D, Ullrich SE, Molina-Cano JL, Cistue L, Clancy JA, Romagosa I (2004) Genetic control of dormancy in a Triumph/Morex cross in barley. Theor Appl Genet 109:62-70

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 Genet 38:904-909

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945-959

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559-575

Qi X, Stam P, Lindhout P (1998) Use of locus specific AFLP markers to construct a high-density molecular map in barley. Theor Appl Genet 96:376-384

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

Rakshit S, Rakshit A, Patil JV (2012) Multiparent intercross populations in analysis of quantitative traits. J Genet 91:111-117

Ranford JC, Bryce JH, Morris PC (2002) PM19, a barley (*Hordeum vulgare* L.) gene encoding a putative plasma membrane protein, is expressed during embryo development and dormancy. J Exp Bot 53:147-148

Reiner L, Loch V (1976) Forecasting dormancy in barley-ten years experience. Cereal Res Commun 107-110

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

Risch N (1992) Genetic linkage: Interpreting LOD scores. Science 255:803-805

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516-1517

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 25:99-119

Rodríguez MV, Margineda M, González-Martín JF, Insausti P, Benech-Arnold RL (2001) Predicting preharvest sprouting susceptibility in barley: a model based on temperature during grain filling. J Agron 93:1071-1079

Rollefson G, Simmons A, Donaldson M, Gillespie W, Kafafi Z, Kohler-Rollefson I, McAdam E, Ralston S, Tubb K (1985) Excavations at the pre-pottery Neolithic B village of 'Ain Ghazal (Jordan), 1983. MDOG 117:69-116

Romagosa I, Han F, Clancy JA, Ullrich SE (1999) Individual locus effects on dormancy during seed development and after ripening in barley. Crop Sci 39:74-79

Rooney TE, Sweeney DW, Sorrells ME (2021) Time series barley germination is predictable using functional principal component analysis or logistic regression and associated with known seed dormancy loci. Crop Sci:doi.org/10.1002/csc2.20638

Saade S, Kutlu B, Draba V, Förster K, Schumann E, Tester M, Maurer A (2017) A donor-specific QTL, exhibiting allelic variation for leaf sheath hairiness in a nested association mapping population, is located on barley chromosome 4H. PloS One 12:e0189446

Samarah NH (2005) Effects of drought stress on growth and yield of barley. Agron Sustain Dev 25:145-149

Samuel MA, Mudgil Y, Salt JN, Delmas F, Ramachandran S, Chilelli A, Goring DR (2008) Interactions between the S-domain receptor kinases and AtPUB-ARM E3 ubiquitin ligases suggest a conserved signaling pathway in Arabidopsis. Plant Physiol 147:2084-2095

Sannemann W, Huang BE, Mathew B, Léon 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 Plant Breed 35:1-16

Sato K, Matsumoto T, Ooe N, Takeda K (2009) Genetic analysis of seed dormancy QTL in barley. Breed Sci 59:645-650

Sato K, Tanaka T, Shigenobu S, Motoi Y, Wu J, Itoh T (2016) Improvement of barley genome annotations by deciphering the Haruna Nijo genome. DNA Res 23:21-28

Sato K, Yamane M, Yamaji N, Kanamori H, Tagiri A, Schwerdt JG, Komatsuda T (2016) *Alanine aminotransferase* controls seed dormancy in barley. Nat Commun 7:1-9

Savin R, Stone PJ, Nicolas ME, Wardlaw IF (1997) Grain growth and malting quality of barley. 1. Effects of heat stress and moderately high temperature. Aust J Agric Res 48:615-624

Schmitz-Linneweber C, Small I (2008) Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci 13:663-670

Schreiber M, Mascher M, Wright J, Padmarasu S, Himmelbach A, Heavens D, Waugh R (2020) A genome assembly of the barley ‘transformation reference’cultivar Golden Promise. G3: Genes Genom Genet 10:1823-1827

Schwarz P, Horsley R, McNamara H (2004) Preharvest sprouting in the 2002 Midwestern barley crop: Occurrence and assessment of methodology. J Am Soc Brew Chem 62:147-154

Semagn K, Bjørnstad Å, Xu Y (2010) The genetic dissection of quantitative traits in crops. Electron J Biotechnol 13:16-17

Seo DH, Ryu MY, Jammes F, Hwang JH, Turek M, Kang BG, Kim WT (2012) Roles of four Arabidopsis U-box E3 ubiquitin ligases in negative regulation of abscisic acid-mediated drought stress responses. Plant Physiol 160:556-568

Seo JS, Zhao P, Jung C, Chua NH (2019) PLANT U-BOX PROTEIN 10 negatively regulates abscisic acid response in Arabidopsis. Appl Biol Chem 62:1-5

Sethi V, Raghuram B, Sinha AK, Chattopadhyay S (2014) A mitogen-activated protein kinase cascade module, MKK3-MPK6 and MYC2, is involved in blue light-mediated seedling development in Arabidopsis. Plant Cell 26:3343-3357

Setter TL, Waters I (2003) Review of prospects for germplasm improvement for waterlogging tolerance in wheat, barley and oats. Plant Soil 253:1-34

Shah SJA, Bashir M, Manzoor N (2012) A review on barley yellow dwarf virus. In Crop Production for Agricultural Improvement Springer, Dordrecht, pp, 747-782

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

Shaveta S, Kaur H, Kaur S (2019) Hulless Barley: A new era of research for food purposes. J Cereal Res 11

Shi J, Shi J, Liang W, Zhang D (2021) Integrating GWAS and Transcriptomics to Identify Genes Involved in Seed Dormancy in Rice.

Shin JS, Chao S, Corpuz L, Blake T (1990) A partial map of the barley genome incorporating restriction fragment length polymorphism, polymerase chain reaction, isozyme, and morphological marker loci. Genome 33:803-810

Shirley BW (1998) Flavonoids in seeds and grains: physiological function, agronomic importance and the genetics of biosynthesis. Seed Sci Res 8:415-422

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

Singh SR, Dalal S, Singh R, Dhawan AK, Kalia RK (2013) Evaluation of genetic fidelity of in vitro raised plants of Dendrocalamus asper (Schult. & Schult. F.) Backer ex K. Heyne using DNA-based markers. Acta Physiol Plant 35:419-430

Sohn SI, Pandian S, Kumar TS, Zoclanclounon YAB, Muthuramalingam P, Shilpha J, Ramesh M (2021) Seed Dormancy and Pre-Harvest Sprouting in Rice—An Updated Overview. Int J Mol Sci 22:11804

Soller M, Brody T, Genizi A (1976) On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor Appl Genet 47:35-39

Spencer CC, Su Z, Donnelly P, Marchini J (2009) Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. PLoS Genet 5:e1000477

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: Join Map. Plant J 3:739-744

Stein N, Prasad M, Scholz U, Thiel T, Zhang H, Wolf M, Graner A (2007) A 1,000-loci transcript map of the barley genome: new anchoring points for integrative grass genomics. Theor Appl Genet 114:823-839

Stich B, Melchinger AE (2010) An introduction to association mapping in plants. CAB Rev 5:1-9

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. Proc Natl Acad Sci USA 100:9440-9445

Strand E (1980) A seed dormancy index for selection of cultivars of cereals resistant to pre-harvest sprouting. Cereal Res Commun 8:175

Sugimoto K, Marzougi S, Yano M (2009) Genetic Control of Seed Dormancy in Rice. In Report of Symposium, pp, 53

Sugimoto K, Takeuchi Y, Ebana K, Miyao A, Hirochika H, Hara N, Yano M (2010) Molecular cloning of *Sdr4*, a regulator involved in seed dormancy and domestication of rice. Proc Natl Acad Sci USA 107:5792-5797

Suzuki Y, Ito S, Otsuka K, Iwasawa E, Nakajima M, Yamaguchi I (2005) Preparation of functional single-chain antibodies against bioactive gibberellins by utilizing randomly mutagenized phage-display libraries. Biosci Biotechnol Biochem 69:610-619

Takahashi F, Yoshida R, Ichimura K, Mizoguchi T, Seo S, Yonezawa M, Shinozaki K (2007) The mitogen-activated protein kinase cascade MKK3–MPK6 is an important part of the jasmonate signal transduction pathway in Arabidopsis. Plant Cell 19:805-818

Takeda K, Hori K (2007) Geographical differentiation and diallel analysis of seed dormancy in barley. Euphytica 153:249-256

Thabet SG, Moursi YS, Karam MA, Graner A, Alqudah AM (2018) Genetic basis of drought tolerance during seed germination in barley. PLoS One 13:e0206682

Thomas SG, Phillips AL, Hedden P (1999) Molecular cloning and functional expression of gibberellin 2-oxidases, multifunctional enzymes involved in gibberellin deactivation. Proc Natl Acad Sci USA 96:4698-4703

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

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

Tordenmalm S, Bason M, Chan S, Foster A, Greer P, Hatcher D, Schwarz P (2004) Sprout damage in barley. J Am Soc Brew Chem 62: 49-53

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

Turuspekov Y, Ormanbekova D, Rsaliev A, Abugalieva S (2016) Genome-wide association study on stem rust resistance in Kazakh spring barley lines. BMC Plant Biol 16:13-21

Ullrich SE (2010) Barley: Production, improvement, and uses (Vol. 12). John Wiley & Sons

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, Molina-Cano JL (2009) Genetic relationships between preharvest sprouting and dormancy in barley. Euphytica 168:331-345

United States Department of Agriculture (USDA) https://apps.fas.usda.gov/psdonline/circulars/grain.pdf

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

Van Hung P (2016) Phenolic compounds of cereals and their antioxidant capacity. Crit Rev Food Sci Nutr 56:25-35

Van Os H, Stam P, Visser RG, Van Eck HJ (2005) RECORD: a novel method for ordering loci on a genetic linkage map. Theor Appl Genet 112:30-40

Vanhala TK, Stam P (2006) Quantitative trait loci for seed dormancy in wild barley (*Hordeum spontaneum* C. *Koch*). Genet Resour Crop Evol 53:1013-1019

Varshney RK, Marcel TC, Ramsay L, Russell J, Röder MS, Stein N, Graner A (2007) A high density barley microsatellite consensus map with 775 SSR loci. Theor Appl Genet 114:1091-1103

Varshney RK, Tuberosa R (2007) Genomics-Assisted Crop Improvement, Vol 2: Genomics Applications in Crops. Springer

Visioni A, Tondelli A, Francia E, Pswarayi A, Malosetti M, Russell J, Comadran J (2013) Genome-wide association mapping of frost tolerance in barley (*Hordeum vulgare* L.). BMC Genomics 14:1-13

Von Bothmer R, Komatsuda T (2011) Barley origin and related species. Barley: production, improvement and uses, pp, 14-62

Von Bothmer R, Sato K, Komatsuda T, Yasuda S, Fischbeck G (2003) The domestication of cultivated barley. Diversity in barley (*Hordeum vulgare*), pp, 9-27

Von Bothmer R, Van Hintum T, Knüpffer H, Sato K (Eds.) (2003) Diversity in barley (*Hordeum vulgare*). Elsevier

Von Bothmer, R, Komatsuda T (2011) Barley origin and related species. Barley: production, improvement and uses, pp, 14-62

Wagner C, Schweizer G, Krämer M, Dehmer-Badani AG, Ordon F, Friedt W (2008) The complex quantitative barley–Rhynchosporium secalis interaction: newly identified QTL may represent already known resistance genes. Theor Appl Genet 118:113

Walker-Simmons M (1987) ABA levels and sensitivity in developing wheat embryos of sprouting resistant and susceptible cultivars. Plant Physiol 84:61-66

Walters DR, Avrova A, Bingham IJ, Burnett FJ, Fountaine J, Havis ND, Newton AC (2012) Control of foliar diseases in barley: towards an integrated approach. Eur J Plant Pathol 133:33-73

Wang JC (1985) DNA topoisomerases. Annu Rev Biochem 54:665-697

Wang M, Jiang N, Jia T, Leach L, Cockram J, Waugh R, Luo Z (2012) Genome-wide association mapping of agronomic and morphologic traits in highly structured populations of barley cultivars. Theor Appl Genet 124:233-246

Wei W, Min X, Shan S, Jiang H, Cao J, Li L, Chang C (2019) Isolation and characterization of TaQsd1 genes for period of dormancy in common wheat (Triticum aestivum L.). Mol Plant Breed 39:1-15

Weir BS (1996) Genetic Data Analysis II. Sinauer, Sunderland, MA

Weir BS (1996) Genetic data analysis II: methods for discrete population genetic data. Sinauer Assoc, Sunderl, MA

Weisstein EW (2004) Bonferroni correction https://mathworld.wolfram.com/BonferroniCorrection.html

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 Natl Acad Sci USA 101:9915-9920

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

Willcox G (1998) Archaeobotanical evidence for the beginnings of agriculture in Southwest Asia. The origins of agriculture and crop domestication, pp, 25-38

Wu R, Ma CX, Casella G (2002) Joint linkage and linkage disequilibrium mapping of quantitative trait loci in natural populations. Genetics 160:779-792

Wu Y, Bhat PR, Close TJ, Lonardi S (2008) Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. PLoS Genet 4:e1000212

Wulff RD (1995) Environmental maternal effects on seed quality and germination. In Kigel, J.; Galili, G. (Eds) Seed development and germination. New York, Marcel Dekker, pp, 491-506

Wulff RD (2017) Environmental maternal effects on seed quality and germination. In Seed development and germination, Routledge, pp, 491-505

Xu Y, Li P, Yang Z, Xu C (2017) Genetic mapping of quantitative trait loci in crops. J Crop Sci 5:175-184

Xu W, Tucker JR, Bekele WA, You FM, Fu YB, Khanal R, Badea A (2021) Genome assembly of the Canadian two-row malting barley cultivar AAC Synergy. G3 11:jkab031

Xue X, Jiao F, Xu H, Jiao Q, Zhang X, Zhang Y, Wang M (2021) The role of RNA-binding protein, microRNA and alternative splicing in seed germination: a field need to be discovered. BMC Plant Biol 21:1-11

Yamaguchi S (2008) Gibberellin metabolism and its regulation. Annu Rev Plant Biol 59: 225-251

Yang J, Yu Y, Cheng J, Tan X, Shen W (2011) Study on pre-harvest sprouting tolerance in *Triticum aestivum* ssp. *yunnanense King*. Journal of Triticeae Crops 31:747-752

Yang T, Shad Ali G, Yang L, Du L, Reddy ASN, Poovaiah BW (2010) Calcium/calmodulin-regulated receptor-like kinase CRLK1 interacts with MEKK1 in plants. Plant Signal Behav 5:991-994

Yee D, Goring DR (2009) The diversity of plant U-box E3 ubiquitin ligases: from upstream activators to downstream target substrates. J Exp Bot 60:1109-1121

Yoshida T, Nakata T, Ichishima E (1991) DNA topoisomerase I from rice: Enzyme synthesis in germination, and partial purification from cultured cells. Phytochemistry 30:3885-3887

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

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178:539-551

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203-208

Zaitlen N, Kraft P (2012) Heritability in the genome-wide association era. Hum Genet 131:1655-1664

Zeng X, Long H, Wang Z, Zhao S, Tang Y, Huang Z, Tashi N (2015) The draft genome of Tibetan hulless barley reveals adaptive patterns to the high stressful Tibetan Plateau. Proc Natl Acad Sci USA 112:1095-1100

Zeng X, Long H, Wang Z, Zhao S, Tang Y, Huang Z, Tashi, N (2015) The draft genome of Tibetan hulless barley reveals adaptive patterns to the high stressful Tibetan Plateau. Proc Natl Acad Sci USA 112:1095-1100

Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics 136:1457-1468

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 M, Pan J, Kong X, Zhou Y, Liu Y, Sun L, Li D (2012) *ZmMKK3*, a novel maize group B mitogen-activated protein kinase kinase gene, mediates osmotic stress and ABA signal responses. J plant Physiol 169:1501-1510

Zhao M, Yang S, Liu X, Wu K (2015) Arabidopsis histone demethylases LDL1 and LDL2 control primary seed dormancy by regulating *DELAY OF GERMINATION 1* and ABA signaling-related genes. Front Plant Sci 6:159

Zhou G, Panozzo J, Zhang XQ, Cakir M, Harasymow S, Li C (2016) QTL mapping reveals genetic architectures of malting quality between Australian and Canadian malting barley (*Hordeum vulgare* L.). Mol Breed 36:70

Zhou Y, Tang H, Cheng MP, Dankwa KO, Chen ZX, Li ZY, Wang JR (2017) Genome-wide association study for pre-harvest sprouting resistance in a large germplasm collection of Chinese wheat landraces. Front Plant Sci 8:401

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

Zhu Y, Wang S, Wei W, Xie H, Liu K, Zhang C, Ma C (2019) Genome-wide association study of pre-harvest sprouting tolerance using a 90K SNP array in common wheat (*Triticum aestivum* L.). Theor Appl Genet 132:2947-2963

Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, Yang J (2016) Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet 48:481-487

Zohary D, Hopf M (2000) Domestication of plants in the Old World: The origin and spread of cultivated plants in West Asia, Europe and the Nile Valley (No. Ed. 3). Oxford university press

Zohary D, Hopf M, Weiss E (2012) Domestication of plants in the Old World: The origin and spread of cultivated plants in West Asia, Europe and the Nile Valley (No. Ed. 4). New York Oxford University Press, pp, 20-74

APPENDIX

Appendix 1. Comparison between number of significant markers among the individual trial environments and the dataset pooled from all trial environments using correction methods FDR at α = 0.05 and lowest 0.1 percentile of the *p* value distribution.

Significance thresholds	Number of markers common in trial environments[a]	Number of markers based on the dataset pooled from all trial environments[a]
FDR adjusted *p* value (-log10 (P) > 4.6)	4 (BRD 2018 and MOR 2019)	16
0.1 percentile of *p* value distribution (-log10 (P) > 3.7)	5 (BRD 2018, BRD 2020 and MOR 2019)	30
	8 (BRD 2018 and BRD 2020)	
	8 (BRD 2018 and MOR 2019)	
	14 (BRD 2020 and MOR 2019)	

[a]BRD18, Brandon 2018; BRD19, Brandon 2019; BRD20, Brandon 2020; MOR19, Morden 2019

Appendix 2. Significant SNPs associated with PHS based on the dataset pooled from all trial environments using lowest 0.1 percentile *p* value threshold[a].

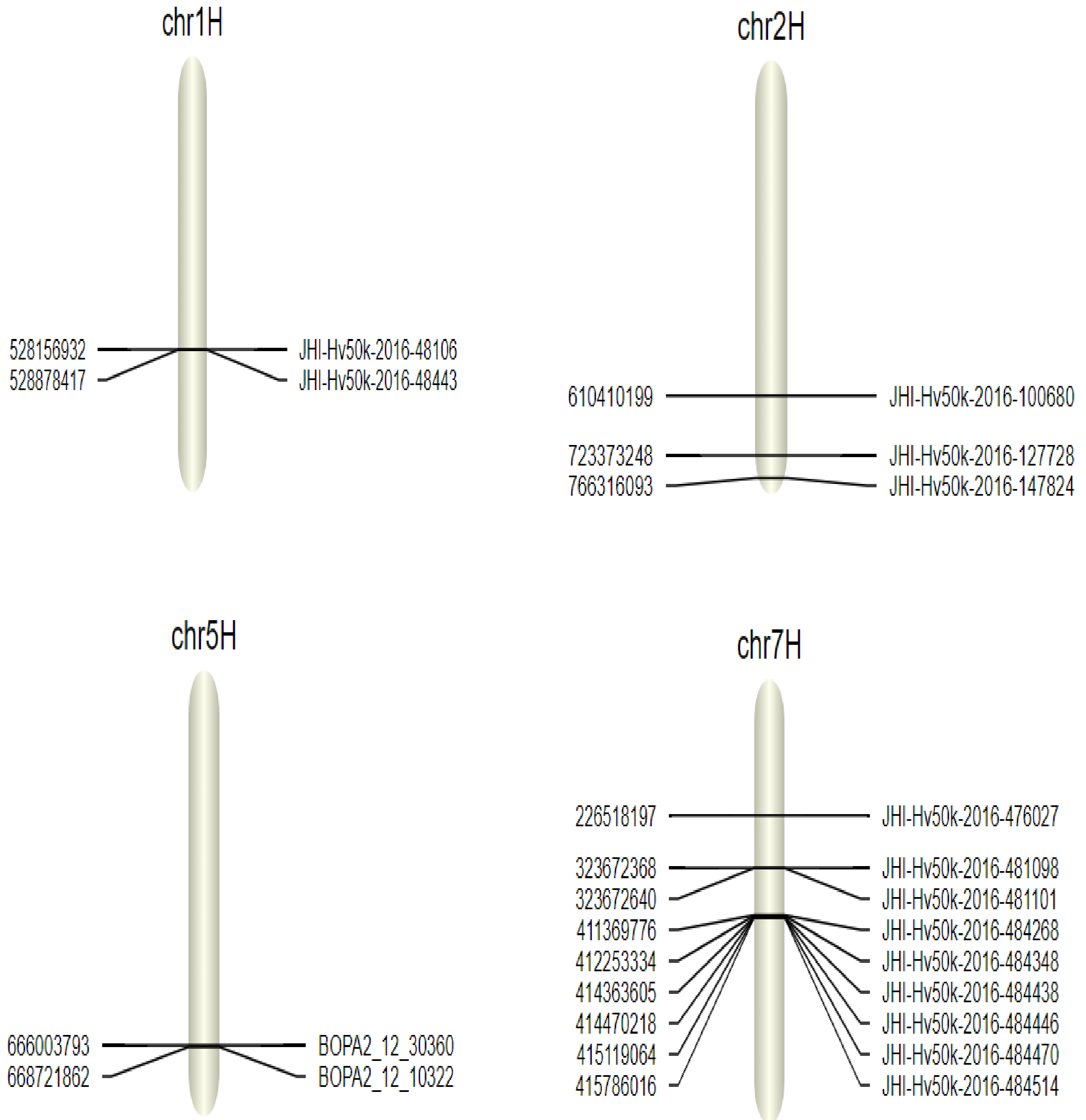
QTL[b]	Marker name	Chr[c]	Position (bp)[d]	Position (cM)[e]	-log10 (*P*)[f]	R^2 (%)[g]
QTL.1H-1	JHI-Hv50k-2016-16507	chr1H	23943414	35.69	3.77	5.80%
QTL.1H-2	JHI-Hv50k-2016-47595	chr1H	526845750	105.81	3.80	5.94%
QTL.1H-3	JHI-Hv50k-2016-48099	chr1H	528156681	103.82	3.82	5.93%
QTL.1H-3	JHI-Hv50k-2016-48106	chr1H	528156932	103.82	5.31	8.71%
QTL.1H-3	JHI-Hv50k-2016-48443	chr1H	528878417	104.75	5.57	9.23%
QTL.1H-4	JHI-Hv50k-2016-100680	chr2H	610410199	61.98	5.38	8.95%
QTL.1H-5	JHI-Hv50k-2016-127728	chr2H	723373248	122.86	5.21	8.55%
QTL.1H-6	JHI-Hv50k-2016-129529	chr2H	727209271	125.21	3.76	6.26%
QTL.1H-7	JHI-Hv50k-2016-147824	chr2H	766316093	149.15	5.88	10.01%
QTL.1H-8	BOPA1_3266-570	chr3H	580441401	N/A	4.40	6.97%
QTL.1H-8	BOPA1_ABC15290-1-3-262	chr3H	580635994	N/A	4.08	6.37%
QTL.1H-9	JHI-Hv50k-2016-165100	chr3H	72344688	45.82	3.77	5.80%
QTL.1H-10	JHI-Hv50k-2016-226832	chr4H	1668357	4.19	3.92	6.04%
QTL.1H-11	BOPA2_12_30360	chr5H	666003793	168.54	5.70	9.45%
QTL.1H-11	BOPA2_12_10322	chr5H	668721862	168.89	5.24	8.74%
QTL.1H-11	BOPA2_12_31123	chr5H	666455425	169.38	4.29	8.48%
QTL.1H-12	JHI-Hv50k-2016-481098	chr7H	323672368	67.92	4.65	7.53%
QTL.1H-12	JHI-Hv50k-2016-481101	chr7H	323672640	67.92	5.12	8.37%
QTL.1H-13	JHI-Hv50k-2016-475948	chr7H	225975685	70.54	3.84	5.98%
QTL.1H-13	JHI-Hv50k-2016-476010	chr7H	226516584	70.54	3.75	6.02%
QTL.1H-13	JHI-Hv50k-2016-476011	chr7H	226516747	70.54	3.91	6.64%
QTL.1H-13	JHI-Hv50k-2016-476024	chr7H	226517763	70.54	3.96	6.19%
QTL.1H-13	JHI-Hv50k-2016-476027	chr7H	226518197	70.54	4.83	9.75%
QTL.1H-13	JHI-Hv50k-2016-483347	chr7H	386940132	70.54	3.78	5.93%
QTL.1H-13	JHI-Hv50k-2016-484268	chr7H	411369776	70.54	6.13	10.41%
QTL.1H-13	JHI-Hv50k-2016-484348	chr7H	412253334	70.54	5.57	9.23%

Appendix 2 *continued*

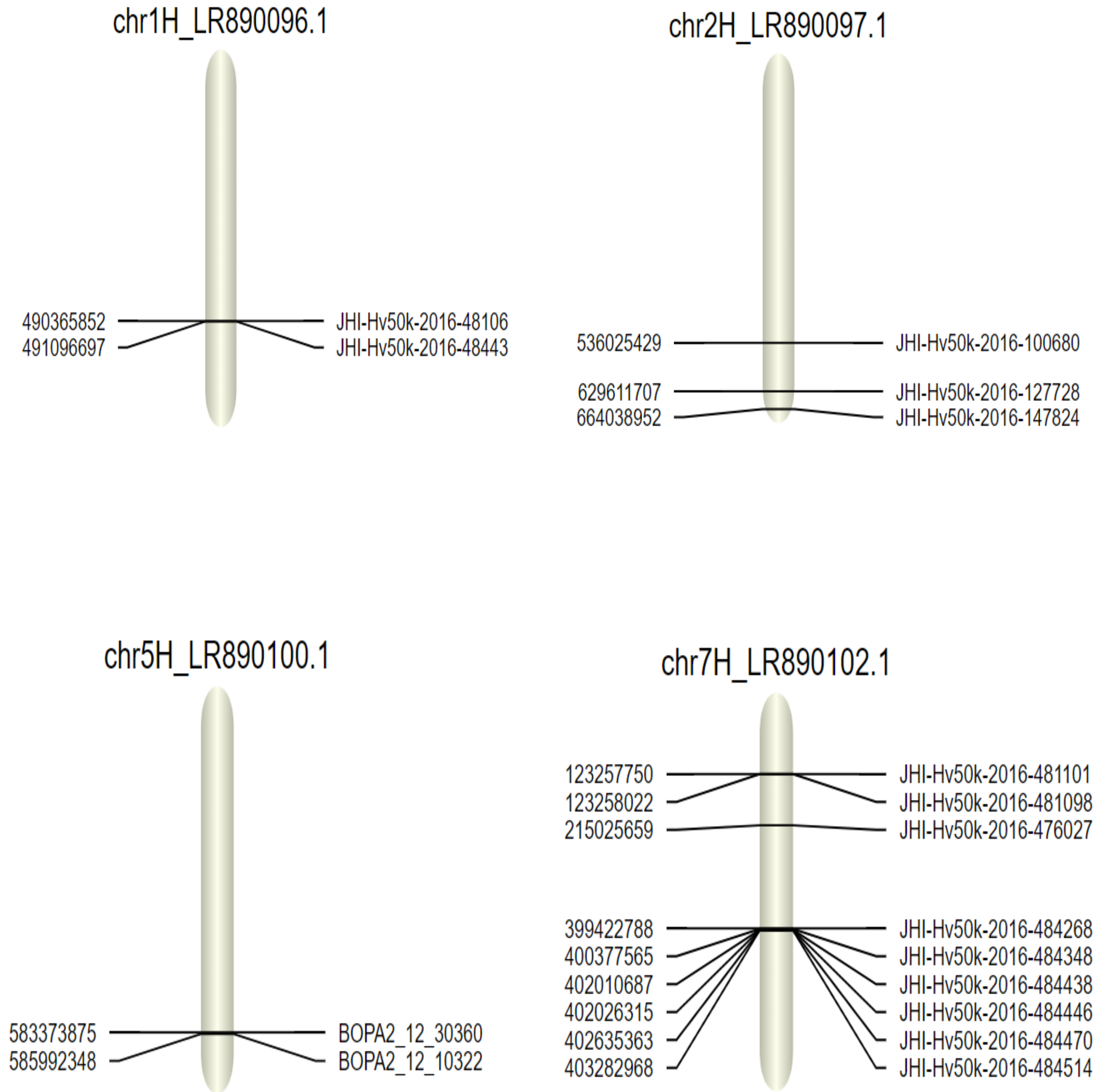
QTL.1H-13	JHI-Hv50k-2016-484438	chr7H	414363605	N/A	4.61	8.19%
QTL.1H-13	JHI-Hv50k-2016-484446	chr7H	414470218	70.54	4.98	8.10%
QTL.1H-13	JHI-Hv50k-2016-484470	chr7H	415119064	70.54	4.98	8.10%
QTL.1H-13	JHI-Hv50k-2016-484514	chr7H	415786016	70.54	4.93	11.69%

[a]Environment as described in Table 3.2
[b] Putative QTL containing one or more than one SNP- described on the basis of LD decay distance
[c]Chromosome
[d]Marker position in bp obtained from the barley reference genome of Morex v1 (Mascher et al. 2017)
[e]Marker positions in cM derived from POPSEQ_2017 genome map using Barleymap (Mascher et al. 2013; http://floresta.eead.csic.es/barleymap/)
[f]Markers selected on the basis lowest 0.1 percentile *p* value threshold of 2.0×10^{-4} [-log10 (*P*) > 3.7]
[g]R^2 (%) denotes the percentage of phenotypic variation explained by each significant SNP

Appendix 3. Physical map of the 16 significant SNPs located on chromosome 1H, 2H, 5H and 7H based on the genome physical map of Morex v1.



Appendix 4. Physical map of the 16 significant SNPs located on chromosome 1H, 2H, 5H and 7H based on the genome physical map of Morex v3.



Appendix 5. Candidate gene IDs in Morex v1 and Morex v2, and the location of Golden Promise v1 contigs overlapping with the candidate genes in Morex v2.

Gene name	Candidate gene IDs in Morex v1	Candidate gene IDs in Morex v2	GP v1 contigs overlapping with the candidate genes in Morex v2
HvMTHD	HORVU1Hr1G082150	HORVU.MOREX.r2.1HG0068060	contig1:453457244-453457567
HvHUP6	HORVU2Hr1G084190	HORVU.MOREX.r2.2HG0146380	contig2:500373790-500374022
HvPRX16	HORVU2Hr1G111610	HORVU.MOREX.r2.2HG0168450	contig2:583599162-583599569
HvPRORP1	HORVU2Hr1G127250	HORVU.MOREX.r2.2HG0180970	contig2:611435280-611436366
HvJMT	HORVU5Hr1G124360	HORVU.MOREX.r2.5HG0446360	contig5:536072526-536072922
HvPM19	HORVU5Hr1G125460	HORVU.MOREX.r2.5HG0447360	contig5:537931456-537932002
HvGlyRs	HORVU7Hr1G055340	HORVU.MOREX.r2.7HG0568920	contig7:196208929-196209447
HvTOP1	HORVU7Hr1G065200	HORVU.MOREX.r2.7HG0557540	contig7:107201705-107202362
HvLDL1	HORVU7Hr1G073020	HORVU.MOREX.r2.7HG0581950	contig7:365550216-365551659
HvPUB8	HORVU7Hr1G073100	HORVU.MOREX.r2.7HG0582070	contig7:366556949-366557837
HvCRLK1	HORVU7Hr1G073300	HORVU.MOREX.r2.7HG0582270	contig7:368117529-368117852
HvGT	HORVU7Hr1G073370	HORVU.MOREX.r2.7HG0582390	contig7:368689365-368690198
HvFIE	HORVU7Hr1G073410	HORVU.MOREX.r2.7HG0582460	contig7:369373983-369374393

Appendix 6. Sequence identity of the coding regions of the candidate genes with their homologs in other plant species.

Genes (Gene IDs)	Homologs (Accession number)	Identity
HvMTHD	*Triticum dicoccoides* LOC119330022 (XM_037603218)	93.49%
(HORVU1Hr1G082150)	*Aegilops tauschii* LOC109781464 (XM_020340069)	92.68%
	Brachypodium distachyon LOC100841577 (XM_024457811)	89.38%
HvHUP6	*Aegilops tauschii* LOC109782660 (XM_020341282)	96.24%
(HvHORVU2Hr1G084190)	*Triticum dicoccoides* LOC119355539 (XM_037622358)	96.00%
	Brachypodium distachyon LOC104585365 (XM_010241745)	91.29%
HvPRX16	*Triticum dicoccoides* LOC119356762 (XM_037623743)	98.40%
(HORVU2Hr1G111610)	*Aegilops tauschii* LOC109786327 (XM_020344905)	96.56%
	Brachypodium distachyon LOC100841484 (XM_003579364)	88.82%
HvPRORP1	*Triticum dicoccoides* LOC 119357339 (XM_037624340)	91.47%
(HORVU2Hr1G127250)	*Aegilops tauschii* LOC 109750542 (XM_020309503)	91.85%
	Brachypodium distachyon LOC 100821676 (XM_003580854)	86.03%
HvJMT	*Aegilops tauschii* LOC 109744228 (XM_040404498)	85.50%
(HORVU5Hr1G124360)	*Triticum dicoccoides* LOC 119269980 (XM_037551923)	80.42%
	Brachypodium distachyon LOC 100821690 (XM_003564985)	77.71%
HvPM19	*Triticum aestivum* PM19/A2 (XM_KP844890)	92.55%
(HORVU5Hr1G125460)	*Aegilops tauschii* LOC 109739175 (XR_005758519)	90.07%
	Triticum dicoccoides LOC 119289339 (XM_037568691)	90.55%
HvGlyRs	*Aegilops tauschii* LOC 109781612 (XR_020340201)	96.99%
(HORVU7Hr1G055340)	*Triticum dicoccoides* LOC 119330230 (XM_037603348)	96.68%
	Brachypodium distachyon LOC 100824721 (XM_040396474)	90.26%

Appendix 6 *continued*		
HvTOP1 (HORVU7Hr1G065200)	*Aegilops tauschii* LOC 109744400 (XM_020303518)	97.85%
	Triticum dicoccoides LOC 119329901 (XM_037603001)	97.73%
	Brachypodium distachyon LOC 100832827 (XM_010229462)	90.45%
HvLDL1 (HORVU7Hr1G073020)	*Triticum dicoccoides* LOC 119338613 (XM_037610908)	96.19%
	Aegilops tauschii LOC 109781192 (XM_020339795)	95.86%
	Brachypodium distachyon LOC 100835742 (XM_003573328)	87.32%
HvPUB8 (HORVU7Hr1G073100)	*Triticum dicoccoides* LOC 119340895 (XM_037612795)	95.75%
	Aegilops tauschii LOC 109733685 (XM_020292908)	95.13%
	Brachypodium distachyon LOC 100840831 (XM_014900206)	83.94%
HvCRLK1 (HORVU7Hr1G073300)	*Aegilops tauschii* LOC 109769088 (XM_020327821)	96.18%
	Triticum dicoccoides LOC 119331992 (XM_037605170)	95.72%
	Brachypodium distachyon LOC 100840227 (XM_010235997)	89.82%
HvGT (HORVU7Hr1G073370)	*Triticum dicoccoides* LOC 119339127 (XM_037611283)	93.79%
	Aegilops tauschii LOC 109762927 (XM_020321811)	93.09%
	Brachypodium distachyon LOC 100828911 (XR_731354)	84.37%
HvFIE (HORVU7Hr1G073410)	*Aegilops tauschii* LOC 109762930 (XM_020321814)	95.44%
	Triticum dicoccoides LOC 119340066 (XM_037611978)	95.38%
	Brachypodium distachyon LOC 100828611 (XM_003573305)	89.55%

ABBREVIATIONS

ABA	abscisic acid
AlaAT	*alanine aminotransferase*
AWPM	ABA-induced wheat plasma membrane polypeptide
BAC	bacterial artificial chromosome
BC	bonferroni correction
BLAST	basic local alignment search tool
BLUE	best linear unbiased estimator
BLUP	best linear unbiased predictor
BRD	Brandon
CRLK	calcium/calmodulin-regulated receptor-like kinase
DH	double haploid
DNA	deoxyribonucleic acid
DOG	delay of germination
DUF	domain of unknown function
FDR	false discovery rate
FIE	fertilization-independent endosperm
GA	gibberellin

GI	germination index
GLM	general linear model
GP	golden promise
GT	glycosyltranferase
GWAS	genome wide association study
HEB	halle exotic barley
HUP	hypoxia response unknown protein
IBSC	International barley sequencing consortium
JMT	jasmonate O-methyltransferase
LD	linkage disequilibrium
LDL	lysine-specific histone demethylase
LOD	logarithm of odds
LOXL	lysyl oxidase-like
MAF	minor allele frequency
MAGIC	multiparent advanced generation inter-cross
MAPK	mitogen activated protein kinase
MKK	mitogen-activated protein kinase kinase
MKKK	mitogen activated protein kinase kinase kinase

MLM	mixed linear model
MOR	Morden
MTHD	mitochondrial hydroxyacyl-thioester dehydratase
NAM	nested association mapping
NCED	nine-*cis*-epoxycarotenoid dioxygenase
PANTHER	protein annotation through evolutionary relationship
PC	principal component
PCA	principal component analysis
qPCR	quantitative polymerase chain reaction
PHS	pre-harvest sprouting
POPSEQ	population sequencing
PPR	pentatricopeptide repeat
PRX	peroxidase
PUB	PLANT U-box
QQ	quantile-quantile
QTL	quantitative trait loci
RFLP	restriction fragment length polymorphism
RIL	recombinant inbred lines

RNA	ribonucleic acid
RVA	rapid visco analyzer
SNP	single nucleotide polymorphism
SSR	simple sequence repeats
TASSEL	trait analysis by association, evolution and linkage
TOP	topoisomerase
USDA	United states department of agriculture