Genetic and transcriptomic analysis for pre-harvest sprouting resistance in red spring wheat

M M Uzzal Ahmed Liton

Thesis submitted to the faculty of Graduate Studies of

the University of Manitoba

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Plant Science

University of Manitoba

Winnipeg

Copyright © 2022 by M M Uzzal Ahmed Liton

ABSTRACT

Liton, M.M.U.A., Ph.D., The University of Manitoba, September 2022. Genetic and transcriptomic analysis for pre-harvest sprouting in red spring wheat. Ph.D. Supervisor: Dr. Belay T. Ayele

Pre-harvest sprouting (PHS) affects wheat production globally as it reduces grain yield and end-use quality. Given PHS resistance is a quantitative trait and strongly influenced by genetic and environment factors. This thesis project is therefore aimed to identify molecular mechanisms regulating dormancy, quantitative trait loci (QTL) and candidate genes for PHS resistance. To this end, this PhD thesis project comprised of a genetic study involving a mapping population and a transcriptomic analysis. The genetic study investigated the dormancy phenotype of a doubled haploid (DH) mapping population derived from dormant/non-dormant parents and evaluated over five environments, and these data along with the genotypic data obtained using a 90 K SNP Chip array was used for detection PHS resistance QTL. A total of four PHS resistance QTLs were detected on chromosome 1D, 4A, 6B and 6D across the five environments; 4A QTL was identified as a major QTL as it explained 40-50% of phenotypic variation while 1D QTL is novel as it was not reported previously. The transcriptomic study investigated selected DH lines showing contrasting dormancy phenotype using RNA-Seq. Our analysis revealed a large number of differentially expressed genes (DEGs) between dormant and non-dormant seeds. The DEGs were found to be enriched in cell, cell part, intracellular and integral component of membrane gene ontology classes. Genes upregulated in dormant seeds are found to be enriched in the carotenoid biosynthesis including ABA metabolism, diterpenoid biosynthesis including gibberellin catabolism and flavonoid biosynthesis as well as ABA signaling pathways while those downregulated are enriched in auxin signaling and GA biosynthesis pathways. Mapping of the DEGs on previously detected QTLs (1D, 4A, 6B and 6D) identified potential candidate genes for seed dormancy. Of these DEGs, *TaCYP71B11* and *TaDOG1L-2* on 1D, *TaPM19-A1* and *A2* on 4A, an uncharacterized gene (Gene ID: TraesCS6B02G325000) on 6B and *TaLEA4* on 6D showed upregulation in dormant seed while two uncharacterized genes (Gene IDs: TraesCS1D02G410700 and TraesCS4A02G317200) on 1D and 4A, *TaRH8* on 6B and *TaHIS4* on 6D QTLs showed downregulation. Furthermore, 2 to 25 SNPs were detected in seed dormancy candidate genes located within the four QTLs. Taken together, the results of this study

foster a deeper understanding of the molecular and genetic mechanisms of seed dormancy and therefore facilitate PHS resistance breeding program in wheat.

ACKNOWLEDGEMENTS

First and foremost, I would like to be grateful to Almighty ALLAH for His mercy and blessings.

A proverb says: "*A person is a person because of other persons*". I, therefore, would like to express my gratitude to some other persons who help, support and guide me to complete my PhD journey. First of all, I would like to express my gratitude to my supervisor Prof. Dr. Belay T. Ayele for his countless support, friendly guidance and timely supervision to complete the PhD program. I convey my special thanks to other advisory committee members Dr. Mark Jordan, Dr. Curt McCartney and Dr. Georg Hausner for their valuable guidance and feedback. I would like to extend my thanks to my past and present lab members for their support especially Dr. Tuan during my stay at CTC lab as well as in the field. I acknowledge and appreciate the support and help of everyone with whom I worked on this project.

I convey my sincere thanks to Drs. Colin Heibert, Santos Kumar, Anita Brûlé-Babel, Doug Cattani and their laboratory members along with Dr. Curt's laboratory members for genotyping and field work for my PhD research. I would like to express my heartfelt thanks to my fellow graduate students at the Department of Plant Science. I am also indebted to the staffs especially Martha Blouw in the Department of Plant Science for their friendly support during my stay.

I sincerely acknowledge the financial support from the University of Manitoba as University of Manitoba Graduate Fellowship (UMGF), Manitoba Crop Alliance (MCA) and National Sciences and Engineering Research Council of Canada (NSERC).

My endeavour would not be completed without support, comfort, company and friendship of all of my family, relatives and friends in Canada and Bangladesh. I am particularly grateful to my wife, Humayra Khatun, for her sacrifice to keep me in the realm of sanity especially during my tough time. Word is not enough to express my gratitude to my family. I am grateful and thankful to my father Latif, mother Tahamina, elder brother Bellal, father in law Din Mohammad and Mother in law Jahanara for their support and encouragement during my PhD journey. I would also like to thanks Dr. Harun and Ehsanul for making my day in Winnipeg and supporting me during the thesis writing process. Lastly, I want to thank all who support me in the accomplishment of this endeavour.

DEDICATIONS

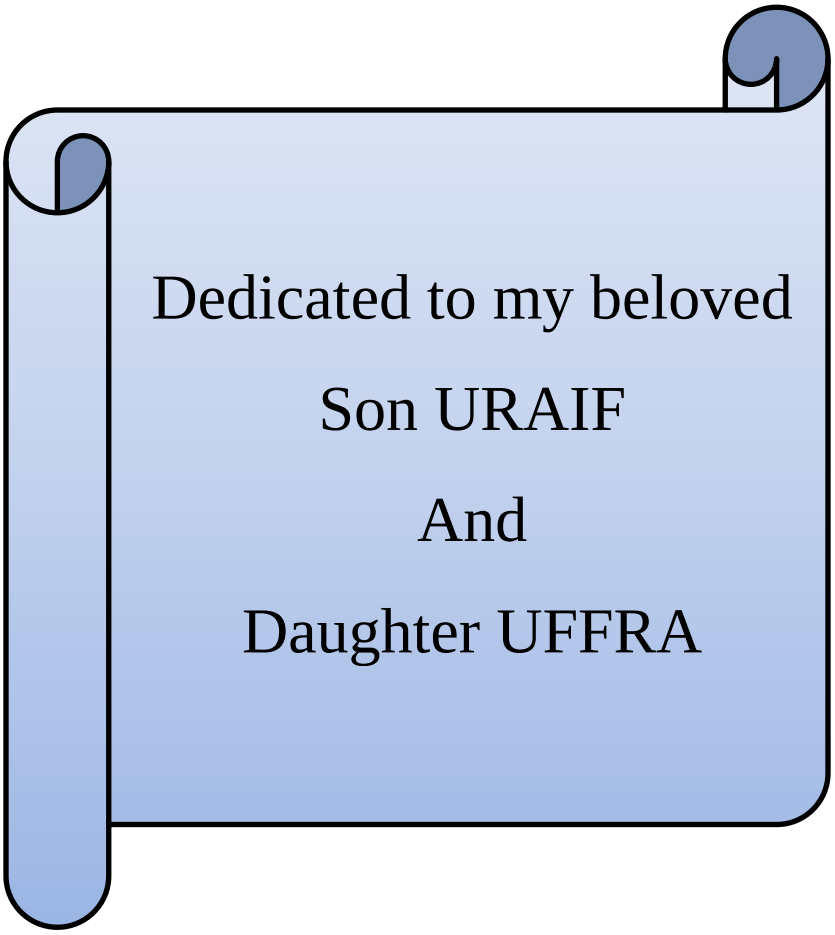


TABLE OF CONTENTS

Contents **Page no.**

ABSTRACT i
ACKNOWLEDGEMENTS iii
DEDICATIONS iv
LIST OF TABLES ix
LIST OF FIGURES x
LIST OF APPENDICES xii
FORWARD xiv
CHAPTER 1 1
INTRODUCTION 1
CHAPTER 2 5
LITERATURE REVIEW 5
2.1. Wheat as a staple food crop 5
2.2. Species of wheat 5
2.3. Evolution of modern wheat 6
2.4. Domestication and cultivation of wheat 8
2.5. Western Canadian wheat classes 8
2.6. Pre-harvest sprouting and its impact 11
2.7. PHS evaluation methods 11
2.8. Wheat grading regulations for PHS in Canada 13
2.9. Effect of PHS on grain quality and end-use application 14
2.10. Seed germination 14
2.11. Seed dormancy 15
2.12. Regulation of seed dormancy and PHS resistance 16
2.11.1. Hormonal regulation of seed dormancy and PHS resistance 16
2.11.2. Environmental regulation of seed dormancy and PHS resistance 21
2.11.3. Grain structure in regulation of dormancy 22
2.11.4. Red grain colour and PHS resistance 22
2.11.5. Secondary metabolites and seed dormancy 23

2.12 QTL mapping for PHS resistance 24
2.12.1. Sources of PHS resistance 24
2.12.2. Mapping populations 25
2.12.3. Genetic markers for genetic mapping in plants 26
2.12.4. Genetic linkage maps 28
2.12.5. Mapping QTL 29
2.13 PHS resistance QTLs and their underlying genes 31
2.14 Important genes controlling PHS resistance in wheat 33
2.14.1. TaMFT/TaPHS1 34
2.14.2. TaMKK3-A 35
2.14.3. TaVP1 36
2.14.4. TaPM19-A1 and A2 36
2.14.5. TaSdr 37
2.14.6. TaQsd1 37
2.14.7. Other candidate genes 38
2.15 Genome-wide gene expression/transcriptomic analysis 38
2.15.1. Wheat genome and transcriptome 39
2.15.2. Differential expression analysis of wheat transcriptomes 40
2.15.3. Gene ontology analysis 41
2.15.4. KEGG pathway analysis 42
2.15.5. Integration of RNA-Seq on QTL data 43
CHAPTER 3 45
IDENTIFICATION OF LOCI FOR PRE-HARVEST SPROUTING RESISTANCE IN THE HIGHLY DORMANT SPRING WHEAT RL4137 45
3.1. Abstract 46
3.2. Introduction 47
3.3. Materials and methods 49
3.3.1. Plant materials and trial environments 49
3.3.2. Assessment of PHS phenotype 49
3.3.3. Genotyping and linkage mapping 49
3.3.4. Statistical analysis 50

3.3.5. QTL analysis 51
3.4. Results 51
3.4.1. Characterization of germination phenotype and heritability 51
3.4.2. Analysis of environmental correlation and variance 54
3.4.3. Genetic mapping 54
3.4.4. Detection of QTL and allelic contribution 55
3.4.5. Combined effects of major QTLs for PHS resistance 61
3.5. Discussion 62
CHAPTER 4 67
TRANSCRIPTOMIC ANALYSIS OF SEED DORMANCY IN A HIGHLY DORMANT GENOTYPE RL4137 REVEALS CANDIDATE GENES FOR PRE-HARVEST SPROUTING RESISTANCE IN WHEAT 67
4.1. Abstract 68
4.2. Introduction 69
4.3. Materials and methods 73
4.3.1. Plant materials and growth conditions 73
4.3.2. Germination assay and allele frequency 73
4.3.3. RNA extraction 73
4.3.4. Library preparation and Illumina sequencing 74
4.3.5. Quality control and expression quantification 74
4.3.6. Differential expression analysis 74
4.3.7. GO and KEGG enrichment analysis of differentially expressed genes 75
4.3.8. Selection of DEGs and functional annotation 75
4.3.9. Identification of transcription factors (TF) and variants calling 75
4.3.10. Validation of RNA-Seq data using RT-qPCR 76
4.4. Results 77
4.4.1. Phenotypic characterization 77
4.4.2. Analysis of allelic frequency in DH lines 77
4.4.3. Quality of transcriptome assembly and gene expression quantification 78
4.4.4. Cluster analysis of expressed genes 79
4.4.5. Analysis of differentially expressed genes (DEGs) 80

4.4.6. GO enrichment analysis of DEGs associated with seed dormancy 82
4.4.7. KEGG pathway enrichment analysis of DEGs associated with seed dormancy 85
4.4.8. Pathways involved in hormone signaling and secondary metabolites biosynthesis 88
4.4.9. Differentially expressed genes within QTLs 99
4.4.10. Expression patterns of previously identified genes of PHS QTLs 105
4.4.11. Discovery of SNPs and InDels 107
4.4.12. Analysis of transcription factors 110
4.4.13. Validation of differentially expressed genes using RT-qPCR 112
4.5. Discussion 114
CHAPTER 5 121
GENERAL DISCUSSION AND CONCLUSION 121
LITERATURE CITED 126
APPENDIX 156

LIST OF TABLES

No.	Caption	Page
2.1	Milling classes of Western Canadian wheat with their characteristics and uses	**10**
2.2	Important genes controlling PHS resistance in wheat	**34**
3.1	Speed of germination index (SGI) of the 'Roblin'/RL4137 double haploid population and its parents in multiple environments	**52**
3.2	Correlation coefficients (r) between the trial environments for speed of germination index (SGI) in the 'Roblin'/RL4137 double haploid population	**54**
3.3	Analysis of variance of speed of germination index in the 'Roblin'/RL4137 double haploid population	**54**
3.4	QTLs for speed of germination index detected by interval mapping in the 'Roblin'/RL4137 double haploid population	**56**
3.5	QTLs for speed of germination index detected by inclusive composite interval mapping in the 'Roblin'/RL4137 double haploid population	**60**
3.6	Mean speed of germination index (SGI) of 16 genotypic classes defined by the four stable QTL in the 'Roblin'/RL4137 DH population	**61**
4.1	Allele frequency of selected dormant and non-dormant DH lines of 'Roblin'×RL4137	**78**
4.2	Differentially expressed genes (≥ 3-fold, P ≤ 0.05) involved in ABA signaling	**91**
4.3	Differentially expressed genes (≥ 3-fold, P ≤ 0.05) involved in auxin and ethylene signaling	**93**
4.4	Differentially expressed genes (≥ 3-fold, P ≤ 0.05) involved in the biosynthesis of secondary metabolites	**97**
4.5	Differentially expressed genes (≥ 2-fold, P ≤ 0.05) genes located in *QPhs.umb-4A*	**100**
4.6	Differentially expressed genes (≥ 10-fold, P ≤ 0.05) genes located in *QPhs.umb-6B*	**102**
4.7	Differentially expressed genes (≥ 10-fold, P ≤ 0.05) genes located in *QPhs.umb-1D* and *QPhs.umb-6D*	**104**
4.8	Expression patterns of previously reported PHS resistance genes in our samples	**106**
4.9	SNPs in DEGs located at 1D, 4A, 6B and 6D QTLs along with their physical positions	**107**
4.10	InDels in DEGs located at 1D, 6B and 6D QTLs along with their physical positions	**109**
4.11	Expression patterns of genes encoding transcription factors (≥ 80-fold, P ≤ 0.05)	**111**

LIST OF FIGURES

No.	Caption	Page
2.1	Evolutionary history of bread wheat modified from Rahman et al. (2020)	**7**
2.2	Hypothetical regulatory mechanisms of hormones and secondary metabolites for controlling seed dormancy modified from Sohn et al. (2021) and Tuan et al. (2018)	**19**
3.1	Frequency distributions of speed of germination index in the 'Roblin'/RL4137 double haploid population in each trial environment and pooled over all environments	**53**
3.2	Major QTL *QPhs.umb-4A* for PHS resistance detected on linkage group 4A.2 by both interval mapping (IM) and inclusive composite interval mapping (ICIM, not shown) in multi-environment experiments	**57**
3.3	Effect of *QPhs.umb-4A* on speed of germination index (SGI) in the 'Roblin'/RL4137 double haploid population	**58**
4.1	Germination percentage of the seeds of dormant and non-dormant double haploid lines evaluated for 7 days	**77**
4.2	Hierarchical clustering analysis based on similarities in expression patterns of genes expressed in the dry and imbibed seeds of dormant (D) and non-dormant (ND) double haploid lines at different time points of imbibition (0, 12 and 24 h)	**80**
4.3	Comparisons of the number of genes differentially expressed between seeds of dormant (D) and non-dormant (ND) double haploid lines at 0, 12 and 24 h after imbibition (D_0/ND_0, D_12/ND_12 and D_24/ND_24)	**82**
4.4	Gene ontology (GO) classes enriched with genes differentially expressed between seeds of the dormant (D) and non-dormant (ND) double haploid lines at 12 h after imbibition	**83**
4.5	Gene ontology (GO) classes enriched with genes differentially expressed between seeds of the dormant (D) and non-dormant (ND) double haploid lines at 24 h after imbibition	**85**
4.6	KEGG pathway enrichment analysis of genes differentially expressed between seeds of dormant (D) and non-dormant (ND) double haploid lines at 12 h after imbibition (D_12/ND_12)	**86**
4.7	KEGG pathway enrichment analysis of genes differentially expressed between seeds of dormant (D) and non-dormant (ND) double haploid lines at 24 h after imbibition (D_24/ND_24)	**87**

4.8	Plant hormone signal transduction pathways enriched with genes differentially expressed between the seeds of dormant and non-dormant double haploid lines	**89**
4.9	Carotenoid biosynthesis pathway enriched with genes differentially expressed between the seeds of dormant and non-dormant double haploid lines	**95**
4.10	Diterpenoids biosynthetic pathway enriched with genes differentially expressed between the seeds of dormant and non-dormant double haploid lines	**98**
4.11	Validation of the RNA-seq data using qPCR	**113**

LIST OF APPENDICES

No.	Caption	Page
1	Primer sequences used for qRT-PCR analysis	**156**
2	Details of the RNA-seq data from all dry and imbibed dormant and non-dormant seed samples	**157**
3	Overview of mapping of the transcriptomes of dry and imbibed seeds of dormant and non-dormant DH lines to the reference genome lines	**158**
4	Number of genes expressed in each sample of the dormant and non-dormant seeds	**159**
5	Pearson correlation coefficients among the four biological replicates of each dormant and non-dormant seed samples used for RNA-seq analysis	**160**
6	Genes differentially expressed between seeds of the dormant (D) and non-dormant (ND) double haploid lines. Comparison of expression levels between D seeds at 0 and 12 HAI (D_0/D_12) (A) and 12 and 24 HAI (D_12/D_24) (B), and between ND seeds at 0 and 12 HAI (ND_0/ND_12) (C) and 12 and 24 HAI (ND_12/ND_24) (D)	**161**
7	Genes differentially expressed between seeds of the dormant (D) and non-dormant (ND) double haploid lines. Comparisons of expression levels between D and ND seeds at 0 (D_0/ND_0) (A), 12 HAI (D_12/ND_12) (B) and 24 HAI (D_24/ND_24) (C)	**162**
8	List of genes differentially expressed (≥300-fold, $P \leq 0.05$) between dry and imbibed dormant and non-dormant seed samples	**163**
9	Number of gene differentially expressed in any of the comparisons and enriched in different GO classes	**164**
10	Flavonoid biosynthesis pathway enriched with genes differentially expressed between the seeds of dormant and non-dormant double haploid lines	**165**
11	Differentially expressed genes (≥2-fold, $P \leq 0.05$)) located in *QPhs.umb-4A* along with their orthologues and GO terms	**166**
12	Differentially expressed genes (≥10, $P \leq 0.05$) located in *QPhs.umb-6B* along with their orthologues and GO terms	**167**
13	Differentially expressed genes (≥ 10, $P \leq 0.05$) located in 1D and 6D QTLs along with their orthologues and GO terms	**168**
14	Number of SNPs and InDels distributed across the 21 chromosomes of wheat	**169**
15	Comparison of the coding sequences (CDS) of an uncharacterized gene (Gene ID: TraesCS6B02G293900) derived from dormant (D) and non-dormant (N) wheat seed	**170**
16	Number of differentially expressed genes encoding transcription factors (TFs)	**171**

and transcriptional regulators (TRs)

17 Gene regulatory networks of *TaDREB2C-B* based on STRING database **172**

18 List of abbreviations **173**

FORWARD

This thesis is written in manuscript style following the University of Manitoba thesis guidelines. The thesis is comprised of a general introduction about the research project and literature review followed by the two main manuscripts as the main body. Each manuscript has an abstract, introduction, materials and methods, results and discussion. The manuscripts are followed by a general discussion and conclusions, literature citation, and appendix sections. The thesis is formatted according to the Theoretical and Applied Genetics (TAG) journal.

CHAPTER 1

INTRODUCTION

Bread wheat (*Triticum aestivum* L.) is an allohexaploid species (2n = 6x = 42) belonging to the family Poaceae (Barnhart), also known as Gramineae (Juss). It is an important cereal crop that continues to feed a large number of populations worldwide (Venske et al. 2019). Currently wheat occupies a large portion of people's diet in many countries and provides energy and protein for humans and animals (Shewry 2009; Hawkesford et al. 2013). It is cultivated from 67° North to 45° South but its cultivation is relatively less in tropical regions (Feldman 1995). In 2019, ~765.8 million tonnes of wheat was produced globally, of which, ~32.4 million tonnes was produced in Canada (FAO 2021). During the same year, Canada exported ~22.8 million tonnes of wheat to different parts of the world (FAO 2021). Canada is known for producing high quality wheat; however, wheat productivity is constantly affected adversely by several biotic and abiotic factors including pre-harvest sprouting (PHS).

PHS is defined as the germination of seeds on the spike before harvest due to wet summer weather conditions and is considered as a major problem of wheat production throughout the world (Rodríguez et al. 2015; Nakamura 2018). PHS causes severe degradation of starch and protein in grain leading to yield and quality loss (Kim and Kim 2021). The annual economic loss due to PHS has been estimated to be approximately 1.2 billion USD globally and $100 million in Canada (DePauw et al. 2012; Shao et al. 2018). In bread wheat, 60-80% of variation in PHS resistance/susceptibility is accounted by difference in the level of seed dormancy (Martinez et al. 2020), which refers to the inability of viable seeds to germinate under conditions favourable/conducive to germination (Finkelstein et al. 2008; Bewley et al. 2013). Seed dormancy is a key quantitative trait controlled by multiple genes and is highly influenced by environmental factors including temperature, rainfall and humidity (Gao and Ayele 2014; Rodríguez et al. 2015). While both high and low levels of seed dormancy are undesirable for crop production (Gao and Ayele 2014; Tuan et al. 2018), the presence of optimum/intermediate level of dormancy in seeds is essential to prevent PHS damage and keep balance between dormancy and the need for germination with subsequent seedling emergence (Tuan et al. 2018; Martinez et al. 2020). Better understanding of the physiological, genetic and molecular

regulatory mechanisms of seed dormancy is necessary to develop cultivars with an optimum level of seed dormancy.

Hormonal regulation of seed dormancy is a well-established physiological mechanism that is crucial for PHS resistance in cereals (Tuan et al. 2018; Sohn et al. 2021). Among the phytohormones, abscisic acid (ABA) and gibberellin (GA) are two key regulators which regulate seed dormancy antagonistically (Tuan et al. 2018; Sano and Marion-Poll 2021; Sohn et al. 2021). ABA induces seed dormancy during seed maturation and regulates its maintenance in mature seed while GA releases seed dormancy and promotes germination (Martinez et al. 2020; Wang et al. 2021). Other hormones like ethylene (ET), jasmonate (JA) and brassinosteroid (BR) regulates seed dormancy and PHS resistance negatively through regulating seed ABA sensitivity while auxin promotes seed dormancy (Shu et al. 2016).

Seed dormancy and PHS resistance are quantitative traits controlled by genes, of which, many are located within quantitative trait loci (QTL) (Ali et al. 2019). A QTL is a segment of DNA highly correlated with variations in a quantitative trait of an organism and it is used as a powerful tool for genetic analysis (Miles and Wayne 2008; Nakamura 2018). The advent of sequence based DNA markers revolutionized the genetic analysis of quantitative traits. To date, numerous PHS resistance and seed dormancy QTLs have been successfully mapped across the 21 chromosomes of hexaploid wheat genome using different types of markers and populations (Ali et al. 2019; Martinez et al. 2020; He et al. 2021). Among the identified QTLs, a few major effect QTLs were detected on chromosome group 3, 4A and 4B (Ali et al. 2019; Wang et al. 2019). Of these, QTLs on chromosome group 3 and 4A have been detected consistently irrespective of the experimental materials and environments, and they explain large extent of phenotypic variations (Nakamura 2018; Ali et al. 2019). In addition, the effect of QTLs varies with epistasis, nature of segregating population and tested environments. QTL epistasis (Q×Q) and interaction with environments (Q×E and Q×Q×E) have been studied previously to uncover the complex genetic structure of PHS resistance in wheat (Ali et al. 2019). However, there is still a need to identify novel QTL or validate previously identified QTLs using new genetic materials for better understanding of the genetic mechanism of PHS resistance.

The advancement of wheat genome sequencing and genome annotation is playing a role in facilitating the identification of causal genes of QTLs. To date, several genes co-localized with seed dormancy/PHS resistance QTLs of wheat including *mother of flowering locus T* and

terminal flower 1/*propyzamide-hypersensitive 1* (*TaMFT*/*TaPHS1*), *mitogen-activated kinase kinase 3* (*TaMKK3*), *viviparous 1* (*TaVP1*), *myb-transcription factor 10* (*TaMyb10*), *ABA induced plasma membrane 19-A1* and *A2* (*TaPM19-A1* and *A2*), *seed dormancy* (*TaSdr*) and *alanine aminotransferase* (*TaQsd*) have been identified using comparative genomics and transcriptomics approaches (Ali et al. 2019; He et al. 2021). Of these, *TaMFT*/*TaPHS1* has been reported as an important causal gene of PHS resistance on 3A QTL, and two single nucleotide polymorphisms (SNPs) on promoter and intron region of this gene are shown to regulate PHS resistance (Nakamura et al. 2011; Liu et al. 2013b). The *TaMKK3-A* is another important causal gene identified at the *Phs1* locus of chromosome 4AL and a functional SNP at +660 bp was reported as regulator of PHS resistance (Torada et al. 2016). Both the *TaMFT*/*TaPHS1* and *TaMKK3-A* genes are reported to regulate PHS resistance through modulating ABA responsiveness of wheat seeds (Lin et al. 2018; Nakamura 2018). The *TaVP1* is another important candidate gene identified on the long arm of chromosome group 3 and orthologous to *ABA insensitive 3* (*TaABI3*), and this gene is hypothesized to regulate seed dormancy in wheat by encoding embryo specific dormancy related transcription factors such as *leafy cotyledon 1* and *2* (*LEC1* and *2*) and *fusca 3* (*FUS3*) (Bailey et al. 1999; Rikiishi and Maekawa 2014; Vetch et al. 2019). Other genes playing crucial roles in regulation of seed dormancy and germination are considered as candidates for PHS resistance in wheat including *TaSdr* and *TaQsd* (He et al. 2021). However, it is possible that many other genes co-localized with seed dormancy QTL regions remain unidentified due to the unavailability of complete genome sequence information.

High-throughput sequencing technologies and availability of genome sequence information are enabling to fill the knowledge gap on genetic and molecular mechanisms of PHS resistance/seed dormancy. Transcriptomic analysis is a good approach to understand molecular mechanisms of genes and their regulatory network in controlling complex biological traits such as seed dormancy (Sohn et al. 2021). Such studies provide information on expression pattern of all expressed genes in a genome and their association with phenotypic variations in a trait of interest (Diouf et al. 2020). Through the application of high-throughput sequencing technologies and high-quality genome sequence information, several dormancy related genes including *TaMFT*, *TaPM19-A1* and *A2* have been identified in wheat along with their molecular mechanisms in regulation of seed dormancy (Nakamura et al. 2011; Barrero et al. 2015). For instance, using RNA-Seq based transcriptomic analysis, previous studies reported several

candidate genes such as *ubiquitin ligase enzyme 3* and variants including SNPs and Insertions-Deletions (InDels) on 3A and 3D QTLs regulating PHS resistance in wheat (He et al. 2021; Wang et al. 2021). Despite the advances made to date, because of the polygenic nature of the trait, there is a need to conduct more genetic and transcriptomic studies to identify novel genomic regions and candidate/causal genes responsible for PHS resistance and gain insights into their underlying regulatory mechanisms. The objectives of this thesis are therefore:

1. to identify loci for PHS resistance in highly dormant spring wheat RL4137
2. to investigate molecular mechanisms associated with seed dormancy in RL4137
3. to identify candidate genes and gene-based SNPs and InDels regulating PHS resistance genes by combining genetic and genome-wide gene expression studies

CHAPTER 2

LITERATURE REVIEW

2.1. Wheat as a staple food crop

Wheat is a collective name for wild or domesticated cereals belonging to the genus *Triticum*. It is more widely grown in the world than any other staple crops, cultivated on 215 million hectares of land (www.wheat.org) covering regions from 67˚ N in southern Russia to 45˚ S in southwestern Australia and Argentina (Feldman 1995). Global trade of wheat is worth approximately $50 billion per year and around 2.5 billion people in 89 countries consume wheat as their staple food (www.wheat.org). The current production volume and consumption of wheat places it as one of the world's most important crop in the 21st century, contributing approximately 20% of world calorie intake (Braun et al. 2010) and 20% of the daily protein (Shewry 2009). Over two-thirds of the wheat produced globally is used for food while 20% is used for livestock feed and about 3% to 5% for seed (Melini and Carcea 2016). In recent years, the amount of wheat produced is not satisfying the demand, triggering price instability and hunger riots. With a predicted world population of 9 billion in 2050, the demand for wheat is expected to increase by 60%. To meet this demand, annual wheat yield increases must rise from the current level of below 1% to at least 1.6% (Willenbockel 2014). All countries share the need to increase wheat yield and quality through increasing tolerance to biotic and abiotic stresses, improving input use efficiency and developing new varieties for a more sustainable wheat production system. Genetically improved wheat varieties and development of innovative cropping systems are also a priority.

2.2. Species of wheat

All the diploid, tetraploid and hexaploid species of wheat belong to the genus *Triticum*, which is divided into five sections and three of these five sections are very common although they belong to different taxonomic classification (Goncharov 2011). *Monococcon* (AA genome) represents section 1, which consists of *Triticum monococcum* and *T. urartu* diploid species. Section 2 is represented by *Dicoccoides* (AABB genome) that consists of *T. turgidum* and *T. durum*, and it also includes *T. timopheevii*, which comprises AAGG genome of tetraploid species. Section 3 is

represented by *Triticum*, which includes different species comprising three different genome compositions of hexaploid wheat such as *T. zhukovskyi* (AAAAGG) and *T. aestivum* (AABBDD) (Matsuoka 2011). Some species of *Triticum* and *Aegilops* genus are combined in some classes and form a common *Triticum* genus due to their close evolutionary relationship (Bowden 1959; Goncharov 2011). Members of the genus *Triticum* exhibits diverse morphological features in their spikes (Fig. 2.1) and have been domesticated around the world over several millenniums (Rahman et al. 2020). They are used for different purposes such as pasta, bread and noodles worldwide, of which *T. aestivum* (bread wheat) and *T. turgidum* (durum wheat) are notable and commonly used for making bread and pasta, respectively.

2.3. Evolution of modern wheat

Modern wheat has undergone long evolutionary events including origin of wild and cultivated emmer, development of spelt wheat and finally formation of modern durum and bread wheat (Rahman et al. 2020). Sporadic changes in wheat genome have happened during this evolutionary process including rapid alteration, hybridization, polyploidization, mutation and domestication (Fig. 2.1). Modern bread wheat (*T. aestivum* subsp *aestivum*) is an allohexaploid (AABBDD) containing 2n=6x=42 chromosomes evolving through hybridization between tetraploid *T. turgidum* subsp *dicoccum* (AABB, cultivated emmer) and the diploid *Aegilops tauschii* (DD) followed by mutations around 8.5 thousand years before prehistoric Stone Age (Fig. 2.1) (Marcussen et al. 2014; Rahman et al. 2020). The A genome donor, *Triticum urartu* (AA, Wild einkorn) and the possible B genome donor *Aegilops speltoides* (SS, Goat grass) were hybridized and polyploidized around 300-500 thousand years ago to produce *Triticum turgidum* subsp. *dicoccoides* (AABB, wild emmer), which was further domesticated as *Triticum turgidum* subsp. *dicoccum* (AABB, cultivated emmer) by subconscious selection around 10 thousand years ago (Chapman et al. 1976; Feldman and Levy 2012; Rahman et al. 2020). The D genome donor, *Aegilops tauschii* (DD genome) was then hybridized with *Triticum turgidum* subsp. *dicoccum* to evolve *Triticum aestivum* subsp. *spelta* (AABBDD, spelt wheat) around 9 thousand years ago, which was further mutated to evolve *T. aestivum* subsp *aestivum* (AABBDD, modern bread wheat) around 8 thousand years ago (Rahman et al. 2020). The origin of the B genome is still controversial whether it was donated by a single *Aegilops* species or through an introgression event of several *Aegilops* species (Salse et al. 2008). As the Sitopsis section of S genome is

conserved in several *Aegilops* species, the closest existing species of the B genome donor is *Aegilops speltoides* (Li et al. 2022).

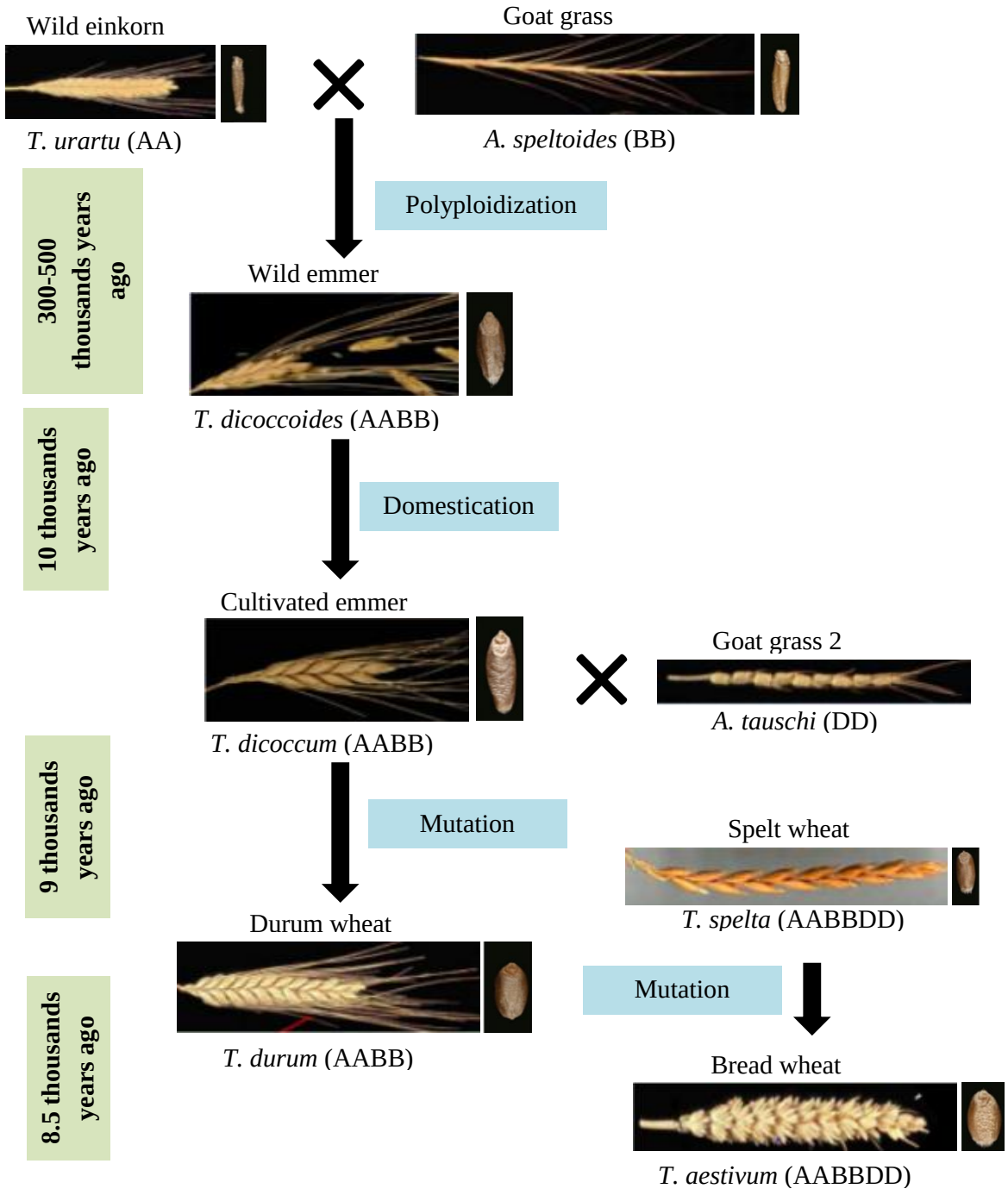


Fig. 2.1 Evolutionary history of bread wheat modified from Rahman et al. (2020). Text in aqua colour boxes represents the events and yellow colour boxes represent approximate times of the events. Changes in spike and grain morphology during the evolution are shown in the picture.

2.4. Domestication and cultivation of wheat

Local hunter-gatherers harvested wild progenitors of wheat in Western Asia in the Epipaleolithic period (Head and Atchison 2012). Those progenitors of earliest domesticated wheat were distributed in Fertile Crescent, an area between eastern shore of Mediterranean and foothills of Zagros Mountains (Weide 2015). A group of fisher-hunter-gatherers in Levant in the Middle East were the first known modern cultures of human to exploit wheat systematically in 23,000 BP (Bar-Yosef 1998). *T. monococcum* (einkorn), the primitive diploid ancestor of wheat was cultivated in Levant; however, the most common species in this area were *T. aegilopoides* and *T. urartu* (Gill and Friebe 2002). *T. turgidum* subsp. *dicoccum*, a tetraploid wheat species, was cultivated in southeastern Turkey and domesticated as durum wheat in 9,000 years ago (Heun et al. 1997; Gill and Friebe 2002). Historical evidences indicated that hexaploid bread wheat species was domesticated and cultivated in Fertile Crescent around 8000 years ago (Rahman et al. 2020; Ghahremaninejad et al. 2021).

Common wheat/bread wheat was first dispersed to Asia, Africa and Europe, and then North America after its origin in the Fertile Crescent (Peterson 1965; Olmstead and Rhode 2011; Marcussen et al. 2014). Cultivation of wheat was confined to Europe, Africa and Asia in the fifteen century. In 1602, wheat was distributed to North America and was brought to South America and Australia afterwards (Olmstead and Rhode 2011; Peterson 1965). The French settlers were the first to grow wheat in Canada, which took place in 1605 at the Port Royal of Nova Scotia, and the Selkirk settlers took the earliest attempts to cultivate wheat in Western Canada in 1812 (Buller 1919). An Ontario farmer, David Fife, received wheat seeds from Glasgow, Scotland in 1842 and selected a hard red spring wheat variety called “Red Fife”. Hard Red Calcutta and Red Fife were crossed to develop cultivar Marquis, which was the foundation of Western Canadian wheat development (Buller 1919; DePauw and Hunt 2001). Later, Thatcher, a western Canadian spring wheat variety, was developed from Marquis and it was widely used to develop western Canadian wheat (Fraser and Whiteside 1956).

2.5. Western Canadian wheat classes

In Western Canada, wheat is classified into 10 market classes including eight milling classes for bread/common wheat, one for durum wheat, and one for special purpose wheat, namely Canada Western Special Purpose (CWSP) (Table 2.1) (Canadian Grain Commission 2019)

(https://grainscanada.gc.ca/en/grain-quality/grain-grading/wheat-classes.html). CWSP class of wheat is not appropriate for milling but for ethanol production and feed due to its high starch and low protein contents, and is also used as a check to evaluate varieties of other classes for registration. Each of the 8 classes of bread wheat is distinct in their characteristics and this helps to distinguish one class from another (Table 2.1). Canada Western Red Spring (CWRS) class of wheat is the largest wheat class among the 8 milling classes of bread/common wheat as it is grown with the largest acreage. In 2021, CWRS wheat was produced in 66% of wheat acreage and it was grown mainly in Saskatchewan followed by Manitoba, Alberta and British Columbia. The top five CWRS wheat varieties in 2021 were AAC Brandon, AAC Viewfield, CDC Landmark, AAC Elie, AAC Redberry (https://www.grainscanada.gc.ca/en/grain-research/statistics/varieties-by-acreage/). CWRS class is well recognized for its quality including milling and baking qualities and contains higher amount of protein. It is widely used to produce high-volume pan breads and various baked products including hearth breads, flat breads and steamed breads due to good dough strength. It is also used for making pasta and noodles. This milling class of wheat is also well known to the industry due to low protein loss during milling (https://canadiancereals.ca/2020-wheat-crop/). Owing to these features, the CWRS wheat class predominates the wheat export by Canada.

Table 2.1 Milling classes of Western Canadian wheat with their characteristics and uses

Wheat Class	Characteristics	Uses
Canada Western Red Spring (CWRS)	Hard red spring wheat, superior milling and baking quality, three milling grades, various guaranteed protein levels	High-volume pan breads, hearth breads, flat breads, steamed breads, noodles, common wheat pasta
Canada Western Extra Strong (CWES)	Hard red spring wheat, extra strong gluten, two milling grades	Ideal for blending, Specially product high gluten strength
Canada Prairie Spring Red (CPSR)	Red spring wheat, medium hard kernels, medium dough strength, two milling grades	Hearth breads, flat breads, steamed breads, noodles
Canada Western Red Winter (CWRW)	Hard red winter wheat, very good milling quality, three milling grades	French breads, flat breads, steamed breads, noodles
Canada Prairie Spring White (CPSW)	White spring wheat, medium dough strength, two milling grades	Flat breads, chapatis, noodles, crackers
Canada Western Soft White Spring (CWSWS)	Soft white spring wheat, low protein content, three milling grades	Cookies, cakes, pastries, crackers, flat breads, chapatis, noodles
Canada Western Hard White Spring (CWHWS)	Hard white spring wheat, superior milling quality producing flour with excellent color, three milling grades	Noodle, flat bread, pan bread, buns
Canada Northern Hard Red (CNHR)	Red spring wheat, medium to hard kernels, very good milling quality, medium gluten strength, three milling grades	Hearth breads, flat breads, steamed breads, noodles
Canada Western Amber Durum (CWAD)	Durum wheat, high yield of semolina, excellent pasta making quality, four milling grades	Semolina for excellent pasta-making quality, couscous

Source: Canadian Grain Commission (https://grainscanada.gc.ca/en/grain-quality/grain-grading/wheat-classes.html)

2.6. Pre-harvest sprouting and its impact

PHS is defined as the precocious germination of physiologically mature seeds of wheat on the spike prior to or during harvest due to wet weather conditions at harvesting season (Derera 1989). Prolonged rainfall with high humidity and high temperature facilitate PHS in the field through enhancing water absorption by the seeds and therefore initiating the germination process. Hormonal levels such as that of GA increases due to water absorption by seeds, triggering enhanced activity of hydrolytic enzymes such as α-amylases that degrade the starch (De Laethauwer et al. 2013). PHS can be assessed visually as it causes swelling of kernel, splitting of seed coat, emergence of root and shoot. PHS decreases test weight and quality of wheat grain, resulting in yield loss and limitation of end-use application (Thomason et al. 2019). Sprouted grains are also difficult to thresh and this contributes to harvest loss. As a result, PHS causes substantial financial loss to wheat producers and industries.

PHS is one of the major problems that occurred frequently in most of wheat producing areas in the world including Europe, Africa, Australia, China, USA and Canada (Rajjou et al. 2012). The first global report for PHS incidence was made in 1990, and this report documented PHS damage only for four years, from 1978 to 1988 (Derera 1990). Subsequently, another study reported that 83% of the wheat planted area in China in 1997 was affected by PHS (Xiao et al. 2002). In Canada, significant PHS damage to wheat occurred four times between 1978 and 1988, and each of the four damages caused a loss of ~$100 million (Derera 1990). The financial loss increased to over $100 million in 2000, when PHS caused high α-amylase activity and therefore undesirable Falling Number (FN) in the wheat grains (Clarke et al. 2005). Globally, the annual financial loss due to PHS has been estimated to be ~$1.2 billion (Shao et al. 2018; Wang et al. 2021).

2.7. PHS evaluation methods

Visual examination of sprouted grains is the only way to evaluate PHS in the field when the sprouting incidence is severe. In the field, PHS is evaluated on a 1 to 9 score scale in which 1 is for no sprouting and 9 is for complete sprouting (visible seminal roots and coleoptiles) (McMaster and Derera 1976; Kulwal et al. 2010). Visible indications of PHS include distinct swelling of germ, splitting of bran, discoloration of grains and emerging of roots and shoots along with clear evidence of germ growth. However, PHS damage is sometimes so precise or

subtle and difficult to analyze and describe due to lack of clear visual evidence. Therefore, visual characterization is not suitable for grading of wheat grains. As a result, PHS should be assessed using other methods such as plating assays and spike-wetting that evaluate the degree of seed dormancy (Trethowan et al. 1996).

PHS damages are caused by starch degrading α-amylase which is initiated by water uptake by wheat seeds before any evidence of germination is visible (Olaerts et al. 2016). This hydrolytic enzyme greatly reduces flour quality of milled grain through digestion of starch. Therefore, for trading purposes, milling industries assess PHS mainly using Hagberg-Perten FN test (Perten 1964) that assesses the activity of α-amylase (Mares and Mrva 2008). The FN test involves grinding of the wheat grains to whole-grain flour to make a slurry with water followed by heating at 100°C. A stirrer is allowed to reach the bottom into the slurry and the FN machine measures the time (in seconds) it takes for the stirrer to fall to the bottom of the slurry. Sprouted grains with elevated α-amylase activity produced thinner slurry and the stirrer falls faster the bottom of the slurry, resulting in a lower FN number. If the FN number is below the standard level, which is 300s, the producers receive lower amount of money for their grain (Martinez et al. 2018) since those grains are not suitable for milling and are sold directly for animal feeding. Previous studies reported significant correlation between FN and sprouting scores, thus, it could be possible to detect similar loci for PHS resistance using FN and sprouting scores although this has not been tested to date (DePauw et al. 2012). Even though the FN test can be used in breeding program to quantify PHS resistance loci; the breeding materials should be assessed using other assays for their PHS resistance. To solve this problem, wheat breeders and researchers introduced temperature and precipitation simulated model to evaluate PHS resistance such as germination test of threshed seeds and sprouting rate in intact spikes, which involves wetting of intact spikes after harvest followed by scoring (Gordon et al. 1977, 1979).

Two kinds of PHS assessment methods have been widely accepted and used by the wheat research community. The first one is germination test of threshed seeds in plating assay in laboratory or temperature controlled chamber and using germination index (GI) as a phenotype for PHS assessment (Knox et al. 2012; Rasul et al. 2012; Chono et al. 2013). For germination test, physiologically mature seeds are harvested from the field, threshed and dried at room temperature to lower the moisture level at approximately 12% in seeds (Humphreys and Noll 2002). The dried seeds are then subjected to germination test using plating assay (Gao et al.

2012) followed by germination count. Finally, germination index is calculated using the formula described by Walker-Simmons (1988).

The second widely used PHS assessment method is spike-wetting followed by scoring under controlled temperature and precipitation in a mist chamber (Paterson et al. 1989). This method evaluates sprouting rate and/or sprouting index (Chen et al. 2008; Rasul et al. 2009; Lin et al. 2015). For sprouting test, five to fifteen physiologically mature and intact spikes will be hand-harvested from fields and dried well before placing in the mist chamber. Intact spikes are selected for this test because those spikes exhibit the effect of awns, spike erectness, tightness of glooms and types of head on sprouting (King and Richards 1984). The spikes are then placed in the mist chamber for a specified duration of time under controlled environment to facilitate sprouting, and then sprouted spikes are either scored based on germination and post-germination growth of roots and shoots (McMaster and Derera 1976) or the score is used to calculate the percentage of germination (Shao et al. 2018).

2.8. Wheat grading regulations for PHS in Canada

Canada is well recognized for exporting good quality wheat grain to different parts of world for a long period of time. The quality of wheat grain in Canada is monitored by Canadian Grain Commission (CGC), an organization of the Federal Government. With respect to grading for PHS, CGC has set various standards of sprouted kernel for different grades in nine wheat classes. For instance, when > 3% in total and > 0.3% in any individual grade of CWRS wheat grain are damaged by sprouting, they are then considered as sprouted and severely sprouted grain, respectively (Official Grain Grading Guide: https://www.grainscanada.gc.ca/en/grain-quality/official-grain-grading-guide/). Kernels are considered as sprouted when there is clear evidence of germ growth on the kernel, splitting of bran due to apparent germ growth, discoloration of kernel in cases when germ is absent, and distinctive germ swelling. Kernels are considered as severely sprouted when the germs extend beyond the seed-coat, severely degenerated kernels exist due to the advanced stage of sprouting, only a portion of broken sprout remains on the kernel, the sprout has disappeared completely with an evidence of extension.

2.9. Effect of PHS on grain quality and end-use application

Moist and humid conditions throughout the harvest season cause PHS, which results in loss of grain quality and limited end-use applications. Major effects of PHS include reduction of test weight, changes in kernel properties such as bulk density, colour, size, shape, and contents of proteins and phenols, and loss of visual quality of grains. The degree of PHS damage ultimately determines the grade of wheat grains (McCaig et al. 2006; Official Grain Grading Guide: https://www.grainscanada.gc.ca/en/grain-quality/official-grain-grading-guide/).

Starch constitutes 64-74% of the dry weight of a wheat grain (Hucl and Chibbar 1996) and degradation of this starch by hydrolytic enzymes induced by PHS causes quality losses, limiting the end-use applications of wheat. For example, flour from sprouted wheat grains produces sticky dough, which is difficult to handle during the bread making process, and noodles prepared from sprouted grains are weak in strength and less bright in colour with the appearance of spots, making it unattractive to customers (Hatcher and Symons 2000). Moreover, significant reduction in gluten strength can occur due to sprouting, which negatively affects the production of wheat-based products (Barbeau et al. 2006). It has been shown that *de novo* synthesis of α-amylase occurs between scutellum and aleurone layer during germination of PHS susceptible wheat varieties (Olaerts et al. 2017). Previous studies reported up to 40-fold increase of α-amylase activity during germination of PHS susceptible varieties of wheat (Noda et al. 2003; Olaerts et al. 2016) and the activity of α-amylase increases with the degree of sprouting with over 80% α-amylase activity detected in severely sprouted grains (Olaerts et al. 2018). As a consequence, severely sprouted grains lose their suitability for human consumption and are considered for animal feed, reducing their market value significantly.

2.10. Seed germination

Seed germination is a complex process that initiates with imbibition (water uptake) and ends with extension of the radicle through the seed coat (Bewley 1997). During imbibition, mature seeds uptake water through seed coat specifically at micropyle end of the seeds (Rathjen et al. 2009). The process of germination is divided into three phases based on water uptake, phase I - the early phase/rapid imbibition, phase II- the middle phase/plateau phase and phase III- the later phase/exponential phase (Weitbrecht et al. 2011). During phase I, dry seeds are subjected to imbibition and absorb water rapidly resulting in hydration of germs. This phase is crucial for

transition from a silent metabolic state of dry seed to active metabolic state and continues until the very early stage of the second phase, which is characterized by a slow and/or constant water uptake by the seed (plateau phase). Radicle protrusion through the seed coat is visible by the end of the second phase. Phase III is a post-germination phase that is characterized by another rapid water uptake event associated with seedling growth (Ma et al. 2017; Gianinetti et al. 2018). Several physical and chemical changes occur during the entire germination process including seed coat/testa rupture, repair of cellular organelles and DNA, and gene transcription and translation (Ma et al. 2017). Imbibition of seeds also triggers the synthesis of biologically active GA which plays an important role for regulating the activity of α-amylase (Olaerts and Courtin 2018).

2.11. Seed dormancy

PHS resistance is determined by the level of seed dormancy (Mares 1993), which is defined as the inability of a mature and viable seed to germinate under favourable conditions of germination (Bewley 1997; Finkelstein et al. 2008). Dormancy is induced during seed maturation and its maintenance depends on genetic factors and environmental factors during the later stages of seed development (Olaerts and Courtin 2018). Genetic loci associated with high degree of seed dormancy in wheat attribute to higher PHS resistance (Martinez et al. 2016).

Based on the physiological and morphological properties of the seed, dormancy is classified as physiological dormancy (PD), morphological dormancy (MD), morphophysiological dormancy (MPD), physical dormancy (PY) and combinational dormancy (PY+PD) (Baskin and Baskin 2004; Nikolaeva 2004). Each of the five classes has many levels and types of dormancy. Among the five classes, PD is the most common class in crops and can be broken by genetic and environmental factors including cold treatment, after ripening and hormonal engineering/treatments. Morphological dormancy is common in underdeveloped embryo while MPD occurs due to a combination of PD and MD. The type of dormancy induced due to the physical strength of seed coat/seed covering structures, which restricts water and gas movement into the seed, is referred to as physical dormancy (Finch-Savage and Leubner-Metzger 2006; Penfield 2017). Dormancy can be further classified into primary and secondary dormancy based on the time of induction. When dormancy induces into seed before dispersing from mother plant due to physical barriers, hormonal regulations and/or other seed related

factors, it is termed as primary dormancy. Whereas, the dormancy reintroduced due to unfavourable environmental conditions is called secondary dormancy (Bewley 1997; Finkelstein et al. 2008). Seed dormancy, based on its cause, can be further divided into two types: embryo dormancy and seed coat imposed dormancy. Embryo dormancy is due to physiological/biochemical block that prevents the growth of embryo and germination. Seed coat imposed dormancy is characterized by physical properties of seed including hard and impermeable seed coat that block entry of water and gas to embryo (Heydecker 1973). The maintenance and release of dormancy depends on the seed's interaction with its surrounding environments including light, temperature, and moisture (Finch-Savage and Leubner-Metzger 2006).

2.12. Regulation of seed dormancy and PHS resistance

Seed dormancy, an important genetic trait for controlling PHS, is regulated by number of genetic, hormonal and environmental factors (Ali et al. 2019; Klupczyńska and Pawłowski 2021). Maintenance of a certain level of seed dormancy is essential in crop production to attain a good balance between achieving uniform seed germination and seedling establishment, and preventing PHS (Gao and Ayele 2014; Tuan et al. 2018). Phytohormones particularly ABA and GA play vital roles in the imposition and release of seed dormancy in cereals (Tuan et al. 2018). Dormancy status of a seed can also be affected by several environmental factors including temperature, humidity, light, oxygen level and nitrate (Finch-Savage and Footitt 2017; Klupczyńska and Pawłowski 2021). Other factors that influence seed dormancy include seed/grain structure, seed coat colour and secondary metabolites other than phytohormones.

2.11.1. Hormonal regulation of seed dormancy and PHS resistance

Seed dormancy and germination are two important biochemical and physiological processes which are controlled by plant hormones, predominantly ABA and GA (Hilhorst and Karssen 1992; Omoarelojie et al. 2021).

2.11.1.1. ABA metabolism and signaling in regulation of seed dormancy

ABA is fundamental for dormancy establishment and maintenance, and its biosynthesis is catalyzed by several enzymes including zeaxanthin epoxidase (ABA1/ZEP), carotenoid cleavage

deoxygenase (NCED), short-chain dehydrogenase reductase/short-chain alcohol dehydrogenase (ABA2/SDR1) and abscisic aldehyde oxidase 3 (AAO3) (Nambara and Marion-Poll 2003).

In the ABA biosynthetic pathway, oxidative cleavage of 9-*cis*-neoxanthin and violaxanthin by 9-cis-epoxycatrotenoid deoxygenase (NCED) is the rate limiting step (Tuan et al. 2018). Previous studies reported the role of different family members of the *NCED* in regulating seed dormancy including *NCED5*, *NCED6* and *NCED9* genes in Arabidopsis and *NCED1* and *NCED2* in cereals including wheat (Frey et al. 2012; Son et al. 2016; Skubacz and Daszkowska-Golec 2017). Besides ABA biosynthesis, ABA homeostasis is controlled mainly by the inactivation of bioactive ABA through 8'-hydroxylation or conjugation. ABA is catabolized through ABA 8′-hydroxylase (ABA8′-OH), which is encoded by cytochrome P450 (*CYP707A*) gene family, and this gene family in Arabidopsis consists of *CYP707A1*, *CYP707A2*, *CYP707A3* and *CYP707A4* (Skubacz and Daszkowska-Golec 2017). Of these, *CYP707A1* and *CYP707A2* have been well studied and their differential expression in various tissues plays an important role in regulation of seed dormancy. For instance, overexpression of *CYP707A1* leads to 3-fold reduction of ABA level and a decrease in dormancy level in mature seeds of Arabidopsis (Kushiro et al. 2004; Millar et al. 2006). Similarly, *CYP707A2* expression is predominant in dry seed and early stage of imbibition, and loss of function mutation in *CYP707A2* results higher accumulation of ABA leading to strong dormancy (Kushiro et al. 2004). In addition, *cyp707a1 cyp707a1* double mutant show strong dormancy in Arabidopsis as compared to other *cyp707a* mutants (Okamoto et al. 2006). Furthermore, *HvABA8'OH1* (*HvCYP707A1*) was highly expressed in the embryo of germinating seeds of barley while mutations in A and D genome copies of *TaABA8'OH1* showed strong dormancy during imbibition of wheat seeds due to higher level of ABA in embryo (Millar et al. 2006; Chono et al. 2006, 2013).

The ABA signaling pathway involves basic components including *pyrabactin resistance1/pyrabactin-like/regulatory components of ABA receptors* (*PYR/PYL/RCAR*), which upon binding with ABA inactivates the negative regulators, *protein phosphatase 2c*s (*PP2C*s), and thereby activates the positive regulators, *snf1-related protein kinase 2*s (*SnRK2*s) (Finkelstein et al. 2008; Tuan et al. 2018). Activation of the *SnRK2*s in turn activates *ABA insensitive*s (*ABI*s) including *ABI3*/*VP1*, *ABI4* and *ABI5* that directly regulates the ABA responsive genes (Tuan et al. 2018). Numerous ABA insensitive mutants of Arabidopsis such as *pyl*, *snrk*s, *abi3*, *abi4* and *abi5* show early germination phenotype, indicating the vital role of

ABA signaling in the induction and maintenance of seed dormancy (Nakashima et al. 2009). Previous studies have reported ABA resistance germination of *abi3, abi4* and *abi5* mutants in Arabidopsis (Söderman et al. 2000).

2.11.1.2. GA metabolism and signaling in regulation of seed dormancy

GA also plays important roles in regulating seed dormancy and germination; it promotes germination by regulating activity of α-amylase and expression of cell wall modifying genes (Linkies and Leubner-Metzger 2012; Tuan et al. 2018; Ali et al. 2019). The level of GA is determined by its biosynthesis and catabolism, which are regulated mainly by GA 20-oxidase (GA20ox) and GA 3-oxidase (GA3ox), and GA 2-oxidase (GA2ox), respectively. Previous studies have shown that higher expression levels of *GA20ox* and *GA3ox* are associated with increased bioactive GA level and therefore germination in after-ripened barley and wheat seeds (Gubler et al. 2008; Liu et al. 2013a; Kashiwakura et al. 2016). While GA inactivating genes, *GA2ox*s, are shown to be repressed in imbibed non-dormant seed of rice and sorghum, leading to increased bioactive GA level and promotion of germination (Tuan et al. 2018).

In GA signaling, the gibberellin insensitive dwarf 1 (GID1), a GA receptor protein, receives GA signal, which further activates GA-dependent MYB transcription factor, GAMYB, and GA responsive genes through degrading DELLA, a negative regulator of GA signaling, leading to activation of GA responses such as seed germination (Ali et al. 2019; Tuan et al. 2018). DELLA orthologues have been identified and characterized in cereals including wheat, and the wheat DELLA is designated as reduced height (RHT) (Ali et al. 2019). Mutation in GID of rice has been reported to result in reduced α-amylase activity (Ueguchi-Tanaka et al. 2005). In addition, transcriptional repression of *TaGID1* and *TaGID2* has been reported in dormant wheat seed developed in low temperature, suggesting reduction in seed sensitivity to GA (Tuan et al.

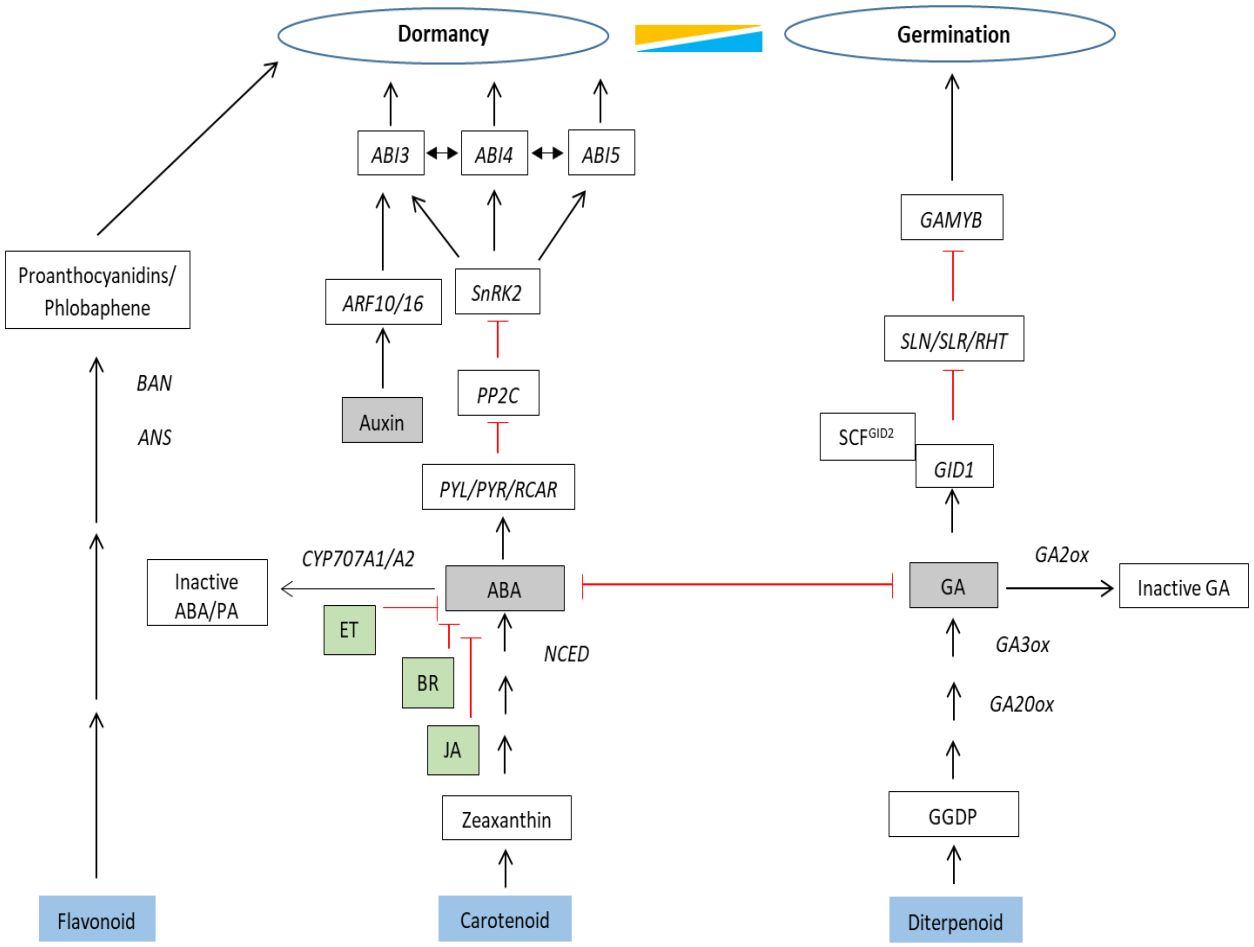


Fig. 2.2 Hypothetical regulatory mechanisms of hormones and secondary metabolites for controlling seed dormancy modified from Sohn et al. (2021) and Tuan et al. (2018). Grey boxes represent major hormones, olive green boxes represent other phytohormones and light blue boxes represent secondary metabolites. Black arrow represents positive interaction and red bar represents negative or repressive effect. ABA, abscisic acid; BR, brassinosteroid; ET, ethylene; GA, gibberellin; GGDP, geranylgeranyl diphosphate; JA, jasmonic acid; PA, phaseic acid. *ABI*, *ABA insensitive*; *ARF*, *auxin responsive factor*; *ANS*, *anthocyanidin synthase*; *BAN*, *BANYULS*; *CYP707A1/A2*, *cytochrome P450 monooxygenase* genes encoding ABA-8′-hydroxylases; *GA20ox*, *GA20 oxidase*; *GA3ox*, *GA3 oxidase*; *GA2ox*, *GA2 oxidase*; *GID1*, *gibberellin insensitive dwarf 1*; *GID2*, *gibberellin insensitive dwarf 2*; *GAMYB*, *GA regulated MYB transcriptional regulator*; *NCED*, *9-cis-epoxycatrotenoid deoxygenase*; *PYR/PYL/RCAR*, *pyrabactin resistance/pyrabactin-like/regulatory components of ABA receptors*; *PP2C*, *protein phosphatase 2c*; *SnRK2*, *snf1-related protein kinase 2*; *SLN/SLR/RHT*, *slender/slender rice/reduced height*.

2021). Inhibition of germination of non-dormant wheat seed was also shown to be associated with increased expression of *RHT1* (Izydorczyk et al. 2018). However, previous studies have also reported lack of positive correlation between expression pattern of GA signaling genes and level of seed dormancy in wheat, indicating the role of post-transcriptional regulation of GA signaling genes in dormancy and germination process (Tuan et al. 2018, 2021; Ali et al. 2019).

2.11.1.3. Other hormones associated with seed dormancy

Although seed dormancy and germination are regulated mainly by the balance between ABA and GA, other hormones such as auxin, ET, cytokinin (CK), BR, JA and salicylic acid (SA) are involved in regulation of seed dormancy and germination either directly or via modulating ABA/GA ratio in seeds (Fig. 2.2) (Sohn et al. 2021).

Auxin is involved in regulation of global events of plant growth and development including seed dormancy. Auxin is directly received by its receptor, transport inhibitor response 1/auxin signaling F-box protein (TIR1/AFB), and binds with Skp1-cullin-F-box (SCF) E3 ubiquitin ligase complex sequentially. Subsequently, interaction of TIR1/AFB with auxin/indole-3-acetic acid (Aux/IAA), the negative regulator of auxin signaling, allows ubiquitination and degradation of Aux/IAA, which in turn activates ARF that directly controls the expression of auxin responsive genes (Leyser 2018). The SCF E3 ubiquitin ligase complex mediated Aux/IAA degradation is controlled by several positive regulators of auxin signaling including auxin-resistant (AXR) and related to ubiquitin (RUB) (Calderon-Villalobos et al. 2010). It has been shown that ARF interacts with ABIs including ABI3, ABI4 and ABI5, the downstream transcriptional factors of ABA signaling, and this interaction increases ABA sensitivity and inhibits seed germination (Brady et al. 2003; Ali et al. 2019). For instance, *auxin response factor 10*/*16* (*ARF10*/*16)* positively regulates the expression of *ABI3* and enhances seed dormancy in Arabidopsis (Liu et al. 2013c). Transcriptional activation of *TaARF6* is proposed to induce seed dormancy in wheat through increasing auxin responsiveness (Tuan et al. 2019). Exogenous application of auxin also increases seed dormancy in wheat through a complementary role of ABA (Ramaih et al. 2003). Likewise, previous studies showed that the impaired ability of Aux/IAA degradation in auxin signaling mutants such as *tir1afb2*, *tir1afb3*, *axr2*, *axr3* causes enhanced germination/reduced dormancy (Kepinski and Leyser 2005; Liu et al. 2013c). However, the mechanisms by which auxin regulates seed dormancy and germination in wheat is still needs to be elucidated.

Ethylene inhibits seed dormancy through counteracting ABA effect and leads to promotion of germination, for example by activating the expression of cell wall modifying genes (Linkies et al. 2009; Linkies and Leubner-Metzger 2012). In presence of ethylene, *ethylene triple response 1* (*ETR1*) inactivates the negative regulator of ET signaling, *constitutive triple response 1* (*CTR1*), which activates downstream transcription factors *ethylene insensitive*s (*EIN*s) including *EIN2* and *EIN3* and the ET responsive genes (Rzewuski and Sauter 2008; Stepanova and Alonso 2009; Yoo et al. 2009) resulting promotion of germination (Wang et al. 2002; Corbineau et al. 2014). Ethylene signaling components such as *ETR* and *EIN2* have been implicated in the repression of ABA level in wheat and Arabidopsis seeds (Chiwocha et al. 2005; Chitnis et al. 2014; Sohindji et al. 2020). Numerous studies have reported that ethylene insensitive mutants (*etr1* and *ein2*) have been shown enhanced dormancy in Arabidopsis due to ABA sensitivity whereas enhanced germination has been observed in the seeds of *ctr1* mutants (Ghassemian et al. 2000; Chiwocha et al. 2005; Subbiah and Reddy 2010).

2.11.2. Environmental regulation of seed dormancy and PHS resistance

Environmental factors including temperature, light, nitrate/nutrient status and oxygen level regulates establishment of dormancy during seed development and maturation in wheat (Ali et al. 2019). Among all the environmental factors, temperature plays an important role in the control of dormancy (Kendall and Penfield 2012). Lower temperature during seed maturation increases seed dormancy in wheat, whereas low temperature during imbibition releases dormancy (Nyachiro et al. 2002; Gualano and Benech-Arnold 2009; Nakamura et al. 2011). Previous studies have shown that the *MFT* gene is involved in the regulation of dormancy induction in wheat seeds due to seed maturation at low temperature (Nakamura et al. 2011; Lei et al. 2013). High temperature also influences ABA/GA ratio through regulating the expression of genes associated with ABA and GA metabolism and signaling (Toh et al. 2008; Izydorczyk et al. 2017). For instance, upregulation of ABA metabolism genes *ZEP* and *NCED*s was observed in Arabidopsis due to high temperature while GA biosynthesis genes *GA20ox* and *GA3ox* was suppressed in same sample leading to induce seed dormancy (Toh et al. 2008). Similarly, Izydorczyk et al. (2018) studied temperature effect on seed dormancy and germination in wheat recently and reported that ABA signaling genes including *TaPYL5*, *TaSnRK2*, *TaABI3* and *TaABI5* showed higher expression in non-dormant imbibed seed under supra-optimal and

suboptimal temperature resulting delay of germination. While GA sensitivity was regulated through inducing *TaRHT1*, a negative regulators of GA signaling.

In addition to temperature, light also influences seed dormancy in cereals. For example, blue, white and far-red lights inhibit seed germination while red light promotes germination in barley and *Brachypodium* (Gubler et al. 2008; Barrero et al. 2012). Blue light-dependent *cryptochrome* genes *Cry1* and *Cry2* positively regulate expression of ABA biosynthetic gene *HvNCED1* and promote light-induced dormancy in barley (Barrero et al. 2014).

2.11.3. Grain structure in regulation of dormancy

The structural composition of a grain is conserved across cereal crops with some variations. Cereal grains in general are comprised mainly of an embryo, an endosperm and a pericarp fused with testa. In addition to the seed covering testa, some cereal grains contain a hull comprising the lemma, palea and glume that is removed by threshing in some cereals such as wheat and rice, and remain intact with the grains in others such as barley and oat (Rodríguez et al. 2015). These grain covering layers act as a physical barrier for entrance of water, air and some water-soluble chemicals conducive for germination into the seed/embryo.

The hull in some cereals grains such as barley and oat also contribute to inhibition of germination by limiting oxygen availability to the seed (Benech-Arnold et al. 2006; Hoang et al. 2013). The molecules in the leachate of the hull could also contribute to the inhibition of seed germination (Gatford et al. 2002). Furthermore, the testa of some plant species contains a waxy layer or cutin layer (a crosslinked fatty acid polymer) that restricts water permeability to the seed (Molina et al. 2008; Ranathunge et al. 2010; Bewley et al. 2013). For example, the outer layer of the endosperm of an Arabidopsis seed is surrounded by cutin layer, which is thicker than cuticular layer at leaf epidermis and therefore, proposed to play an important role in the regulation of seed dormancy and germination (De Giorgi et al. 2015).

2.11.4. Red grain colour and PHS resistance

Grain pericarp colour is an important genetic factor associated with flour quality and PHS resistance in wheat (Ali et al. 2019; Lang et al. 2021). Red grained wheat contains more phenolic compounds than white grained wheat and produces less bright flour with bitter taste (Kim et al. 2006). In immature grains of wheat, grain colour correlates positively with the amount of

phenolic compounds such as catechin and catechin tannin deposited in the pericarp (Miyamoto and Everson 1958). Proanthocyanidins (PA) is another phenolic compound found in the pericarp of immature seeds and acts as a precursor for the synthesis of anthocyanidins (Mccallum and Walker 1990). Both catechin and PA have been produced via the flavonoid biosynthesis pathway and are converted to red phlobaphene and anthocyanidins in mature seed at maturity (Ali et al. 2019).

It has been shown that white grained wheat is more susceptible to PHS than red grained wheat (Flintham 2000; Warner et al. 2000; Himi et al. 2002). The red grain colour gene, designated as *R* gene, located on group 3 chromosome controls the grain colour in wheat through transcriptional activation of the flavonoid biosynthesis genes (Ali et al. 2019). This *R* gene has been mapped on the same region of Myb-type transcription factor loci (*TaMyb10-A*, *TaMyb10-B* and *TaMyb10-D*) (Himi and Noda 2005; Himi et al. 2011) and reported as a causal gene for red pericarp pigmentation and therefore PHS resistance. In support of this, previous genetic studies have reported co-localization of PHS resistance QTLs and red grain colour QTLs in wheat (Kumar et al. 2009; Santos et al. 2010; Lin et al. 2016). Among the three homeologs of *TaMyb10*, *TaMyb10-B* controls grain colour significantly and is strongly associated with PHS resistance (Vetch et al. 2019). A recent study also reported the importance of *TaMyb10-D* in regulating red grain colour and PHS resistance in wheat (Lang et al. 2021). Red pigmentation in the seed coat, due to increased expression of flavonoid biosynthesis genes, forms a barrier of water uptake and therefore influences seed dormancy positively while increased expression of the ABA biosynthesis gene *NCED* promotes ABA biosynthesis, which increases ABA level and therefore delay germination in wheat (Lang et al. 2021). In agreement, mutation in *TaMyb10* led to a change of seed coat colour from red to white and an increase in the rate of dormancy release (Mares and Himi 2021). However, the molecular mechanisms linking red grain colour and PHS resistance are still unclear.

2.11.5. Secondary metabolites and seed dormancy

Secondary metabolites are natural organic products produced in plants through secondary metabolism. Although they are not directly required for plant growth and development (Li et al. 2016), secondary metabolites have important roles in a number of traits such as disease resistance and seed dormancy and germination (Keswani et al. 2014). Among the many

secondary metabolites produced by plants, phenylpropanoids, diterpenoids, carotenoids and flavonoids are known for their regulation of seed dormancy and germination as well as PHS resistance (Ali et al. 2019; Stanley and Yuan 2019; Eksi et al. 2020; Wang et al. 2021).

A recent study has identified a receptor-like kinase (*RLK*) gene, which is involved in phenylpropanoid biosynthesis within PHS resistance QTL on chromosome 3A of wheat (Wang et al. 2021). Phenylpropanoids are also reported to be involved in the regulation of primary dormancy in Arabidopsis seeds (Stone et al. 1994). Diterpenoids are a large class of secondary metabolites produced by plants (Hu et al. 2021). Diterpenoid metabolism is associated with the early steps of the biosynthesis of GA, a major phytohormone that promotes dormancy release and germination (Zi et al. 2014), whereas, carotenoids are critical precursors for the biosynthesis of ABA, which induces to seed dormancy (Standly and Yuan 2019). However, carotenoids are reported to protect seeds from damage by reactive oxygen species (ROS) and promote germination (Howitt and Pogson 2006). Flavonoids are important secondary metabolites for providing red pericarp pigmentation in wheat, which contributes to PHS resistance, as discussed above.

2.12 QTL mapping for PHS resistance

PHS is expressed as a quantitatively inherited trait which is governed by either multiple genes or alleles of a single gene (Kulwal et al. 2010; Vetch et al. 2019) and is strongly affected by environmental factors (Anderson et al. 1993). Thus, it is imperative to identify QTLs for PHS resistance for unrevealing genetic mechanism of PHS resistance. Genetic mapping of QTLs for PHS resistance requires a segregating mapping population for reliable phenotypic data and genotyping of the population with appropriate genetic markers. Several factors detailed below are needed for mapping PHS resistance QTL.

2.12.1. Sources of PHS resistance

Available sources for PHS resistance vary with wheat growing regions. In Canada, RL4137, which is a red grained wheat breeding line that is highly resistant to PHS has served as a source of PHS resistance in germplasm such as ‘AC Domain’, ‘Columbus’ and CN10955 (Matus-Cádiz and Hucl 2004; Ogbonnaya et al. 2008; DePauw et al. 2009, 2012). Irrespective of environments, RL4137 exhibits resistance to PHS consistently, and CWRS wheat cultivars with RL4137 in

their ancestry show lower PHS score than those cultivars that do not have lineage with RL4137. Red grain colour associated PHS resistance loci have been identified on chromosome group 3 of RL4137, confirming the role of grain colour in PHS resistance (DePauw et al. 2009). 'AC Domain', a red grained cultivar, is another well-known source of PHS resistance, and a double haploid (DH) population derived by crossing 'AC Domain' with PHS susceptible genotype RL4452 has been used for genetic analysis of PHS resistance in wheat (Cabral et al. 2014). Extremely dormant germplasm OS21-5 derived from 'Tordo' also serves as a good source of PHS resistance in Japan (Derera 2018). In addition, 'Gifu-Komugi', 'Zenkouji-komugi' (Zen) and OS38 have been used for genetic study of PHS and develop PHS resistance wheat cultivars (Hoshino et al. 1989; Osa et al. 2003; Mori et al. 2005; Osanai et al. 2005). For example, 'Zen' was crossed with Australian white wheat cultivar 'Spica' to develop a recombinant inbred line (RIL) population to identify QTLs and improve PHS resistance in white grained wheat (Kottearachchi et al. 2006). Another PHS resistance cultivar 'AUS1408' also known as 'Transval', a South African landrace, reported as a major PHS resistance allele contributor in several populations (Mares et al. 2005; Tan et al. 2006; Ogbonnaya et al. 2007; Lin et al. 2018). Other species like *Aegilops tauschii* (Xiu-Jin et al. 1997) and *Triticum spelta* (Zanetti et al. 2000) could also serve as sources of PHS resistance in wheat; however, introgression could be an issue. For example, a PHS resistance QTL in *T. spelta* is associated with ear morphology which is undesirable for cultivated wheat (Zanetti et al. 2000).

2.12.2. Mapping populations

Populations derived from experimental crosses are important tool for genetic mapping to dissect the genetics of complex traits and enhance plant breeding (Scott et al. 2020). A segregating population is the pre-requisite for QTL mapping derived from parents with contrasting phenotype of the trait of interest. Several distinct type of mapping populations are used to study heritable quantitative variations in crops species including F_2 populations, backcross (BC) populations, DH, near isogenic lines (NIL), RIL, nested association mapping (NAM) population and multi-parent advanced generation inter-cross (MAGIC) population (Boopathi 2020; Scott et al. 2020). Of these, F_2 and BC populations are considered as temporary populations as their propagation is indefinite. They are easy to construct and highly heterozygous, and are used to construct preliminary linkage maps (Boopathi 2020). All other mapping populations are

considered as permanent and immortal populations as they are propagated without changing their genetic constitutions over generations. In addition, DH, RIL and NIL populations are bi-parental while NAM and MAGIC populations are multi-parental (Scott et al. 2020). Among the bi-parental mapping populations, DH populations are widely used for QTL mapping due to the short period of time needed to generate the populations and their high power to detect QTL. Owing to these advantages, DH populations have been used for developing a consensus map of bread wheat (Wang et al. 2014) and winter triticale (Tyrka et al. 2018). They are also used to detect QTLs for PHS resistance in wheat (Cabral et al. 2014) and structural diversity in chromosomes of wheat (Gutierrez-Gonzalez et al. 2019), and novel loci for controlling spikelet number in bread wheat (Li et al. 2021).

2.12.3. Genetic markers for genetic mapping in plants

Genetic markers are important tools in plant breeding that have been used to characterize diversity in plants and improvement of crops (Kadirvel et al. 2015). They are distinguishable in morphological traits or protein/DNA and serve as landmarks for detecting trait associated QTLs. Genetic markers are of two types such as classical markers and DNA markers (Moose and Mumm 2008). DNA markers are fragments of DNA on known chromosomal position of an organism while classical markers are morphological, cytological and protein markers (Jiang 2013). Morphological markers such as seed coat colour and plant height could be used in breeding program when they are linked to the trait of interest (Kadirvel et al. 2015). However, the number of morphological markers is limited and sometimes has undesirable effects on plant growth and development (Jiang 2013). Cytological markers are related with chromosomal number, size and banding pattern, and they are used to study genetic and genomic relationship between organisms through identifying chromosomes as well as detecting mutation on the chromosome (Santos et al. 2006). Cytological markers have been used to distinguish *Aegilops* chromosomes in *Aegilops-Triticum* introgressions and the utilization of C-banding pattern of such markers has been reported (Kwiatek et al. 2019); C-banding pattern is used to detect pericentric inversion/chromosomal recombination since the size of C-band may differ between two individuals/groups (Wijayanto et al. 2005). However, application of cytological markers is laborious, expensive and less sensitive to study genomic constitution of hybrids. Protein markers are enzyme variants which are distinguished by their size and molecular weight. They are used to

initially map the chromosome and then the DNA markers are used to tap the gene, thus, protein markers are limited in their application in plant genetic research (Jiang 2013).

DNA markers, which are small polymorphic regions of DNA in a genome, are used to identify polymorphism between two individuals and alleles of a gene associated with that region (Jiang 2013). Based on throughput of the data generated, DNA markers can be grouped into low-throughput, medium-throughput, high-throughput and ultra high-throughput (Amiteye 2021). Restriction Fragment Length Polymorphism (RFLP) is the only low-throughput and first generation molecular/DNA marker (Grodzicker et al. 1975; Jones et al. 2009). Sequence-Tagged Site (STS) (Olson et al. 1989), Random Amplified Polymorphic DNA (RAPD) (Williams et al. 1990), Expressed Sequence Tags (EST) (Adams et al. 1991), Simple Sequence Repeats (SSRs) (Akkaya et al. 1992) and Amplified Fragment Length Polymorphisms (AFLPs) (Vos et al. 1995) are all medium throughput DNA markers, in which, SSR markers are widely used in plant science research. High-throughput markers include Kompetitive Allele-Specific PCR (KASPar) Assays, Diversity Array Technology (DArT), GoldenGate Assays and SNPs while ultra high-throughput markers include whole genome genotyping Infinium Assays and Genotyping-By-Sequencing (GBS) (Gupta et al. 2008; Amiteye 2021).

Among the high-throughput and ultra high-throughput markers, SNPs are widely used in genetic and association studies due to their high abundance and uneven distribution across the genome (Mariotti et al. 2020). Currently, preparation of SNP library/SNP discovery is easy and cheap due to low cost of next-generation sequencing, sequencing and re-sequencing of genomes/transcriptomes (Thudi et al. 2012). The sequence variation of SNPs is highly abundant in genomes of organisms, thus, gene derived SNP array technology is playing an important role in genotyping populations in genetic studies including association mapping, bi-parental mapping and genomic selections (Mourad et al. 2020). Several technologies have been used to develop high-throughput and multiplex SNP array including Illumina iSelect array and Affymetrix Axiom array. Using these technologies, several high-throughput SNP genotyping platforms have been developed including the Illumina Wheat 9K iSelect SNP array (Cavanagh et al. 2013), the Illumina Wheat 90K iSelect SNP array (Wang et al. 2014), the Wheat 15K SNP array (Boeven et al. 2016), the Axiom® Wheat 660K SNP array (Sun et al. 2020), the Wheat 55K SNP array (Liu et al. 2018) and the Axiom® HD Wheat 820K SNP array (Winfield et al. 2016), the Wheat Breeders' 35K Axiom array (Allen et al. 2017), the Wheat 50K Triticum TraitBreed array

(Rasheed and Xia 2019). Of these, the 90K SNP array of wheat was developed from almost 2,500 wheat accessions consisting of approximately 90,000 SNPs associated with genes. The wheat Infinium iSelect 90K SNP array is a standard genotyping technology that meets the need for studying complex traits due to its dense coverage of wheat genome as it comprises 81,587 SNPs (Wang et al. 2014).

2.12.4. Genetic linkage maps

Molecular markers associated with genes controlling a trait of interest can be detected through either linkage mapping based on bi-parental population or association studies of unrelated individuals (Mariotti et al. 2020). Construction of accurate linkage map with high genome coverage is the pre-requisite to localize QTLs and comparative genomic studies using bi-parental mapping population (Waiho et al. 2019). Genetic linkage map acts as a road map to locate the position of QTLs and/or genes on the chromosome and also indicate the interval distance between two markers (Collard and Mackill 2007). Genetic distances between markers, which is calculated using meiotic recombination fractions and recombinant genotypic frequencies, is used to order the markers on the chromosome (Paterson 2002; Collard et al. 2005).

The genetic distance between markers is measured in centimorgan (cM), which is obtained through converting recombination fraction to cM with a mapping function as recombination fraction and cM are not related linearly (Hartl and Jones 2009). In general, Kosambi and Haldane mapping functions are used for genetic mapping (Phillips et al. 2012; Kosambi 2016), in which, Kosambi is broadly used to adjust for interference between cross-over events (Hartl and Jones 2009). Logarithm of odds (LOD score) is used to determine the linkage between markers and positively correlate with accuracy of linkage map (Risch 1992; Collard et al. 2005). Several computer softwares including MapDisto 2.0 (Heffelfinger et al. 2017) MAPMAKER/EXP 3.0 (Lander et al. 1987) and JoinMap® 3.0 (van Ooijen and Voorrips 2001) are currently used to develop accurate genetic linkage maps.

RFLP markers were initially used to develop genetic maps from various mapping populations (Chao et al. 1989; Cadalen et al. 1997; Begna and Yesuf 2021). However, their application in linkage analysis has been limited due to the requirement of large quantities of DNA and they possess low level of polymorphism, and their application involves expensive and laborious processes (Begna and Yesuf 2021). For example, less than 10% polymorphism has

been reported in linkage analysis of interspecific crosses using RFLP markers (Röder et al. 1998). Advances in marker development technology have enabled the development and application of high-throughput, polymorphic, and abundant markers such as SSR/microsatellite, SNPs for linkage map construction. The first linkage map of wheat was constructed using 214 SSR markers (Röder et al. 1998). Linkage maps of 3,441 and 2,873 cM of wheat were also constructed using SSR and AFLP markers, respectively (Torada et al. 2006; Båga et al. 2007).

The markers currently used for linkage mapping include DArT, SNP and GBS (Semagn et al. 2006; Cabral et al. 2014; Li et al. 2015a). Using 90K SNP array, three high-density linkage maps of wheat consisting of 5,950, 4,861, and 8,376 SNPs have been developed using three different RIL populations (Liu et al. 2016). A linkage map containing 18,601 SNPs has also been developed using a MAGIC population (Gardner et al. 2016). In addition to SNP markers, a high density consensus map of 977.47 cM containing 9,793 DArT markers was developed using three different RIL populations of lentil (Ates et al. 2018). Using genotyping-by-sequencing (GBS), a linkage map of 1,959 cM consisting of 3,641 SNPs was generated using a RIL population of wheat (Hussain et al. 2017). A combination of different types of markers (12,201 SNP, SSR, DArT and EST markers) has also been used to construct a linkage map of 2569.4 cM using a DH population (Cabral et al. 2014). Currently, the Wheat 90K Infinium iSelect SNP array is widely used to develop high-density linkage maps of wheat (Wen et al. 2017). Therefore, SNP markers are popular tools to develop high density linkage maps and to identify QTLs for polygenic agronomic traits of wheat.

2.12.5. Mapping QTL

A QTL is a genomic region of an individual associated with phenotypic variations. QTL mapping is the genome-wide interference of a relationship between genotype and phenotype of a tested population through estimating genomic locations, numbers, phenotypic effect and gene action (Paterson 2002; Miles and Wayne 2008). The association between phenotype and genotype to detect QTL can be analyzed statistically using single-marker analysis and interval mapping (IM) (Tanksley 1993; Zhang et al. 2012). Of these, t-test based single-marker mapping is the simplest method of QTL detection and does not require a linkage map (Doerge 2002). In addition, analysis of variance (ANOVA) and linear regression are commonly used in QTL studies as they provide information about segregation of a population and QTL linked

phenotypic variation associated with markers (Shao et al. 2018; Rasul et al. 2009). However, single-marker analysis cannot estimate the accurate location of QTL on the chromosome. Thus, IM is used to identify QTL locations through estimating the linkage of both linked and unlinked genes on the chromosome (Zobaer et al. 2019).

Interval mapping based on maximum likelihood and regression is broadly used in mapping QTL of interested traits. This method requires marker based genetic linkage map and can estimate the location of QTLs. In principle, IM tests a model of presence of a QTL by scanning the interval between two adjacent marker loci. The fitness of the model correlates with maximum likelihood, which approximates the distributions of quantitative traits linked with the markers. The results are expressed as LOD scores and the significant LOD scores above the threshold value are reported as QTLs. Three types of IM exist which includes simple interval mapping (SIM), composite interval mapping (CIM)/multiple interval mapping (MIM) and inclusive composite interval mapping (ICIM) (Wang 2009; Zobaer et al. 2019). Of these, SIM considers each locus once and calculates the maximum likelihood ratio between intervals of two adjacent markers associated with logarithm of odds ratio/Pearson correlation coefficient between phenotype and markers to test the existence of QTL (Lander and Botstein 1989; Xu 1998; Kao 2000). The CIM and MIM represent multiple QTL models. Both methods estimate the location of QTLs more accurately and determine the effect of each QTL by scanning the whole genome (Zobaer et al. 2019). Several computer packages including QTL IciMapping V4.1 (Meng et al. 2015), QGene (Nelson 1997), MAPMAKER/QTL (Lincoln 1992) and MapQTL® 5 (van Ooijen 2004) are used for these mapping functions to detect QTLs on the chromosomal locations. However, the combined effects of two linked QTLs are difficult to estimate using CIM/MIM.

This problem associated with estimating the combined effects of two linked QTLs can be addressed by using another type of interval mapping called ICIM, a modified method of CIM. The ICIM can detect QTLs more efficiently and have lower rate of false discovery with almost unbiased estimation of QTL effect and locations (Li et al. 2015b). It can estimate additive, dominant and epistatic QTLs in a bi-parental mapping population more precisely (Zhang et al. 2008; Wang 2009; Li et al. 2015b). ICIM detects QTLs in two steps, which include selection of markers and adjustment of phenotypic value. Marker selection can be done through stepwise regression considering the information of all markers simultaneously. After that, the phenotypic values are adjusted for all markers except for those two markers flanking the current mapping

interval. In this way, ICIM reduces the background noises and increases the accuracy of estimating stable QTLs with their effect and location in multi-environmental trials (Zhang et al. 2012). Larger size of mapping population increases the accuracy of this method. For example, ICIM produces less biased estimation of QTL when the population size is 200 or more. Software packages such as QTL IciMapping V4.1 (Meng et al. 2015) are available to locate multiple QTL with additive effects.

For detecting QTLs for PHS resistance, several studies have used more than one method of assessing PHS including FN, GI, sprouting rate and index and compare these methods for effectiveness in the detection of QTLs (Singh et al. 2008; Rasul et al. 2009; Cabral et al. 2014). For example, comparison between GI and sprouting rate methods of PHS evaluation revealed variability in their effectiveness to detect QTL depending on a number of factors including genotypes, storage time, properties of inflorescence, and environmental factors such as temperature, light, oxygen and seed moisture levels (Paterson et al. 1989). Another study that compared sprouting test, GI and FN methods in diverse wheat line reported that sprouting test under rain simulator is an effective method of sprouting assessment for PHS susceptible materials; however, the accuracy and reliability of GI was found to be more than sprouting test and FN for assessment of PHS in the individual genotypes (Trethowan 1995). However, studies that used GI, sprouting index and FN to detect QTLs for PHS resistance found similar QTLs on chromosome 4A (Cabral et al. 2014; Rasul et al. 2009).

2.13 PHS resistance QTLs and their underlying genes

Significant efforts have been made to map QTLs and underlying genes for seed dormancy in wheat to understand better the genetic mechanisms of PHS and identify genetic markers for marker-assisted breeding of PHS resistance in wheat. To date, many PHS resistance and seed dormancy QTLs have been reported on almost all 21 chromosomes of bread wheat using different types of markers, populations, phenotyping methods and trial environments (Kulwal et al. 2010; Gao et al. 2013; Ali et al. 2019b; Liton et al. 2021). Among the previously identified QTLs, those on chromosome group 3 and chromosome 4AL have been reported as major QTLs due to their large effect on PHS resistance irrespective of genotypic differences of the mapping population and trial environments. Several causal and candidate genes have also been reported on those QTLs such as *TaMFT*/*TaPHS1*, *TaMKK3-A* and *TaVP1, TaMyb10, TaPM19-A1* and *A2*

(Barrero et al. 2015; Nakamura 2018; Ali et al. 2019b; Vetch et al. 2019). A recent study that involved the use of a DH population derived from 'Roblin'×RL4137 reported the contribution of up to 50% of phenotypic variation by a PHS resistance QTL on chromosome 4A (Liton et al. 2021). Whereas a PHS resistance locus on chromosome 3AS has been reported to contribute up to 38% of phenotypic variation in a RIL population derived from a cross between two red-grain parents (Osa et al. 2003). Due to their large effect on PHS resistance in wheat, the two loci on chromosome 3AS and 4AL have been subjected to fine mapping and transcriptomic analysis to clone/identify the underlying candidate/causal genes (Liu et al. 2011, 2013b; Nakamura et al. 2011; Barrero et al. 2015; Torada et al. 2016). As a result, some causal genes including *TaMKK3-A* (Torada et al. 2016), *TaMFT*/*TaPHS1* (Nakamura et al. 2011; Liu et al. 2013b), *TaVP1* (Yang et al. 2007; Chang et al. 2010) and *TaMyb10* (Lin et al. 2016) have been reported on chromosome group 3 and 4AL. Some other candidate genes including *TaPM19-A1/A2* (Barrero et al. 2015), *TaGA20ox* (Cabral et al. 2014) and *enhanced response to ABA 8* (*TaERA8*) (Martinez et al. 2020) have also been reported on chromosome 4A.

Among the other QTLs, those on chromosome 1D, 2B, 2D, 4B, 5D, 6A, 6B, 6D, 7A and 7D play an important role in regulation of PHS resistance and seed dormancy in wheat (Cabral et al. 2014; Kumar et al. 2015; Wang et al. 2019; Liton et al. 2021). Of these, *Qphs.umb-1D* is a newly identified PHS resistance QTL with 2.3% phenotypic variation (Liton et al. 2021). *Qphs.spa.-2B* and *Qphs.spa.-2D* have been shown to explain 9% and 14% of phenotypic variation in terms of GI, respectively, in DH population derived from a cross between SC8021-V2 and 'AC Karma' and thus reported as PHS resistance QTLs (Knox et al. 2012; Zhang et al. 2014a; Kumar et al. 2015). Moreover, a PHS resistance candidate gene, *seed dormancy* (*TaSdr*), has been identified to be located on chromosome 2B (Zhang et al. 2014a). A study on multiple pairs of NILs has led to the identification of a major PHS resistance QTL on chromosome 4BL, and the resistance allele of this QTL increased PHS resistance on an average by 70% in the studied population (Wang et al. 2019). Major QTLs that contribute to a decrease in PHS have also been reported on chromosome 5A, 5D, 6A, 6B, 6D, 7A and 7D (Kumar et al. 2015). Of these QTLs, the one detected on chromosome 5A was found to be sensitive to environment and independent of grain colour (Fofana et al. 2009) whereas another QTL on chromosome 6D was mapped near the ABA catabolism gene *TaABA8'OH1*. Two minor and consistent QTL have also

been reported on chromosome 6B and 6D by another study through evaluating PHS resistance under five different environments (Liton et al. 2021).

PHS resistance is also influenced by complex interactions between QTLs (Q×Q) and trial environments (Q×Q×E) (Ali et al. 2019; Kumar et al. 2015; Kulwal et al. 2010). Epistatic interaction between 1D, 4A, 6B and 6D QTLs to increase PHS resistance in 'Roblin'/RL4137 derived DH population has been reported recently (Liton et al. 2021). Another study that combined three PHS resistance QTLs (*Qphs.hwwg-3A.1*, *Qphs.hwwg-3B.1* and *Qphs.hwwg-5A.1*) to determine the effect of their combination on PHS resistance in a population derived from 'Danby' and 'Tiger', has shown a significant increase of PHS resistance due to Q×Q interactions. However, the effect of individual QTL and combined effect of QTLs in controlling PHS resistance in wheat are still unclear.

2.14 Important genes controlling PHS resistance in wheat

As a follow up to QTL analysis of PHS resistance, enormous efforts have been made to identify the associated genes. As the result, several genes associated with PHS resistance have been identified (Table 2.2), and these genes facilitate our understanding of the genetic mechanisms for controlling PHS resistance in wheat. Among the identified genes, some genes that have large impact on PHS resistance including *TaMFT*/*TaPHS1*, *TaMKK3-A* and *TaVP1* (Table 2.2) are considered as causal genes and their roles have been validated through multiple studies (Ali et al. 2019b; Vetch et al. 2019; Carrillo-Barral et al. 2020). While other genes those are believed to have both major and minor effects including *TaPM19-A1* and *A2*, *TaSdr* and *TaERA8* are considered as candidate genes since their roles have not been validated in detail.

Table 2.2 Important genes controlling PHS resistance in wheat

Gene name	Short name	QTL/ Chromosome	Methods	References
MITOGEN-ACTIVATED KINASE KINASE 3	*TaMKK3-A*	Phs1 locus on chromosome 4AL	Map-based approach (Fine mapping followed by molecular mapping)	Torada et al. 2016
ABA INDUCED PLASMA MEMBRANE 19	*TaPM19-A1* and *A2*	4A-1 locus on chromosome 4AL	Genetic study using MAGIC population along with transcriptomic analysis using RNA-Seq	Barrero et al. 2015
VIVIPAROUS 1	*TaVP1*	Homoeologous chromosome group 3AL	Southern blot analysis	Bailey et al. 1999
MOTHER OF FT AND TFL1	*TaMFT*	*QPhs.ocs-3A.1*	Transcriptomic analysis using microarray followed by genetic analysis	Nakamura et al. 2011
PROPYZAMIDE-HYPERSENSITIVE 1	*TaPHS1*	*Qphs.pseru-3AS*	Comparative mapping and map-based cloning	Liu et al. 2013b
SEED DORMANCY	*TaSdr*	Homoeologous chromosome group 2AS	Comparative genomics	Zhang et al. 2014a
ENHANCED RESPONSE TO ABA 8	*TaERA8*	*TaMKK3* locus of chromosome 4AL	BSA-exome-seq[a]	Martinez et al. 2020
ALANINE AMINOTRANSFERASE	*TaQsd1*	Homoeologous chromosome group 5	Genomic cloning and transcript sequencing	Onishi et al. 2017
GIBBERELLIN 20 OXIDASE	*TaGA20ox*	Chromosome 4AL	Comparative genomics	Cabral et al. 2014

[a]BSA-exome-seq, bulk segregant analysis-exome sequencing

2.14.1. TaMFT/TaPHS1

TaMFT, also known as *TaPHS1*, is an important causal gene for seed dormancy and PHS resistance reported on *Qphs.ocs-3A.1* of chromosome 3AS (Nakamura et al. 2011). This gene was identified through transcriptomic study on the effect of temperature on seed dormancy and showed upregulation in dormant seeds matured under low temperature. The observed differential

expression of this gene was shown to be associated with the presence of a SNP at -222 in the promoter region (Nakamura et al. 2011). A subsequent study cloned this gene at the same locus using map-based cloning and reported the presence of two SNP mutations at +646 and +666 position of the coding region of the gene, resulting in a premature stop codon that produces a truncated protein. Moreover, a 33-bp insertion in the promoter region of the gene (-194), which acts as a binding site of APETALA 2 transcription factor, was reported and this impacted its expression level (Jiang et al. 2018). It is hypothesized that *TaMFT* is a positive regulator of ABA sensitivity, and this was validated through mutational studies (Nakamura et al. 2011; Liu et al. 2013b; Vetch et al. 2019). Among the three homologs, *TaMFT-3A* is considered to be the most important one in the regulation of PHS resistance, although other homologs encoded nearly identical proteins. Further study is needed to investigate the contribution of each homeolog of *TaMFT* in PHS resistance in wheat.

2.14.2. TaMKK3-A

Several previous studies detected *Phs1* as a major locus for PHS resistance on chromosome 4AL (Shorinola et al. 2016; Torada et al. 2016; Nakamura 2018; Ali et al. 2019; Vetch et al. 2019). Using a map-based cloning approach, Torada et al. (2016) identified *TaMKK3* as a causal gene of the *Phs1* locus. Fine mapping of the large-effect 4A QTL also confirmed the presence the *TaMKK3* at this locus (Shorinola et al. 2016). *TaMKK3* encodes a mitogen activated protein kinase kinase 3 and plays a vital role in phosphorylation of proteins in many signaling pathways including ABA signaling, pathogen resistance signaling, jasmonic acid signaling, blue light signaling, and osmotic stress signaling (Nakamura 2018; Vetch et al. 2019). In ABA signaling, phosphorylation of *PP2C*s activates *SnRK2*s that positively regulates the downstream signaling components. Thus, *TaMKK3* enhances PHS resistance in wheat through positively regulating ABA responsiveness (Torada et al. 2016). A SNP identified at 660 bp downstream of the initiation codon of *TaMKK3* has been shown to cause nonsynonymous allelic substitution (missense mutation) in the kinase domain, resulting in increased PHS resistance (Torada et al. 2016; Vetch et al. 2019; Martinez et al. 2020). In addition, the level of PHS resistance in wheat genotypes varies with the two alleles of *TaMKK3-A* such as *TaMKK3-Aa* or *TaMKK3-Ab*, (Shorinola et al. 2017). For example, *TaMKK3-Aa* has higher positive effect on PHS resistance since the genotypes containing *TaMKK3-Aa* allele shows lower GI value (Yiwen et al. 2022).

The effect of *TaMKK3-A* on PHS resistance was confirmed by evaluating the individual and combined contributions of *TaMKK3-A* and *TaPHS1* (Lin et al. 2018), and this study reported that no single gene has adequate power to protect wheat from PHS due to genotype-environment interaction, however, their combined/additive effect significantly reduced PHS in most of the tested environments.

2.14.3. TaVP1

TaVP1 is a wheat homolog of maize *Viviparous-1* gene and an ortholog of *ABI3* (Yang et al. 2007; Vetch et al. 2019). It was identified approximately 30 cM apart from the *R* loci on the long arm of chromosome group 3 (Bailey et al. 1999). The *VP1* gene plays important roles during late embryogenesis in maize and the later stage of embryo development in bread wheat (McCarty et al. 1991). It has been shown the induction of *VP1* increases seed ABA sensitivity, which contributes increased level of seed dormancy (Nakamura and Toyama 2001). Extensive variations in the sequence of *TaVP1*, except in the *TaVP-1D* homeolog, have been observed in modern wheat cultivars (Yang et al. 2014). However, all the variations/alleles are not functional and some of them produce truncated proteins through mis-splicing (McKibbin et al. 2002). At least five alleles of *TaVP-1A* and *TaVP-1B* designated as *TaVP-1Agm*, *TaVP-1Aam*, *TaVP-1Aim*, *TaVP-1Ba*, *TaVP-1Bb* and *TaVP-1Bc* have been found to be associated with PHS resistance (Ali et al. 2019; Yang et al. 2014). Stacking of these alleles in a single genotype incrementally increases PHS resistance (Vetch et al. 2019). An increase in PHS resistance has also been reported due to the combined effect of *TaVP-1Agm*/*TaVP-1Ba*, *TaVP-1Agm*/*TaVP-1Bb*, *TaVP-1Aam*/*TaVP-1Bb* and *TaVP-1Aim*/*TaVP-1Bb* (Yang et al. 2014).

2.14.4. TaPM19-A1 and *A2*

Two adjacent PHS resistance candidate genes, *TaPM19-A1* and *TaPM19-A2*, have been identified as main candidate genes on *Phs-A1* locus controlling seed dormancy in Australian wheat germplasm through genetic analysis using MAGIC population followed by transcriptome and gene functional analysis (Barrero et al. 2015). This study showed that variation in the expression level of these two genes is associated with seed dormancy. However, their location and function relative to *TaMKK3* is still unclear (Shorinola et al. 2016). This is because although both the *TaPM19-A1* and *TaPM19-A2* genes were reported to be located 0.3 cM distal of the

Phs-A1 locus in the UK wheat germplasms, they were not detected as the main PHS/seed dormancy candidate genes underlying *Phs-A1* (Shorinola et al. 2016). In addition to genetic analysis, characterization of *TaPM19-A1* through ectopic expression in Arabidopsis has reported higher expression of *PM19L1* in seeds of dormant C24 ecotype of Arabidopsis (Barrero et al. 2019). On the other side, T-DNA insertion lines/mutant lines (*pm19l1-1*, producing a knock-out allele) showed higher dormancy than wild-type Col-0 indicating its negative role in seed dormancy which is in contrast to the previous results. The discrepancy between the results of the two experiments can be explained by either species specific expression of that gene or influence of other ABA responsive genes to increase dormancy level in mutants (Barrero et al. 2019).

2.14.5. TaSdr

TaSdr is an ortholog of *OsSdr4*, a gene associated with seed dormancy in rice; however, the exact function of *TaSdr* in wheat is unknown (Vetch et al. 2019). In rice, *OsSdr4* is directly regulated by *OsVP1* and in turn regulates *OsDOG1* (Sugimoto et al. 2010). *TaSdr* has been identified on homoeologous chromosome group 2 through comparative genomic approach (Zhang et al. 2014a). Of the three homoeologs of *TaSdr*, *TaSdr-A1* and *TaSdr-B1* have been cloned and found to have allelic variations. Among the identified alleles associated with seed dormancy and PHS resistance, *TaSdr-A1a* and *TaSdr-B1a* are reported to be prominent in wheat cultivars and land races in China, Japan and Australia (Zhang et al. 2014a, 2017). These studies also reported presence of SNPs at -11 bp and +643 bp of initiation codon of *TaSdr-A1* and *TaSdr-B1*, respectively, which control seed dormancy through regulating the expression of *TaSdr-A1* and *TaSdr-B1*.

2.14.6. TaQsd1

TaQsd1, an ortholog of barley *Qsd1*, has been identified recently on homoeologous chromosome group 5 of bread wheat (Onishi et al. 2017). In barley, *Qsd1* encodes alanine aminotransferase and it is hypothesized to be involved in grain colour formation, however, the exact mechanism of *Qsd1* in regulating dormancy in wheat and barley seeds is still unknown (Nakamura 2018). It has been shown recently that *TaQsd1* is represented by a gene with the International Wheat Genome Sequence Consortium (IWGSC) Gene ID of, TraesCS5A02G216200.1 (Mares et al. 2021). Of

the homeologs of *TaQsd1*, *TaQsd1-5B* shows higher level of expression in dormant genotypes that exhibit strong-dormancy (Wei et al. 2019).

2.14.7. Other candidate genes

Some genes such as *TaGA20ox* and *TaERA8* have been reported as PHS candidate genes on chromosome 4AL (Martinez et al. 2020; Cabral et al. 2014). The *TaGA20ox*, which is a well-known GA biosynthesis gene, has been mapped on PHS QTLs of *QGi.crc-4A* (Cabral et al. 2014). In agreement with this transcriptional regulation of *TaGA20ox1* and *TaGA20ox2* influences seed dormancy in wheat (Tuan et al. 2021). The *TaERA8*, which is also mapped on *TaMKK3-A* locus of chromosome 4AL, has been shown to regulate seed dormancy by modulating ABA sensitivity (Martinez et al. 2020). However, further study is needed to validate and understand the mechanisms through which *TaERA8* controls seed dormancy and consequently PHS resistance.

2.15 Genome-wide gene expression/transcriptomic analysis

Gene expression is the process of converting genetic information to observable characters/phenotype via gene transcription and translation of the transcripts (Ghorbani et al. 2018). It is a vital process causing cellular differentiation, morphogenesis, versatility and adaptability of any organisms (O'Connor et al. 2010). Thus, gene expression analysis provides information about transcriptional behaviour of an organism and is a commonly used method to study biological systems. Gene expression analysis is of two types based on intention of study: genome-wide gene expression and targeted gene expression analysis. When the gene of interest is known, targeted gene expression analysis can be performed using approaches such as qPCR and/or RT-qPCR (Murphy et al. 1990). When the gene(s) of interest is not known or when large-scale/genome-wide gene expression analysis is needed; approaches such as microarray (Augenlicht and Kobrin 1982) and/or RNA-Sequencing (Mortazavi et al. 2008) analysis can be applied. Genome-wide gene expression analysis examines the expression of all genes in a given genome through measuring the frequency of transcripts. Transcriptome, complete set of transcripts in a cell at a certain time, is the vital mediator between information encoded in the genome and phenotype (Wang et al. 2009).

2.15.1. Wheat genome and transcriptome

Understanding the molecular mechanisms of important agronomic traits has lagged in wheat due to its large genome size (approximately 17 Gb), presence of high amount of repetitive DNA sequences (>85%) in its genome and lack of complete reference genome sequence information of wheat (Appels et al. 2018; Walkowiak et al. 2020). The IWGSC has made enormous efforts since 2005 to provide genome-assembly information and released the first reference genome sequence of wheat in 2017 using Chinese Spring landrace, which was updated in 2021. The updated wheat reference genome sequence is a milestone to facilitate the study of the biology of important agronomic traits using advanced molecular techniques including comparative genomics and transcriptomics. It also provides a better understanding of the genomic composition and physical structure of wheat genome, which facilitates mapping of actively transcribed regions of the genome during plant growth and development under different environmental factors including biotic and abiotic stresses.

Transcriptome is the sum total of all messenger RNAs transcribed from genes of an organism, which in general is further translated to protein. Transcriptome analysis is widely used to uncover transcriptional behaviour of biological systems in living organisms and thereby assess gene functions at the genome level and can also identify novel genes controlling important traits. Every cell of multicellular organisms contains the same extent of genetic information and genes; however, every gene is not transcriptionally active at all times as gene expression varies with tissue, stage and the surrounding environmental conditions. Microarray has been used to perform large scale gene expression analysis in wheat for a long time as it was a cost-effective method to compare the expression of thousands of genes at a time (Wan et al. 2008; Zhang et al. 2014b). Previous microarray based studies have shown that 6,106 upregulated and 6,521 downregulated genes associated with DNA synthesis, carbohydrate metabolism, hormone metabolism and protein degradation were found between dormant/ABA treated and non-dormant wheat seeds during germination (Yu et al. 2016; Cheng et al. 2021). However, the application of this technology in transcriptome analysis has its own limitation mainly due to the limited number of probe sets/genes represented on the chip and the indirect measurement of gene expression through hybridization (Wang et al. 2009). With the advancement of next generation sequencing (NGS) technologies, transcriptome analysis through NGS has become an approach of choice. This is because high throughput sequencing technologies can characterize and quantify complete

set of RNAs in a given tissue accurately (Park et al. 2021; Sohn et al. 2021). As a result, RNA-seq based transcriptome analysis provides deeper insight into the molecular mechanisms regulating important traits (Li et al. 2022). To date 850 RNA-Seq samples of wheat from 32 different tissues collected from different stages of growth and development and in response to different stress factors have been made publicly available (https://www.ebi.ac.uk/gxa) and this resource facilitates gene function studies and analysis of complex gene regulatory network of complex biological processes (Jia et al. 2018; Appels et al. 2018).

The advancement of NGS technologies along with updated wheat reference genome sequence are providing wheat scientists the opportunity to facilitate positional cloning, analyze sub-genome dominance, and examine genetic and epigenetic modifications (Pfeifer et al. 2014). It also helps to identify molecular mechanisms, biological pathways and QTL localized genes for different traits (He et al. 2021; Wang et al. 2021). For example, transcriptomic analysis of *Brachypodiuym* leaf tissues infected with *Fusarium pseudograminearum* followed by homologous gene identification in the sub-genomes of wheat revealed the dominance of the B and the D sub-genomes (Powell et al. 2017). In addition, the prominent role of the D sub-genome in controlling grain weight and size in wheat is affected through altered gene expression. Similarly, the A sub-genome has been reported to be important for controlling different traits. For instance, several important genes for PHS resistance were reported on the A sub-genome of wheat including *TaMKK3-A*, *TaPM19-A1* and *A2*, *TaMFT* and *TaPHS1* (Nakamura et al. 2011; Liu et al. 2013b; Barrero et al. 2015; Torada et al. 2016). Of these, *TaMFT* and *TaPM19* were identified as QTL localized genes using microarray- and RNA-seq based transcriptomic analysis, respectively (Barrero et al. 2015; Nakamura et al. 2011).

2.15.2. Differential expression analysis of wheat transcriptomes

Transcript abundance is an indicator of gene expression; higher amount of transcripts of a gene indicates higher level of expression of that gene. In RNA-seq based genome-wide gene expression analysis, amount of transcripts is estimated by counting reads mapped on the gene or by Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced (FPKM) value or both units (Zhao et al. 2021). Read count also represents the length of gene and depth of sequencing while FPKM is used to compare level/pattern of gene expression in different samples in a given experiment (Trapnell et al. 2010). Genes differentially expressed between samples

provides important ideas in their role of controlling the trait of interest. A previous transcriptomic study in wheat has identified 14,550 genes as differentially expressed between two stages of wheat caryopsis development (6 days after anthesis vs. 42 days after anthesis), suggesting their involvement in wheat grain developmental processes (Wan et al. 2008). With respect to wheat seed dormancy, microarray-based studies have shown differential expression of hormone metabolism and signaling genes between dormant and after-ripened seeds, for example, overrepresentation of ABA-responsive genes in dormant seeds, suggesting the role of ABA-responsive genes in inducing and maintaining seed dormancy in wheat (Gao et al. 2012; Liu et al. 2013a; Chitnis et al. 2014). Differential expression of genes involved in flavonoid biosynthesis, L-phenylalanine metabolic processes and ABA transport have also been reported between PHS resistance and PHS susceptible wheat genotypes, indicating the importance of those DEGs in the regulation of seed dormancy (Lee et al. 2021; Wang et al. 2021). Gene differential expression analysis between samples can be performed at individual gene or gene set level to identify biological relevancy of the differentially expressed genes (DEGs) in controlling plant growth and development. Gene set level analysis has some advantages over individual level analysis (Khatri et al. 2012; Creixell et al. 2015) as it has better power of explaining the importance of the selected genes representing a biological processes or pathways. However, different forms of analysis that make use of established databases such as gene ontology (GO) and biological pathway (KEGG) analyses are pre-requisites to predict the function of gene sets in controlling biological important traits.

2.15.3. Gene ontology analysis

GO is a major bioinformatics initiative to unify the representation of gene and gene product attributes across all species (Ashburner et al. 2000). It is an important tool to organize biological terms in a hierarchical structure within three broad categories of gene function, which are cellular component, biological process and molecular function. In each category, the GO terms form a structure of parental node and child node, in which, child nodes are more specialized. Each GO term represents a set of genes that may fall under several parental GO terms since genes can be involved in several biological functions (The Gene Ontology Consortium 2021). Enrichment analysis of DEGs is an important tool to enable their functional classification under the three broad GO categories (The Gene Ontology Consortium 2021). A GO term is considered

as enriched when a set of DEGs from a given comparison are present more frequently in the gene list representing that specific GO term (Khatri et al. 2012). Enrichment analysis of DEGs between dormant and non-dormant seed of rice has shown that genes upregulated in dormant seeds are enriched in biosynthesis and polymerization of flavonoid and those upregulated in non-dormant seeds are enriched in mitochondrial electron transport. GO overrepresentation analysis between dormant and after-ripened Arabidopsis seed has reported that genes upregulated in dormant seeds are overrepresented in GO terms associated with response to biotic and abiotic stresses (Dekkers et al. 2016). A study in wheat showed that genes downregulated in imbibing after-ripened seeds are enriched in ABA response and response to stress GO terms while those upregulated are enriched in GO terms of carbohydrate metabolism, nitrogen metabolism, cytoplasmic membrane-bound vesicle and cell wall modification (Gao et al. 2012). A recent RNA-Seq based transcriptomic analysis in wheat reported that similar number of up- and downregulated genes in involved in PHS resistance near isogenic lines (NILs) at 15 and 25 days post anthesis (DPA), and these DEGs are enriched in cell, cell part, organelle and membrane part GO terms under cellular component; catalytic activity, binding and transporter activity GO terms under molecular function; and metabolic process, cellular process and single organism process GO terms under biological process (Wang et al. 2021). In addition, GO categories including flavonoid biosynthesis, trehalose biosynthesis, abscisic acid transport and response to oxidative stress have been shown to be enriched with genes upregulated in ABA treated PHS resistant wheat cultivar (Lee et al. 2021).

2.15.4. KEGG pathway analysis

Kyoto Encyclopedia of Genes and Genomes (KEGG) is the collection of genes and proteins in a genome from publicly available databases, which is used to identify interaction of multiple genes to regulate a certain biological function. It is also used to identify biological pathways associated with traits of interest through enrichment analysis. Using KEGG pathway analysis, previous studies reported the association of seed dormancy and germination with some biological pathways including the biosynthesis of secondary metabolites, plant hormone signal transduction, general metabolism, phenylpropanoid biosynthesis, protein processing in endoplasmic reticulum, photosynthesis, MAPK signaling, amino sugar and nucleotide sugar metabolism and spliceosome pathways (Kanjana et al. 2016; Zhang et al. 2019; Wu et al. 2020;

Wang et al. 2021; Li et al. 2022). For example, phenylpropanoid biosynthesis, flavonoid biosynthesis and amino sugar and nucleotide sugar metabolism pathways are overrepresented in genes upregulated by ABA treatment of quinoa seeds, while protein processing in endoplasmic reticulum was enriched in downregulated genes (Wu et al. 2020). General metabolic, plant hormone signal transduction, secondary metabolite biosynthesis and spliceosome pathways have been reported as biological pathways highly enriched with genes differentially expressed between PHS resistant and susceptible genotypes of rice and wheat (Li et al. 2022; Wang et al. 2021). Of these, plant hormone signal transduction and secondary metabolite biosynthesis pathways are enriched with genes upregulated in dormant seeds than their non-dormant counterparts suggesting their important role in inducing and maintaining seed dormancy. For example, genes involved in ABA metabolism and signaling pathways are overrepresented in gene sets upregulated in dormant as compared to non-dormant seeds (Ali et al. 2019). Previous studies have identified several genes in hormone metabolism and signalling pathways such as those involved in ABA and GA metabolism and signaling including *NCED*, *GA3ox*, *SnRK2*, *ABI3* and *ABI5* as differentially expressed between dormant and non-dormant seeds of different plant species such as Arabidopsis, wheat, rice and barley (An and Lin 2011; Gao et al. 2012; Nelson and Steber 2017; Wang et al. 2021; Li et al. 2022).

2.15.5. Integration of RNA-Seq on QTL data

Previous studies reported that combining QTL mapping with genome-wide gene expression analysis is a robust tool to detect candidate genes controlling complex trait in crops (Xu et al. 2015; Zhang et al. 2017; Jian et al. 2019). Integrating QTL mapping with gene expression in *Brassica napus* led to the identification of 45 flowering time-related genes including a key flowering gene, *pseudo response regulator 7* (*PRR7*), in the QTL region of flowering time (Jian et al. 2019). Utilizing this new approach, six genes that are associated with hormone signaling and energy metabolism, and showed differential expression (four upregulated and two downregulated genes) between contrasting NILs of wheat have been reported as candidate genes for a PHS resistance QTL *QPhs.ccsu-3A.1* (Wang et al. 2021). A recent study that utilized comparative transcriptomic analysis of *Aegilops tauschii–Triticum aestivum* chromosome segment substitution lines identified a gene encoding RING finger-dependent E3 ubiquitin-

protein ligase and located on chromosome 3DL of wheat as a novel candidate gene for PHS resistance (He et al. 2021).

Wheat breeding programs have made significant efforts to integrate a moderate level of seed dormancy/PHS resistance in commercial varieties worldwide. However, their success have been restricted due to unavailability of a complete genome sequence and therefore lack of sufficient information on the genetic and molecular basis of seed dormancy and PHS resistance in wheat. Thus, a better understanding of the genetic loci, molecular mechanisms and causal/candidate genes responsible for PHS resistance is critical to develop PHS resistant wheat cultivars through marker-assisted gene introgression and gene pyramiding.

CHAPTER 3

IDENTIFICATION OF LOCI FOR PRE-HARVEST SPROUTING RESISTANCE IN THE HIGHLY DORMANT SPRING WHEAT RL4137

M M Uzzal A. Liton[1], Curt A. McCartney[2], Colin W. Hiebert[2], Santosh Kumar[3], Mark C. Jordan[2], Belay T. Ayele[1*]

[1]Department of Plant Science, University of Manitoba, Winnipeg, MB, R3T 2N2, Canada

[2]Morden Research and Development Center, Agriculture and Agri-Food Canada, Morden, MB, R6M 1Y5, Canada

[3]Brandon Research and Development Center, Agriculture and Agri-Food Canada, Brandon, MB, R7A 5Y3, Canada

This chapter was published as Liton MMUA, McCartney CA, Hiebert CW, Kumar S, Jordan MC, Ayele BT (2021) Identification of loci for pre-harvest sprouting resistance in the highly dormant spring wheat RL4137. Theoretical and Applied Genetics 134, 113–124. https://doi.org/10.1007/s00122-020-03685-y

Author contributions: BTA and MCJ conceived, designed and supervised the research; UAL and CAM performed the experiments and analyzed the data; CWH generated the genotypic data; SK supervised the management of the field trial at Brandon; UAL wrote the first draft of the manuscript; BTA and CAM edited and revised the manuscript; and all authors read and approved the final manuscript.

Identification of loci for pre-harvest sprouting resistance in the highly dormant spring wheat RL4137

3.1. Abstract

Pre-harvest sprouting (PHS) significantly reduces wheat grain yield and quality. Therefore, identifying quantitative trait loci (QTL) for PHS resistance is key to facilitate marker-assisted breeding. To this end, we studied PHS in a population of 330 doubled haploid (DH) lines derived from 'Roblin'/RL4137. The parental and DH lines were examined for their PHS phenotype based on speed of germination index in five environments and genotyped using the wheat Infinium 90 K SNP array. A total of five QTLs were detected on linkage groups 1D, 4A.2, 6B.1, 6D and 7A over the five environments. The QTL *QPhs.umb-4A* on linkage group 4A.2 was the most consistent across all environments and explained 40–50% of phenotypic variation. The QTL on 1D is a novel QTL and explained 1.99–2.33% of phenotypic variation. The QTLs on 6B.1 and 6D each explained 3.09–4.33% and 1.62–2.45% of phenotypic variation, respectively. A combination of four stable QTLs on linkage groups 1D, 4A.2, 6B.1 and 6D greatly increased PHS resistance. Allelic effects for the QTLs *QPhs.umb-4A, QPhs.umb-6B* and *QPhs.umb-6D* were contributed by RL4137, whereas 'Roblin' contributed the resistant allele for *QPhs.umb-1D. QPhs.umb-4A* was required for strong dormancy in the 'Roblin'/RL4137 DH population, and the presence of QTLs *QPhs.umb-1D, QPhs.umb-6B* and *QPhs.umb-6D* incrementally increased dormancy; DH lines carrying all four QTLs are considerably more dormant than those carrying only *QPhs.umb-4A* or none of the four QTLs. Thus, the QTLs identified in this study have the potential to improve PHS resistance in spring wheat.

Keywords: Pre-harvest sprouting, double haploid population, speed of germination index, quantitative trait loci, SNP marker, seed dormancy

3.2. Introduction

Hexaploid wheat (*Triticum aestivum* L.) is the second most produced cereal crop worldwide. Its economic importance is reflected by the fact that over 40% of the world population uses it as their staple food, contributing approximately 20% of the total calories consumed by humans (Braun et al. 2010). Among the major factors that limit wheat productivity is pre-harvest sprouting (PHS), which refers to the germination of grains while on the spike due to the occurrence of wet and humid summer condition prior to harvest. This occurs because the seeds of many commercial cultivars do not have a sufficient level of dormancy, a trait that blocks the germination of seeds under apparently optimal conditions (Gao and Ayele 2014). The annual economic loss in wheat production due to PHS has been estimated to be ~ 1.2 billion USD worldwide (Shao et al. 2018). Therefore, it is necessary to identify genetic elements that can enhance the development of PHS-resistant wheat cultivars.

The progenitor species of hexaploid wheat exhibit high levels of seed dormancy. However, domestication of wheat has been associated with selection against strong dormancy in order to develop wheat cultivars with uniform and rapid rate of seed germination and seedling establishment (Tuan et al. 2018). This has made the modern wheat cultivars to be susceptible to PHS, which results in substantial grain quality losses (Vetch et al. 2019a). Therefore, the development of wheat cultivars with intermediate level of seed dormancy has been the focus of wheat breeding programs around the world.

PHS is a complex trait governed by different genetic and environmental factors (Gao and Ayele 2014). The inheritance of genetic components for PHS resistance has been studied using classical Mendelian and molecular approaches, and these studies reported that seed dormancy and PHS resistance are regulated by two or more independent genes (Mares 1996; Lawson et al. 1997). This is well supported by the findings of previous studies which detected QTLs/genes for PHS resistance on all the 21 chromosomes of wheat through utilizing diverse mapping populations and different PHS phenotyping methods (Mohan et al. 2009; Kulwal et al. 2010). However, some non-dormancy related factors such as spike morphology including spike structure and erectness, and tenacity of glumes act as physical barriers for water penetration into the spike/seed and thereby affect sprouting (King and Richards 1984). Thus, the use of a phenotyping method such as speed of seed germination index, which involves threshed seeds but not intact spikes, is important to determine the precise genetic mechanisms underlying seed

dormancy and thereby PHS resistance. Evaluation of PHS phenotype based on seed germination index along with the use of DNA markers has led to the identification of important QTLs for PHS resistance on chromosomes 3B (Fofana et al. 2009) and 4A (Ogbonnaya et al. 2008; Cabral et al. 2014).

Genetic dissection of traits is difficult in common wheat due to its complex genome (allohexaploid, 2n = 6x = 42, AABBDD), large genome size (~ 17 GB) and presence of higher amount of repetitive DNA sequences (Wu et al. 2015). Molecular markers play an important role in determining the genetic bases of traits of agronomic importance in wheat by facilitating the identification and tagging of genetic regions or quantitative trait loci (QTLs) consisting of one or more genes regulating the trait of interest (Semagn et al. 2006; Collard and Mackill 2007) through construction of high-density genetic linkage maps. Given that PHS is a quantitatively inherited trait (Rodriguez et al. 2015; Tuan et al. 2018; Wang et al. 2019), identification of QTLs for this trait and markers tightly linked to such QTLs is required to facilitate marker-assisted breeding of PHS-resistant cultivars.

Previous studies that involved the use of different mapping populations identified QTLs that affect PHS resistance/seed dormancy across all the 21 chromosomes of wheat (Kulwal et al. 2010; Vetch et al. 2019b; Wang et al. 2019). Among the different QTLs identified, detailed genetic analyses have been conducted with QTLs on group 3 chromosomes and chromosome 4AL as they have been reported to have major effects on PHS (Kulwal et al. 2010; Rodriguez et al. 2015; Vetch et al. 2019b). Some of the causal genes underlying significant QTLs on group 3 chromosomes and 4AL have been identified and their effects on regulating PHS/seed dormancy have been characterized, including *TaPHS1* (Nakamura et al. 2011; Liu et al. 2013b; Lin et al. 2018), *TaMKK3-A* (Shorinola et al. 2016; Torada et al. 2016; Shorinola et al. 2016), *TaVP1* (Yang et al. 2007) and *TaMyb10* (Lin et al. 2016; Wang et al. 2016).

RL4137 has been used as a source of PHS resistance in wheat breeding programs around the world, and particularly in Canada (Matus-Cádiz and Hucl 2004; DePauw et al. 2009, 2012). The objective of this study was to determine the genetic basis of seed dormancy in the highly dormant spring wheat breeding line RL4137 using a large doubled haploid (DH) population.

3.3. Materials and methods

3.3.1. Plant materials and trial environments

A mapping population consisting of 330 DH lines was developed from a cross between 'Roblin' and RL4137 using the maize pollination method described previously (Fedak et al. 1997). 'Roblin' is a former Canada Western Red Spring (CWRS) cultivar that is PHS susceptible. 'Roblin' was developed from the cross RL4302/RL4356// RL4359/RL4353 (Campbell and Czarnecki 1987). RL4137 is a widely used PHS-resistant breeding line that was developed from the cross RL2520//Thatcher*6/Kenya Farmer, and it is an ancestor of many PHS-resistant wheat cultivars (DePauw et al. 2009). The DH population was grown in five different environments with replicated trials over 3 years, in a greenhouse (2016, designated as GH16) and field sites located in Winnipeg (2017, designated hereafter as WPG17), Brandon (2017, designated hereafter as BRD17) and Morden (2017 and 2018, designated hereafter as MOR17 and MOR18). Field tests were conducted in alpha lattice designs with 19 incomplete blocks and 3 replications, and greenhouse test was conducted in randomized complete block design (RCBD) with 3 replications. Harvesting of the seeds was performed at their harvest maturity; when the peduncle of ~80% of the spikes turned yellow and the seeds were hard enough to be dented with thumbnail.

3.3.2. Assessment of PHS phenotype

In order to assess PHS phenotype in the mapping population, physiologically mature seeds harvested from the greenhouse and field trials were subjected to germination tests as described in Gao et al. (2012). Seeds were imbibed between filter papers in Petri plates at room temperature under darkness. Germination was monitored daily for 21 days and a seed was considered germinated when the coleorhiza emerged beyond the seed coat. Speed of germination index (SGI) was calculated according to Maguire (1962) using the formula shown below.

$$Speed\ of\ Germination\ Index\ (SGI) = \left[\frac{No.\ of\ germinated\ seeds}{Days\ of\ first\ count}\right] + \cdots\cdots + \left[\frac{No.\ of\ germinated\ seeds}{Days\ of\ final\ or\ last\ count}\right]$$

3.3.3. Genotyping and linkage mapping

The mapping population was genotyped using a 90 K Infinium iSelect Custom Wheat Beadchip (Wang et al. 2014). Genotype calling was performed using Genome Studio v.2.0.2 (Illumina).

Monomorphic SNPs, SNPs with > 10% missing data, and SNPs with three or more clusters were excluded. Genotypic data consisting of 8751 polymorphic SNPs on 316 DH lines were used for generating linkage maps using the method described in Cabral et al. (2014) with minor modifications. Linkage bins were determined with QTL IciMapping V4.1 software (Meng et al. 2015) using the 'BIN' module, and a single marker with the least missing data was selected per linkage bin for linkage analysis. Linkage analysis was conducted with MapDisto v1.7.7 (Lorieux 2012). A minimum logarithm of odds (LOD) value of 4.0 and maximum recombination fraction of 0.25 were used to identify linkage groups. Markers were ordered with a combination of the 'Automap,' 'Order sequence' and 'Compare all orders' functions. The 'Branch and Bound II' and 'Seriation II' ordering methods were used in combination with the sum of adjacent recombination fractions and count of crossover events as fitting criteria. Once the linkage map was developed, a single marker with the least missing data was selected per linkage bin for QTL analysis, resulting in a linkage map consisting of 1248 markers.

3.3.4. Statistical analysis

Descriptive statistical and frequency distribution analyses were performed using SPSS16.0 (SPSS Inc. 2007). A general linear model was used for analysis of variance using the following formula:

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

where G_i is the effect of the *i*th genotype; E_j is the effect of the *j*th environment; $B_{k(j)}$ is the blocking effect; GE_{ij} is the interaction effect between genotype and environment; and *eijkl* is the random error in individual plots (Imtiaz et al. 2008; Shao et al. 2018).

To account for the missing data, type III sum of squares were used. Relationship among the tested environments was determined through analysis of correlation coefficients among the environments using META-R (Alvarado et al. 2015). The broad-sense heritability across the five experiments was also estimated by META-R using the following equation (Imtiaz et al. 2008; Shao et al. 2018):

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

where *Vg* is the genotypic variance; *Vge* is the variance of genotype by environment; *Ve* is the error variance; *r* is the number of replications; and *e* is the number of environments.

3.3.5. QTL analysis

The mean SGI of three replications from each environment and best linear unbiased prediction (BLUP) value of all environment for the 316 DH lines were used for individual and pooled environment QTL mapping using QTL IciMapping V4.1 software (Meng et al. 2015). Interval mapping (IMA-DD) and inclusive composite interval mapping (ICIM-ADD) QTL analyses were conducted using the SGI and molecular phenotypic data of the 316 DH lines along with 1248 markers. The LOD significance thresholds for declaring QTL ($\alpha = 0.05$) were determined from 10,000 permutations. The thresholds were 3.13 for IM and 3.14 for ICIM. LOD scores greater than 2.5 were reported and QTLs were named using standard nomenclature for wheat.

3.4. Results

3.4.1. Characterization of germination phenotype and heritability

SGI was evaluated in the parental and DH lines under multiple environments over 3 years. Our data indicated that SGI value of the parental and DH lines varied across the environments tested (Table 3.1). The DH population had the highest SGI in WPG17 (14.44) followed by BRD17 (14.01). SGI was lowest in MOR18 (12.68) and MOR17 (12.76). The difference between the maximum and minimum SGI values also varied across the trial environments and years. MOR18 showed the lowest range (16.46) followed by BRD17, while GH16 showed the highest range (22.66) for SGI. The PHS susceptible parent ‘Roblin’ showed similar SGI value across all trials, whereas the SGI value for the resistance parent RL4137 varied between trial locations and years. The highest SGI (6.86) for RL4137 was observed in BRD17, while the lowest (1.90) was observed in GH16. Broad-sense heritability was calculated for each of the five trial environments and across all the five environments in which the highest heritability was found in GH16 followed by MOR17 and BRD17, while WPG17 trial showed the lowest heritability followed by the MOR18 trial (Table 3.1). Heritability across all the five trial environments was 0.91.

Table 3.1 Speed of germination index (SGI) of the 'Roblin'/RL4137 double haploid population and its parents in multiple environments

	Parental lines		DH lines			
Environment[a]	'Roblin' Mean SGI	RL4137 Mean SGI	Minimum SGI	Maximum SGI	Mean SGI	Heritability
GH16	19.93	1.90	2.34	25.00	13.86	0.99
WPG17	18.06	3.22	2.90	20.28	14.44	0.76
BRD17	17.28	6.86	3.87	20.33	14.01	0.96
MOR17	16.95	2.55	1.95	21.17	12.76	0.97
MOR18	17.61	4.08	2.98	19.44	12.68	0.89
Pooled	17.97	3.72	4.93	19.02	13.01	0.91

[a]GH16, greenhouse 2016; WPG17, Winnipeg 2017; BRD17, Brandon 2017; MOR17, Morden 2017; MOR18, Morden 2018

Frequency distributions of DH lines based on their respective SGI values revealed that their distribution is skewed toward PHS susceptibility for all trial environments (Fig. 3.1). The distribution of DH lines with intermediate SGI values followed an almost normal distribution for all environments except WPG17. Bidirectional transgressive segregation and bimodal distribution were also observed in some environments. For example, the transformed SGI data of WPG17 had a bimodal distribution (Data available at: https://link.springer.com/article/10.1007/s00122-020-03685-y).

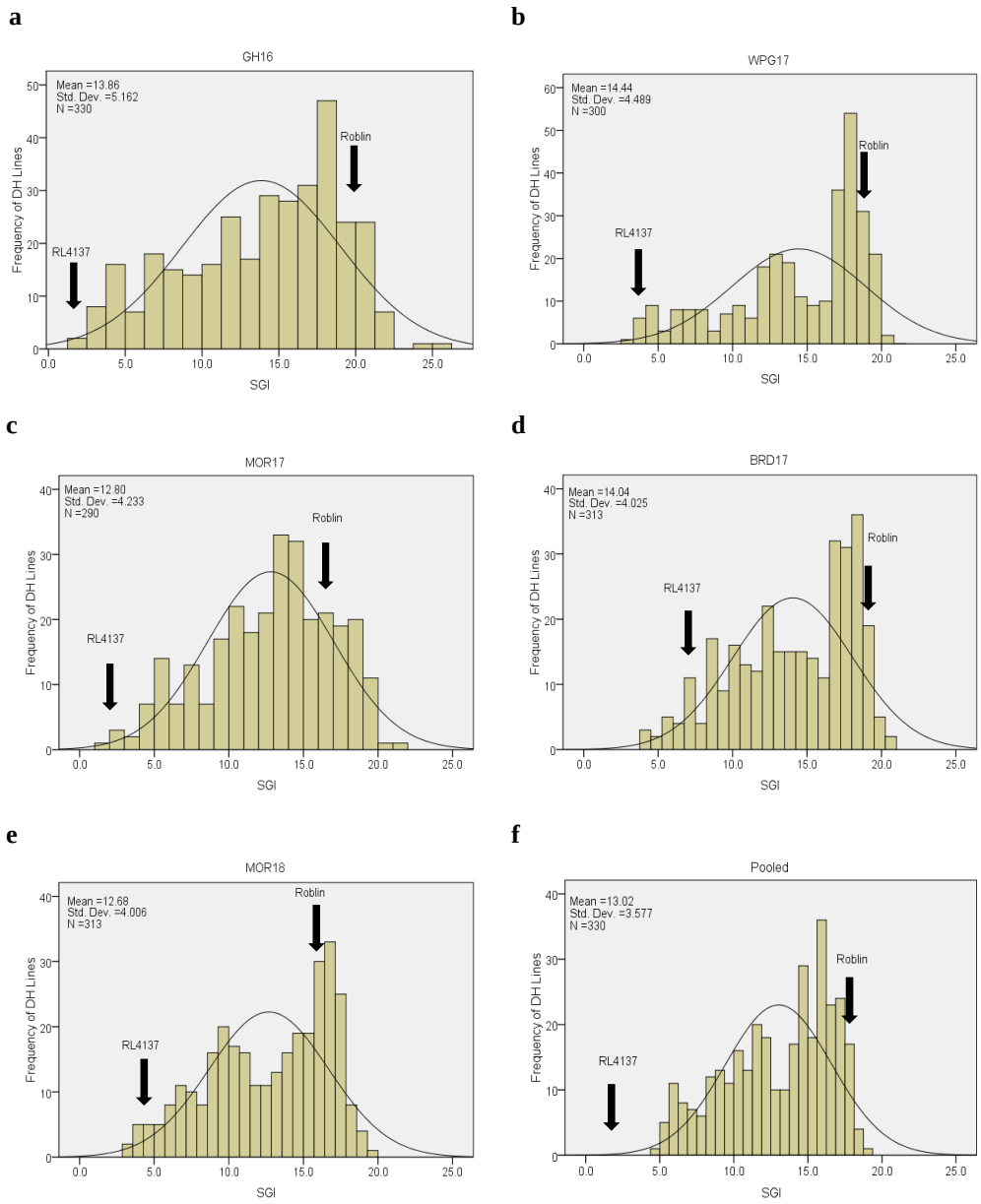


Fig. 3.1 Frequency distributions of speed of germination index in the 'Roblin'/RL4137 double haploid population in each trial environment and pooled over all environments. Black arrows indicate the mean SGI of the parental lines, RL4137 and Roblin, and the solid black curves indicate normal distribution. Overall mean, standard deviation and number of lines tested in each environment are indicated in each graph. The trial locations and years are indicated at the top of each graph. GH16, greenhouse trial in 2016; WPG17, Winnipeg trial in 2017; BRD17, Brandon trial in 2017; MOR17, Morden trial in 2017; MOR18, Morden trial in 2017; Pooled, data pooled from all trials.

3.4.2. Analysis of environmental correlation and variance

Correlation analysis examined the relationship among the trial environments with respect to SGI. Phenotypic correlation was positive and significant for all trial locations and years (Table 3.2).

Table 3.2 Correlation coefficients (r) between the trial environments for speed of germination index (SGI) in the 'Roblin'/RL4137 double haploid population

Environment[a]	GH16	WPG17	MOR17	BRD17
WPG17	0.62***			
MOR17	0.71***	0.65***		
BRD17	0.74***	0.65***	0.68***	
MOR18	0.70***	0.64***	0.72***	0.68***

[a]Environments are as described in Table 1
Asterisks (***) represent significance at $P<0.001$

The highest positive and significant correlation (0.74) was observed between GH16 and BRD17, while the lowest positive correlation (0.62) was found between GH16 and WPG17. Analysis of variance (ANOVA) identified statistically significant variance for both genotypes (DH lines) and environments, and the interaction between genotype and environment (Table 3.3).

Table 3.3 Analysis of variance of speed of germination index in the 'Roblin'/RL4137 double haploid population

Source	DF	Type III SS	Mean square	Expected mean square	*F* value	P>*F*
Environment[a] (*E*)	4	2221.65	555.40		86.78	<0.0001
Genotype (*G*)	330	66153.37	200.50	Ve+rVge+reVg	31.33	<0.0001
G×*E*	1210	24524.88	20.27	Ve+rVge	3.17	<0.0001
Error	2858	18299.36	6.40	Ve		
Total	4404	879266.60				

[a]Environments are as described in Table 1
DF, degrees of freedom; SS, sum of square; Ve, error variance; Vg, genotypic variance; Vge, variance due to interaction between genotype and environment; r, number of replications; e, number of environments; P, probability

3.4.3. Genetic mapping

A high-density linkage map of 2388.6 cM was generated using genotypic data of the DH population, which was derived using the wheat 90 K SNP array. A total of 8751 SNPs was mapped on 37 linkage groups (LGs) covering the 21 wheat chromosomes with some

chromosomes represented by more than one linkage group. For instance, chromosome 2A consisted of linkage groups 2A.1 and 2A.2 in which 2A.1 consists of part of chromosome arm 2AS, whereas 2A.2 consisted of chromosome arm 2AL. The SNPs were distributed unevenly across the chromosomes and the genomes (Data available at: https://link.springer.com/article/10.1007/s00122-020-03685-y). The highest number of SNPs (4720) was distributed across the B genome and a total of 2855 SNPs were mapped on the A genome (Data available at: https://link.springer.com/article/10.1007/s00122-020-03685-y). The lowest number of SNPs (1145) was mapped on the D genome. With respect to the distribution of SNPs on chromosomes, the highest number of SNPs (1340) was distributed on chromosome 2B and the lowest was found on chromosome 4D (a total of 8 SNPs; 5 SNPs on LG 4D.1 and 3 SNPs on LG 4D.2) (Data available at: https://link.springer.com/article/10.1007/s00122-020-03685-y).

3.4.4. Detection of QTL and allelic contribution

QTLs were detected by IM and ICIM QTL analyses using genotypic (90 K SNP) and phenotypic (SGI) data of 316 DH lines (Tables 3.4, 3.5; Fig. 3.2; Data available at: https://link.springer.com/article/10.1007/s00122-020-03685-y). IM and ICIM detected a total of five QTLs on linkage groups 1D, 4A.2, 6B.1, 6D and 7A. The QTLs individually explained 1.79–51.84% phenotypic variance with LOD value ranging from 3.68 to 65.45, and additive value ranging from −0.83 to 3.84. Four QTLs on linkage groups 1D (*QPhs.umb-1D*), 4A.2 (*QPhs.umb-4A*), 6B.1 (*QPhs.umb-6B*) and 6D (*QPhs.umb-6D*) were reproducibly detected in three or more trials with higher LOD values (Table 3.5; Fig. 3.2; Data available at: https://link.springer.com/article/10.1007/s00122-020-03685-y).

Table 3.4 QTLs for speed of germination index detected by interval mapping in the 'Roblin'/RL4137 double haploid population

Environment[a]	QTLs	LG	Position (cM)	Left Marker	Right Marker	LOD[b]	PVE (%)	Additive effect[c]
GH16	*QPhs.umb-4A*	4A.2	55.7	BS00068243_51	BS00037019_51	44.68	41.11	3.60
WPG17	*QPhs.umb-4A*	4A.2	55.1	BS00068243_51	BS00037019_51	29.91	38.48	2.84
MOR17	*QPhs.umb-4A*	4A.2	55.7	BS00068243_51	BS00037019_51	36.04	41.55	2.88
	QPhs.umb-7A	7A	73.4	Ku_c8523_166	Kukri_c38390_218	3.82	5.69	1.06
BRD17	*QPhs.umb-4A*	4A.2	55.4	BS00068243_51	BS00037019_51	42.81	29.12	2.85
MOR18	*QPhs.umb-4A*	4A.2	55.5	BS00068243_51	BS00037019_51	28.82	32.22	2.41
Pooled	*QPhs.umb-4A*	4A.2	55.3	BS00068243_51	BS00037019_51	48.68	46.15	2.58

[a]Environments are as described in Table 1

[b]LOD score value exceeding the threshold value, which is 3.13

[c]Estimated value of additive effect of the QTL, positive additive values indicate dormancy allele contributed by RL4137

LG, linkage groups; LOD, logarithm of odds; PVE, phenotypic variation explained by QTL

Of these four QTLs, *QPhs.umb-4A* was the most consistent across the trials with the highest LOD value (65.45) and explained the highest amount of phenotypic variation (51.84%). As a result, this QTL was considered a major QTL controlling PHS resistance (Fig. 3.2).

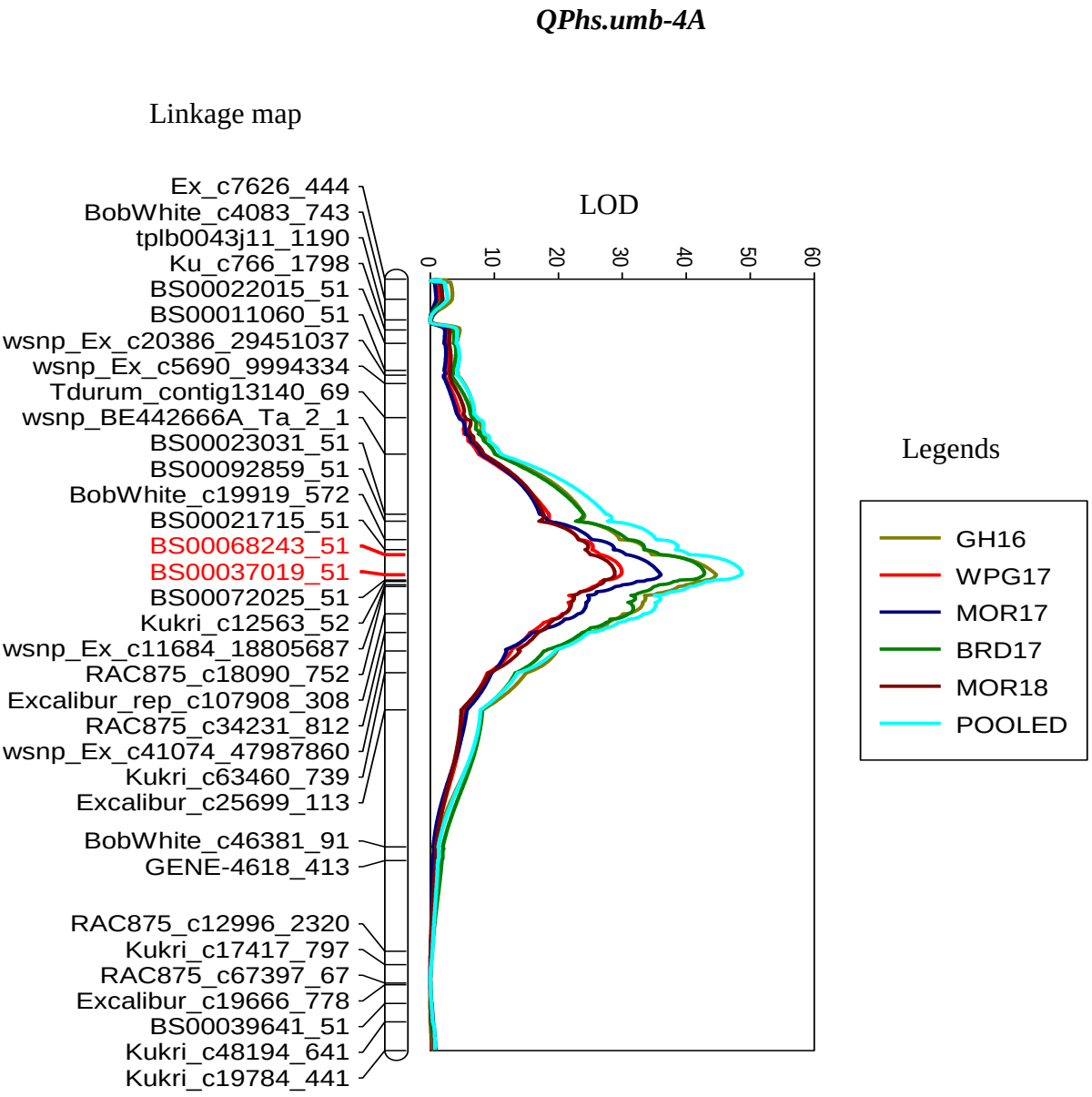


Fig. 3.2 Major QTL *QPhs.umb-4A* for PHS resistance detected on linkage group 4A.2 by both interval mapping (IM) and inclusive composite interval mapping (ICIM, not shown) in multi-environment experiments. The LOD value for the QTL identified at each environment is shown by the numbers at the top of the graph and the curves with different colours indicate the different trial environments as described in Figure 1. Markers in red color indicate the closest markers or marker interval for detected QTL.

This QTL was identified at approximately 55 cM on linkage group 4A.2 with flanking markers BS00068243_51 and BS00037019_51 in all trial environments. BLAST search located SNP BS00068243_51 at 603.067 Mbp on chromosome 4A of Chinese Spring RefSeq v1.0. The additive value for this QTL was approximately +3 for all trials, indicating that the PHS-resistant allele was contributed by the resistant parent RL4137. The effect of *QPhs.umb-4A* is presented separately due to its strong influence on PHS resistance in this population (Fig. 3.3). The mean SGI for the RL4137 allele at *QPhs.umb-4A* was 10.48 (*n* = 162 DH lines), while the mean SGI for the 'Roblin' allele was 15.61 (*n* = 151 DH lines). The bimodal nature of SGI in the DH population can be attributed to QTL *QPhs.umb-4A*.

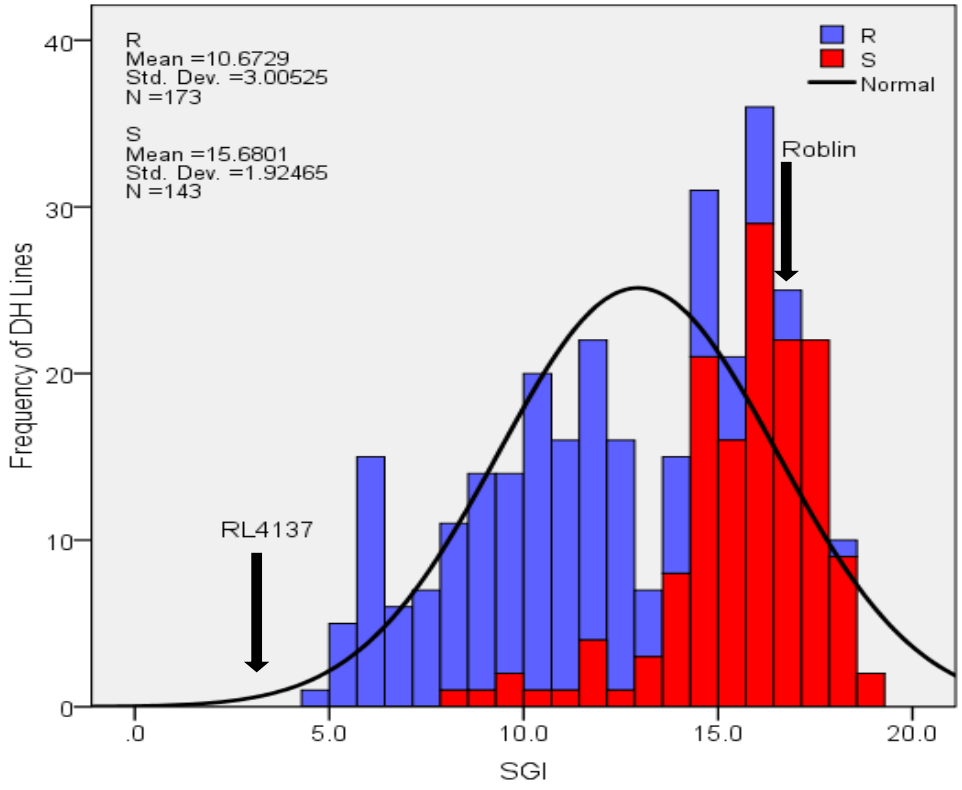


Fig. 3.3 Effect of *QPhs.umb-4A* on speed of germination index (SGI) in the 'Roblin'/RL4137 double haploid population. The black arrows indicate the mean SGI values of the parental lines. R, DH lines carrying the RL4137 allele; S, DH lines carrying the 'Roblin' allele at the BS00037019_51 locus; Normal, normal distribution curve.

QPhs.umb-6B, *QPhs.umb-1D* and *QPhs.umb-6D* are the three significant and stable minor QTLs associated with SGI and located at 142.5 cM, 117.7 cM and 1.9 cM on their respective chromosomes. These three stable QTLs had moderate effects on PHS resistance in the DH population compared to *QPhs.umb-4A*. Of these three QTLs, *QPhs.umb-6B* explained 3.1–4.3% of phenotypic variation and resistance came from RL4137. *QPhs.umb-1D* explained approximately 2% of phenotypic variation and the additive value was −0.6 indicating the 'Roblin' allele increased PHS resistance. Another QTL, *QPhs.umb-6D* was statistically significant using data pooled over environments (pooled) and had nearly significant LOD values in three individual environments. Though *QPhs.umb-6D* didn't have significant LOD values in each individual environment, it was considered as a stable but minor QTL due to its consistent effect in three environments. This QTL explained 1.62–2.45% of phenotypic variation and resistance allele was contributed by RL4137. The final QTL *QPhs.umb-7A* was detected by IM and ICIM and had minor effects on SGI in the 'Roblin'/RL4137 DH population (Tables 3.4, 3.5). The QTL *QPhs.umb-7A* was only detected in MOR17 and explained 4.0–5.7% of phenotypic variation. The resistance allele was contributed by the PHS resistance parent RL4137.

Table 3.5 QTLs for speed of germination index detected by inclusive composite interval mapping in the 'Roblin'/RL4137 double haploid population

Environment[a]	QTL	LG	Position (cM)	Left Marker	Right Marker	LOD[b]	PVE (%)	Additive effect[c]
GH16	*QPhs.umb-1D*	1D	142.5	BS00022188_51	Kukri_c10485_1346	3.92	2.33	-0.83
	QPhs.umb-4A	4A.2	55.7	BS00068243_51	BS00037019_51	55.52	49.56	3.84
	QPhs.umb-6B	6B.1	116.1	BobWhite_c1509_622	RFL_Contig3609_1678	5.16	3.09	0.96
WPG17	*QPhs.umb-4A*	4A.2	55.0	BS00068243_51	BS00037019_51	36.41	39.82	3.05
	QPhs.umb-6B	6B.1	115.8	RAC875_c63209_154	BobWhite_c1509_622	4.17	3.41	0.89
	QPhs.umb-6D	6D	1.9	IAAV7251	RFL_Contig4424_654	3.02	2.45	0.76
MOR17	*QPhs.umb-1D*	1D	142.5	BS00022188_51	Kukri_c10485_1346	3.76	2.04	-0.63
	QPhs.umb-4A	4A.2	55.7	BS00068243_51	BS00037019_51	53.57	45.80	2.96
	QPhs.umb-6B	6B.1	117.7	Tdurum_contig77500_369	Tdurum_contig47140_138	7.70	4.33	0.91
	QPhs.umb-6D	6D	3.7	IAAV7251	RFL_Contig4424_654	2.98	1.62	0.56
	QPhs.umb-7A	7A	73.4	Ku_c8523_166	Kukri_c38390_218	7.09	3.97	0.87
BRD17	*QPhs.umb-4A*	4A.2	55.7	BS00068243_51	BS00037019_51	43.47	48.26	2.81
MOR18	*QPhs.umb-1D*	1D	142.5	BS00022188_51	Kukri_c10485_1346	2.83	2.18	-0.63
	QPhs.umb-4A	4A.2	55.7	BS00068243_51	BS00037019_51	37.27	37.89	2.61
	QPhs.umb-6B	6B.1	117.4	BobWhite_c1059_518	Tdurum_contig77500_369	5.30	4.16	0.86
	QPhs.umb-6D	6D	1.9	IAAV7251	RFL_Contig4424_654	3.00	2.31	0.65
Pooled	*QPhs.umb-1D*	1D	142.5	BS00022188_51	Kukri_c10485_1346	4.12	1.99	-0.54
	QPhs.umb-4A	4A.2	55.5	BS00068243_51	BS00037019_51	65.45	51.84	2.75
	QPhs.umb-6B	6B.1	115.8	RAC875_c63209_154	BobWhite_c1509_622	7.32	3.63	0.73
	QPhs.umb-6D	6D	1.6	Jagger_c211_207	IAAV7251	3.68	1.79	0.51

[a]Environments are as described in Table 1

[b]LOD score value exceeding the threshold value, which is 3.14

[c]Estimated value of additive effect of the QTL, positive and negative additive values indicate dormancy allele contributed by RL4137 and 'Roblin', respectively

LG, linkage groups; LOD, logarithm of odds; PVE, phenotypic variation explained by QTL

3.4.5. Combined effects of major QTLs for PHS resistance

The DH lines were categorized into different genotypic classes based on eight SNP loci (BS00068243_51, BS00037019_51, RAC875_c63209_154, Tdurum_contig47140_138, BS00022188_51, Kukri_c10485_1346, Jagger_c211_207 and RFL_Contig4424_654) flanking the stable QTLs *QPhs.umb-4A*, *QPhs.umb-6B*, *QPhs.umb-1D* and *QPhs.umb-6D* (Tables 3.4, 3.5; Data available at: https://link.springer.com/article/10.1007/s00122-020-03685-y). Pooling of the SGI values of all genotypic class over all environments showed a gradual increase in SGI value from resistance to susceptible genotypic classes (Tables 3.6; Data available at: https://link.springer.com/article/10.1007/s00122-020-03685-y). The frequency distribution of DH lines was skewed slightly toward PHS susceptibility (Fig. 3.1; Data available at: https://link.springer.com/article/10.1007/s00122-020-03685-y) and this phenotypic skewing was mostly the result of QTL *QPhs.umb-4A*. The three significant and stable QTLs *QPhs.umb-6B*, *QPhs.umb-1D* and *QPhs.umb-6D* had little impact on SGI without the R allele (i.e., RL4137) at QTL *QPhs.umb-4A*. These three QTL played an important role in fine-tuning SGI in this population. The lowest GI value was observed in genotypic class R1R2R3R4 in all environments (except BRD17). The genotypic class with the highest SGI value varied between the different environments, but consistently did not carry the R1 allele.

Table 3.6 Mean speed of germination index (SGI) of 16 genotypic classes defined by the four stable QTL in the 'Roblin'/RL4137 DH population

Genotypic classes	No. of DH lines	Mean SGI					
		GH16[a]	WPG17	BRD17	MOR17	MOR18	Pooled
R1 R2 R3 R4	20	7.27	8.33	8.65	10.82	7.84	8.73
R1 R2 R3 S4	19	10.31	10.37	10.22	10.21	9.52	9.91
R1 R2 S3 R4	20	9.82	11.45	9.39	11.28	10.16	10.27
R1 S2 R3 R4	16	10.09	11.03	8.94	10.96	9.74	9.92
R1 R2 S3 S4	33	11.18	11.91	9.66	10.94	10.19	10.65
R1 S2 S3 R4	19	12.37	13.34	11.24	12.11	11.40	11.64
R1 S2 R3 S4	29	12.27	13.38	11.69	12.03	11.98	11.94
R1 S2 S3 S4	17	12.26	14.47	11.30	12.85	12.18	12.53
S1 R2 R3 R4	33	16.86	17.21	14.17	16.83	14.42	15.29
S1 R2 R3 S4	22	17.42	16.94	16.06	15.96	14.99	15.32
S1 R2 S3 R4	10	17.47	15.03	14.91	15.91	12.69	14.56
S1 S2 R3 R4	21	17.12	17.22	15.93	17.11	15.00	15.58
S1 R2 S3 S4	19	18.59	18.28	16.46	17.72	16.01	16.32
S1 S2 R3 S4	8	17.97	18.77	15.33	17.21	15.82	15.89
S1 S2 S3 R4	15	18.41	16.99	16.11	16.53	16.17	16.07
S1 S2 S3 S4	15	17.99	17.58	17.35	17.58	16.47	16.63

[a]GH16, greenhouse 2016; WPG17, Winnipeg 2017; BRD17, Brandon 2017; MOR17, Morden 2017; MOR18, Morden 2018

SGI, speed of germination index; R, resistance allele; S, susceptible allele; the numbers 1, 2, 3 and 4 after R and S indicate *QPhs.umb-4A*, *QPhs.umb-6B*, *QPhs.umb-1D* and *QPhs.umb-6D*, respectively

3.5. Discussion

The frequency distribution of SGI in the DH population was continuous with slight skewness toward PHS susceptibility, suggesting the involvement of multiple alleles in controlling PHS resistance. The contributions of such multiple minor allelic effects on PHS resistance and seed dormancy QTLs have also been reported previously (Mohan et al. 2009; Rasul et al. 2009; Kumar et al. 2015; Shao et al. 2018; Su and Li 2018). In the 'Roblin'/ RL4137 population, segregation at the major QTL *QPhs. umb-4A* was responsible for the bimodal tendency in the SGI frequency distributions, as illustrated in the stacked histogram (Fig. 3.3). The RL4137 allele at BS00037019_51locus increased PHS resistance by 19.1%, while the 'Roblin' allele at the same locus decreased PHS resistance by 20.4% from the population SGI mean value (12.96),

indicating that the 4A QTL have a great influence on PHS resistance of DH lines studied. In agreement with our findings, the influence of 4A QTL in controlling PHS resistance in wheat has been reported by a number of previous studies (Kato et al. 2001; Mares and Mrva 2001; Flintham et al. 2002; Kulwal et al. 2005; Mares et al. 2005; Mori et al. 2005; Torada et al. 2005; Chen et al. 2008; Imtiaz et al. 2008; Ogbonnaya et al. 2008; Rasul et al. 2009; Cabral et al. 2014). Our data also showed a highly significant and positive phenotypic correlation among the trial environments (Table 3.2) with high broad-sense heritability (Table 3.1). These results indicate similarity in the ranking of DHs in each environment and the greater contribution of genetic factors in controlling the investigated trait as described previously (Kulwal et al. 2005; Kumar et al. 2006; Imtiaz et al. 2008; Mohan et al. 2009; Su and Li 2018).

ICIM QTL analysis detected four significant and stable QTLs for PHS resistance on linkage groups 1D, 4A.2, 6B.1 and 6D; *QPhs.umb-4A* in all environments (Fig. 3.2), and *QPhs.umb-1D*, *QPhs.umb-6B* and *QPhs.umb-6D* in three or more environments (Data available at: https://link.springer.com/article/10.1007/s00122-020-03685-y). Among these four stable QTLs, IM QTL analysis also detected *QPhs.umb-4A* in all environments (Fig. 3.2). Several PHS resistance QTLs have previously been identified across the 21 chromosomes of common wheat (Kulwal et al. 2010; Tyagi and Gupta 2012; Vetch et al. 2019b; Wang et al. 2019). The *QPhs.umb-1D* is unique since it has not been reported by any previous study, while *QPhs.umb-4A*, *QPhs.umb-6B* and *QPhs.umb- 6D* have been detected before (Sorrells and Anderson 1996; Roy et al. 1999; Cabral et al. 2014; Kumar et al. 2015). The *QPhs.umb-4A* is likely to be same locus as *QGi.crc-4A* (Cabral et al. 2014) based on its mapping position and flanking SNPs. Kumar et al. (2015) reported two QTLs for PHS resistance on chromosome 6B and 6D using SSR markers and DH population from PHS resistance genotype SC8021- V2 and moderately susceptible genotype 'AC Karma'. However, comparing the previously reported QTLs (*QPhs. spa-6B* and *QPhs.spa-6D*) with those identified in this study (*QPhs.umb-6B* and *QPhs.umb-6D*) is difficult since the two studies involved different types of markers and coverage of genetic linkage maps.

QPhs.umb-4A was the most consistent across all trials, indicating that this QTL is a stable QTL with great influence in controlling PHS resistance in our population as reported in previous studies (Kulwal et al. 2010; Tyagi and Gupta 2012; Vetch et al. 2019b; Wang et al. 2019). Since this QTL was detected at almost the same location of chromosome 4A (linkage group 4A.2 at 55

cM), it likely explains the highest degree of phenotypic variation observed in all the trial environments. Likewise, Cabral et al. (2014) reported a consistent QTL for PHS resistance at a similar location of chromosome 4A using a population derived from a cross between RL4452 and 'AC Domain', for which RL4137 is an indirect parent (Townley-Smith and Czarnecki 2008). This study also identified *gibberellin 20 oxidase* (*GA20ox*) and abscisic acid-responsive element binding factors (*ABF2* and *ABF3*) as candidate genes associated with the QTL. Other studies have also reported QTLs on chromosome arm 4AL as regulators of PHS resistance through controlling seed dormancy (Kulwal et al. 2010; Tyagi and Gupta 2012; Vetch et al. 2019b; Wang et al. 2019). A Mitogen-Activated Protein Kinase Kinase 3 (*TaMKK3-A*) and a plasma membrane protein gene (*TaPM19*) have been proposed as the gene(s) responsible for the chromosome arm 4AL seed dormancy QTLs (Shorinola et al. 2016; Torada et al. 2016; Lin et al. 2018). BLAST searching the *TaMKK3-A* and *TaPM19* genes against RefSeq v1.0 from Chinese Spring revealed that these genes are positioned at 609.074 Mbp (LC091368.1) and 604.056 Mbp (KP844899.1), respectively, on chromosome 4A, which corresponds to our marker data for *QPhs.umb-4A*. It is therefore likely that the major QTL *QPhs.umb-4A* is controlled by the same underlying PHS regulator gene(s). This QTL is flanked by BS00068243_51 and BS00037019_51 markers with a distance of 3.8 cM and the resistance alleles are contributed by RL4137 parental line.

Given that each of the other three stable QTLs for PHS resistance, namely *QPhs.umb-1D*, *QPhs.umb-6B* and *QPhs.umb-6D* (Data available at: https://link.springer.com/article/10.1007/s00122-020-03685-y), individually explains a significant portion of phenotypic variation (1.62–4.33%) in multiple test environments, they are considered as significant and stable minor QTLs. In agreement with our results, QTLs for PHS resistance have been reported on chromosomes 6B and 6D (Sorrells and Anderson 1996; Roy et al. 1999; Kumar et al. 2015). The *QPhs.umb-1D* on chromosome 1D, however, has not been reported as a major PHS QTL before, indicating that it is a unique QTL controlling PHS resistance in the studied population. The *QPhs.umb-1D* was detected almost at the same chromosomal location in all environments, and it is flanked by BS00022188_51 and Kukri_c10485_1346 loci that are located at 462.203 Mbp and 475.724 Mbp, respectively, on chromosome 1D of Chinese Spring RefSeq v1.0. Though the other two QTLs *QPhs.umb-6B* and *QPhs.umb-6D* were flanked by different SNPs in different environments, their chromosomal

locations were similar in all environments. While the PHS resistance alleles for *QPhs.umb-6B* and *QPhs.umb-6D* were contributed by the resistant parental line RL4137, PHS resistance at the QTL *QPhs.umb-1D* was interestingly contributed by the susceptible parental line 'Roblin', indicating the role of both parents in allelic contribution toward PHS resistance in the studied population.

Since genetic interaction is critical for crop improvement, combination of QTLs plays an important role in breeding programs. The three stable QTLs (*QPhs.umb-1D, QPhs.umb-6B* and *QPhs.umb-6D*) showed moderate to minor individual effects on PHS resistance but progressively increased PHS resistance when combined with *QPhs.umb-4A* (Tables 3.6; Data available at: https://link.springer.com/article/10.1007/s00122-020-03685-y). While *QPhs.umb-4A* alone reduced SGI by 32% (pooled, R1_ _ _ versus S1__ _), a combination of all the four QTLs decreased SGI by up to 53% (pooled, genotypic class R1R2R3R4 versus S1S2S3S4), indicating a considerable enhancement of PHS resistance. Such an effect of a combination of QTLs on resistance against PHS has also been reported previously (Imtiaz et al. 2008; Kumar et al. 2015; Shao et al. 2018). Our analysis revealed that the RL4137 parental line contributed resistant alleles for *QPhs.umb-4A*, *QPhs.umb- 6B* and *QPhs.umb-6D*, and the resistant allele for *QPhs.umb-1D* came from 'Roblin' (Table 3.5). These results highlight the presence of complementary alleles in the two parental lines; however, RL4137 contributes more genetic factors for PHS resistance than 'Roblin'. Previous studies have also reported similar type of parental genetic contributions for PHS resistance (Knox et al. 1998; Kulwal et al. 2004; Imtiaz et al. 2008; Fofana et al. 2009; Mohan et al. 2009; Kumar et al. 2015). This study, furthermore, emphasizes the importance of pyramiding QTLs in breeding since there was considerable variation in SGI among the carriers of the RL4137 allele at *QPhs.umb-4A.*

In summary, a high-density genetic map was used to detect QTL for PHS resistance in common wheat. Four significant and stable QTLs were identified on chromosomes 1D, 4A, 6B and 6D evaluating a doubled haploid population of 'Roblin'/RL4137 in greenhouse and field experiments. Detection of QTL on chromosome 4A was consistent with previous reports and explained the largest amount of phenotypic variation within the population. This study is the first to identify QTL for PHS resistance on chromosome 1D (*QPhs.umb-1D*). To develop wheat cultivars with enhanced PHS resistance, all four QTLs are important, but the RL4137 allele at *QPhs.umb-4A* is critical for wheat breeders. RL4137 also contributed other PHS resistance

alleles at *QPhs.umb-6B* and *QPhs.umb-6D*, and ‘Roblin’ contributed PHS resistance at *QPhs.umb-1D*. The highest PHS resistance genotypic class exhibited allelic combination from both RL4137 and ‘Roblin’ parents. Therefore, pyramiding of PHS resistance QTLs from different sources can significantly improve resistance of spring wheat to PHS.

CHAPTER 4

TRANSCRIPTOMIC ANALYSIS OF SEED DORMANCY IN A HIGHLY DORMANT GENOTYPE RL4137 REVEALS CANDIDATE GENES FOR PRE-HARVEST SPROUTING RESISTANCE IN WHEAT

M M Uzzal A. Liton[1], Pham Anh Tuan[1], Curt A. McCartney[1], Colin W. Hiebert[2], Mark C. Jordan[2], Belay T. Ayele[1*]

[1]Department of Plant Science, University of Manitoba, Winnipeg, MB, R3T 2N2, Canada

[2]Morden Research and Development Center, Agriculture and Agri-Food Canada, Morden, MB, R6M 1Y5, Canada

Author contributions: BTA and MCJ conceived, designed and supervised the research; UAL and PAT performed the experiments and analyzed the data; CWH developed the double haploid lines; CAM was involved in the selection of the double haploid lines used for the RNA-seq analysis; UAL wrote the first draft of the manuscript; BTA and PAT edited and revised the manuscript; and all authors read and approved the final manuscript.

Transcriptomic analysis of seed dormancy in a highly dormant genotype RL4137 reveals candidate genes for pre-harvest sprouting resistance in wheat

4.1. Abstract

To better understand the molecular bases of dormancy in wheat, the present study conducted comparative transcriptomic analysis between dormant and non-dormant double haploid (DH) lines selected from a DH population used for genetic study. Our analysis identified a large number of genes as differentially expressed between dormant and non-dormant seeds. Gene Ontology (GO) enrichment analysis reported that cell, cell part, intracellular, integral component of membrane under the cellular component domain are the major GO terms enriched with genes differentially expressed between dormant and non-dormant seeds. Pathway enrichment analysis revealed that carotenoid biosynthesis, which includes ABA metabolism, diterpenoid biosynthesis including gibberellin catabolism and flavonoid biosynthesis as well as ABA signaling pathways are enriched with genes upregulated in dormant seed while those involved in auxin signaling and GA biosynthesis pathways are enriched with genes downregulated in same sample. Mapping of the differentially expressed genes on previously detected QTLs (1D, 4A, 6B and 6D) identified potential candidate genes for seed dormancy. Of these DEGs, *TaCYP71B11* and *TaDOG1L-2* on 1D, *TaPM19-A1* and *A2* on 4A, an uncharacterized gene (Gene ID: TraesCS6B02G325000) on 6B and *TaLEA4* on 6D showed upregulation in dormant seed while two uncharacterized genes (Gene IDs: TraesCS1D02G410700 and TraesCS4A02G317200) on 1D and 4A, *TaRH8* on 6B and *TaHIS4* on 6D QTLs showed downregulation in same sample. In addition, AP2-EREBP transcription factor family member, *TaDREB2C*, and C/EBP family member, *histone H2A* showed up- and downregulation in dormant seed, respectively. Furthermore, 2 to 25 SNPs that appear to be associated with seed dormancy candidate genes were detected within the four QTLs. Taken together, the results of this study have the potential to foster our understanding of the molecular and genetic mechanisms of seed dormancy and therefore facilitate PHS resistance breeding program in wheat.

Keywords: pre-harvest sprouting, plant hormone, seed dormancy, single nucleotide polymorphism, transcriptomic analysis, wheat.

4.2. Introduction

Bread wheat (*Triticum aestivum* L.) is one of the important cultivated cereals worldwide and the grains are used for different purposes including food, feed and raw materials for industrial applications. However, the productivity and quality of wheat is significantly affected by various biotic and abiotic factors including PHS, which refers to the germination of seeds on spikes prior to harvest due to wet and humid weather conditions (Groos et al. 2002). PHS causes significant yield and quality reduction in wheat (Mares et al. 2005; Kulwal et al. 2012). Annual financial loss due to PHS, has been estimated to be approximately $1.2 billion worldwide (DePauw et al. 2012; Shao et al. 2018). Therefore, it is imperative to enhance PHS resistance to reduce wheat yield and quality losses.

PHS resistance is a quantitative trait which is controlled by several genetic and environmental factors. Seed dormancy, a trait that makes viable seeds unable to germinate under favourable conditions, is highly associated with PHS resistance (Gao and Ayele 2014; Tuan et al. 2018; Wu et al. 2020). Domesticated crop cultivars possess low level of seed dormancy than their wild ancestors since such a trait is required for uniformity in germination and seedling establishment (Tuan et al. 2018; Wu et al. 2020); however, such a low level of dormancy in seeds leads to PHS. In contrast, higher degree of dormancy leads to non-uniform germination and seedling establishment (Gao and Ayele 2014; Tuan et al. 2018). Thus, understanding the molecular mechanisms regulating seed dormancy is needed to develop cultivars with intermediate level of dormancy and PHS resistance.

Seed germination is a sequential time event of resumption of embryo growth which starts with uptake of water and ends with radicle protrusion involving three phases of water uptake (Bewleyl 1997; Gianinetti et al. 2018; Song et al. 2019). The first phase is characterized by rapid uptake of water through imbibition and activation of metabolic processes (Gianinetti et al. 2018). Modification of transcript abundance through degradation of stored mRNA and resumption of cellular and molecular processes start at this stage (Née et al. 2017; Gianinetti et al. 2018). In the second phase, water uptake continues very slowly or stops as the seed is hydrated sufficiently and reactivation of metabolic processes, repair of membrane and cellular components, and synthesis of proteins and metabolites occur (Bewleyl 1997; An and Lin 2011; Rajjou et al. 2012). At the third phase, rapid water uptake resumes with radicle protrusion and seedling growth (Bewley 1997). In dormant seed, the third phase is postponed and the seed remains in the

metabolically active second phase (Gianinetti et al. 2018). Previous studies have reported that activation of transcriptional regulatory programs starts within 3 hours of imbibition and extends up 12 hours (Howell et al. 2009; An and Lin 2011; Rajjou et al. 2012).

Global gene expression analyses have provided important information about the molecular mechanisms that control seed dormancy. Previous transcriptomic studies of seed dormancy in different plant species such as Arabidopsis, wheat, rice and barley have reported the involvement of several biological processes in of the regulation of this trait, which include nutrient reservoir, carbon metabolism, biosynthesis of protein, chromatin remodelling, cell wall modification, catalytic activity, transportation activity and stress tolerance, secondary metabolite biosynthetic and plant hormone signal transduction pathways (An and Lin 2011; Gao et al. 2012; Nelson et al. 2017; Wu et al. 2020; Wang et al. 2021).

Plant hormones, mainly ABA and GA, are important regulators of seed dormancy and germination (Tuan et al. 2018; Wu et al. 2020). Embryo produced ABA is fundamental to induce and maintain dormancy during seed development and imbibition (Finkelstein et al. 2002). The seed ABA level is controlled by its production and catabolism that are regulated mainly by NCED and ABA8′-OH, encoded by *CYP707A* genes, respectively (Nambara et al. 2010; Tuan et al. 2018). Previous studies, for example, have shown higher and lower expression levels of the *NCED*s and *CYP707A*s, respectively, in dormant seeds of different plant species including Arabidopsis and wheat (Izydorczyk et al. 2018; Tuan et al. 2018). In addition to the amount of ABA, the response of seeds to ABA is important in the control of dormancy. ABA signaling is regulated mainly by PYR/PYL/RCAR, PP2C and SnRK2, which regulate ABA signaling negatively and positively, respectively. In addition, the downstream transcription factors that directly regulate the ABA responsive genes, which are designated as ABA insensitive proteins (ABIs include ABI3, ABI4 and ABI5) are critical components of ABA signaling (Nambara et al. 2010; Tuan et al. 2018). Previous studies have shown that genes encoding *TaPYL*s *TaSnRK2*, *TaABI3*, *TaABI4* and *TaABI5* are upregulated in dormant seeds (Izydorczyk et al. 2018; Tuan et al. 2021), reflecting their important role in the regulation seed ABA responsiveness and dormancy.

GA promotes seed germination and increase embryo growth by degrading DELLA protein, a plant growth inhibitor (Sun and Gubler 2004). The amount of GA in plant tissues is regulated mainly by the GA biosynthetic enzymes, GA20ox and GA3ox, and a GA catabolic enzyme,

GA2ox (Tuan et al. 2018). Upregulation of genes encoding GA20ox and GA3ox and downregulation of those encoding GA2ox have been reported in after-ripened/non-dormant seeds (Gubler et al. 2008; Izydorczyk et al. 2018; Tuan et al. 2021). In addition to amount of GA in seeds, GA signaling is important to control germination. GA signaling is regulated by GID1, which degrades DELLA protein and activates downstream transcription factor GAMYB, leading to increase GA responsiveness and seed germination (Tuan et al. 2018). Previous studies have also shown the presence of higher expression level of GA signaling genes such as *TaGID*s and *TaGAMYB* in non-dormant/germinating seeds during imbibition but lower expression level of *TaRHT1* (encoding wheat DELLA protein) in the same seed sample (Izydorczyk et al. 2018; Tuan et al. 202).

Other plant hormones such as auxin, ET, JA and BR also play roles in the regulation of seed dormancy and germination (Shu et al. 2016). Of these, auxin promotes seed dormancy by altering ABA response through regulating the downstream ABA signaling component including *ABI3*, *ABI4* and *ABI5* (Ali et al. 2019). For example, the ARF controls the expression of *ABI3* and therefore seed dormancy (Liu et al. 2013c). In contrast, ET promotes seed germination by antagonizing the effects of ABA (Ghassemian et al. 2000; Arc et al. 2013; Corbineau et al. 2014). For instance, seeds of ET insensitive mutants of Arabidopsis such as *etr1*, *ein2* and *ein6* showed hypersensitivity to ABA and enhanced dormancy and this was associated with repression of genes encoding endosperm weakening enzyme such as β-glucanase and chitinase (Leubner-Metzger et al. 1998; Corbineau et al. 2014). In addition, upregulation of the *ACO1* and *ERS1* genes have been reported in after-ripened wheat seeds (Chitnis et al. 2014). The other two hormones, JA and BR reduce seed ABA sensitivity and therefore promote germination (Skubacz and Daszkowska-Golec 2017; Ali et al. 2019). It has been shown previously that the JA biosynthetic genes including *lipoxygenase 5* (*LOX5*), *12-oxophytodienoate reductase 1* (*OPR1*) and *allene oxide synthase* (*AOS*) are upregulated in imbibed after-ripened wheat seeds (Liu et al. 2013a) while JA-insensitive mutants such as *jin4* and *jar1* exhibit higher ABA sensitivity in Arabidopsis (Berger et al. 1996). With respect to BR, upregulation of the BR biosynthetic and signaling genes in imbibed after-ripened/non-dormant wheat seeds have been shown to be associated with increased expression of levels of BR responsive genes including those involved in cell elongation, a process of required for the completion of seed germination (Zhang et al. 2009).

Several QTLs that control PHS have been identified on all the 21 chromosomes of bread wheat (Wang et al. 2021; Ali et al. 2019). Of these QTLs, those located on chromosome group 3 and 4A explain large phenotypic variation consistently (Mares and Mrva 2001; Chen et al. 2008; Somyong et al. 2014). However, only a few candidate and causal genes underlying the QTLs associated with the PHS have been reported to date, which include *TaPM19-A1* and *A2*, *TaMKK3*, *TaMFT/TaPHS1*, *TaVp-1* and *Tamyb10* (Ali et al. 2019; Vetch et al. 2019; He et al. 2021). Of these genes, *TaMFT/TaPHS1* was reported as a causal gene for *QPhs.pseru-3A* QTL, which has been identified as a major PHS resistance QTL in multiple studies (Vetch et al. 2019). Three SNPs in the promoter and the third intron of *TaMFT*, which are located at -222 bp and 646/666 bp positions, respectively, are critical for PHS resistance (Jiang et al. 2018; Nakamura 2018). The *TaMKK3* was also identified as a causal gene for Phs1 locus of 4AL QTL, and it encodes a mitogen-activated kinase kinase 3 protein (Torada et al. 2016; Lin et al. 2018), which has been reported to positively regulate seed ABA responsiveness (Danquah et al. 2015). A SNP located at 660 bp downstream of initiation codon of *TaMKK3*, which causes a non-synonymous substitution of amino acids, has been identified as important regulator of PHS (Torada et al. 2016; Lin et al. 2018). Other genes that have been identified as candidate genes regulating PHS resistance include *TaDOG1* and *TaSdr* (Nishimura et al. 2018; He et al. 2021). However, many other causal genes underlying PHS resistance QTLs and the associated SNPs remain to be identified.

Our previous genetic study, which involved a DH population derived from a cross between 'Roblin' and RL4137, has also detected a major 4A QTL with other three consistent/stable QTLs on chromosome 1D, 6B and 6D (Liton et al. 2021). Of these, the *QPhs.umb-4A* explains up to 50% of phenotypic variation while *QPhs.umb-1D* has never been reported previously. Although one causal (*TaMKK3*) and two candidate genes (*TaPM19* and *TaGA20ox*) have been identified in the 4A QTL region (Cabral et al. 2014; Barrero et al. 2015; Torada et al. 2016), it is likely that more genes controlling PHS are present in the four QTLs. Thus, detailed study of those loci using different approaches such as transcriptomic analysis between PHS resistant and susceptible lines is crucial to identify more candidate genes associated with PHS resistance. The main objective of this transcriptomic study was therefore to investigate the molecular mechanisms regulating seed dormancy in a highly dormant genotype RL4137 and to identify candidate genes

controlling PHS resistance including those associated with the QTLs detected by our genetic study of the double haploid population.

4.3. Materials and methods

4.3.1. Plant materials and growth conditions

Five highly dormant (DH1-DH5) and five highly non-dormant (DH6-DH10) lines of the DH population derived from 'Roblin'/RL4137 (Liton et al. 2021) were used for this study. In brief, 'Roblin' is a CWRS wheat class variety with PHS susceptibility while RL4137 is a PHS resistant breeding line. The selected DH lines along with their parents were grown in a growth chamber under a 22°C/18°C (day/night) with a 16/8 h photoperiod. Seeds were harvested at harvest maturity and stored at -80°C until further use.

4.3.2. Germination assay and allele frequency

Seeds of each DH line were surface sterilized and subjected to germination assay using Petri-plates exactly as described previously (Gao et al. 2012). Imbibing seeds were incubated at room temperature under darkness and germination was counted every day for 7 days; a seed was considered germinated when the coleorhiza emerged through the seed coat.

For transcriptomic analysis, seeds of each DH line were collected at 12 and 24 h after imbibition (HAI) and their embryonic tissues were separated. Embryonic tissues were also separated from the dry seeds of each line. The tissues were stored at -80°C until they are needed for RNA extraction. Allelic frequency of the selected DH lines at QTLs detected in our previous genetic study was analyzed using marker data derived from genotypic analysis of the respective DH population (Liton et al. 2021).

4.3.3. RNA extraction

Total RNA was isolated from the embryonic tissues of the highly dormant and highly non-dormant seed samples as described previously (Izydorczyk et al. 2018). RNA integrity and quantitation were examined using the RNA Nano 6000 Assay Kit and the Agilent 2100 Bioanalyzer system (Agilent Technologies, Santa Clara, CA, USA). Only RNA samples with RNA integrity number (RIN) higher than 9.0 were used for sequencing library preparation.

4.3.4. Library preparation and Illumina sequencing

Preparation of the sequencing libraries was performed using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (New England BioLabs, Ipswich, MA, USA) following manufacturer's protocols. In brief, poly-T oligo-attached magnetic beads were used to isolate mRNA from total RNA and the mRNA was fragmented to get an RNA insert size of ~200 nt in NEBNext First Strand Synthesis Reaction Buffer using divalent cations under elevated temperature. Synthesis of first strand complementary DNA (cDNA) was performed using M-MuLV Reverse Transcriptase (RNase H) and random hexamer primer. Second strand cDNA was synthesized subsequently using DNA Polymerase I and RNase H followed by conversion of overhangs to blunt ends, 3' adenylation and adapter ligation. AMPure XP system (Beckman Coulter, Brea, CA, USA) was used to purify fragments with 150~200 bp in length. The fragments were then amplified by PCR followed by quality assessment using Agilent Bioanalyzer 2100 system (Agilent Technologies). Good quality fragments were selected for sequencing on an Illumina platform (Hiseq 2500) to generate 150 bp pair-end reads.

4.3.5. Quality control and expression quantification

In order to get clean data (clean reads), reads containing adapter sequences and poly-N, and low-quality reads were removed from the raw data. At the same time, Phred quality score (Q) of 20 and 30 (Q20 and Q30) and GC content of the clean data were calculated. High quality reads were then mapped to bread wheat reference genome CS RefSeq v1.0 (https://wheat-urgi.versailles.inra.fr/) using HISAT2 v2.1.0 (Kim et al. 2015). Mapped reads were counted using HTSeq v0.6.1 (Anders and Huber 2012) to estimate the level of expression of an individual gene. Read count was also used to determine gene length and depth of sequencing. In addition, FPKM of each gene, which was calculated based on the length of the gene and read counts mapped to this gene. FPKM was used to estimate gene expression level (Trapnell et al. 2010). Pearson correlation coefficient (R^2) was calculated among the four replicates of each sample using the gene expression data.

4.3.6. Differential expression analysis

Differential gene expression analysis between samples was performed with the DESeq R package v1.18.0 using a model that is based on the negative binomial distribution (Anders and Huber 2010). The resulting P-values were adjusted using the Benjamini and Hochberg's

approach and an adjusted P-value ≤ 0.05 was considered as a cut-off P-value to determine differential expression of genes.

4.3.7. GO and KEGG enrichment analysis of differentially expressed genes

GO enrichment analysis of the DEGs was performed using the GOseq R package (Young et al. 2010), in which gene length bias was corrected. GO terms with corrected P-value ≤ 0.05 were considered significantly enriched with the DEGs. KOBAS software v3.0 (Mao et al. 2005) was used to determine the statistical enrichment of DEGs in KEGG pathways. KEGG pathway enrichment was determined using rich factor, q-value and number of genes in the enriched pathways. Rich factor is the ratio of number of DEGs and total number of annotated genes in a certain pathway. The q-value is the corrected P-value after multiple hypothesis tests and it ranges between 0 and 1. The lower the value is the more significant the pathway.

4.3.8. Selection of DEGs and functional annotation

The DEGs (≥2-fold, P≤0.05) involved in different biological processes and those identified to be localized in previously identified QTL regions (Liton et al. 2021) were annotated. To this effect, the IWGSC ID of the selected DEGs were blast searched against the *Triticum aestivum* PANTHER v16.0 database of protein families, subfamilies, functions and pathways (Mi et al. 2021; http://pantherdb.org/). Other databases including TAIR (https://www.arabidopsis.org/), Ensemble Plants (https://plants.ensembl.org/) and The Rice Annotation Project (https://rapdb.dna.affrc.go.jp/) were also used to annotate the DEGs.

4.3.9. Identification of transcription factors (TF) and variants calling

The DEGs encoding transcription factors (TFs) were identified using iTAK software v1.2 and its default parameters. Annotation of the TFs was done using Plant Transcription Factor Database (PlantTFDB) v5.0 (Tian et al. 2020). Gene interaction network was generated using STRING v11.5 (Szklarczyk et al. 2019). Picard-tools v1.96 and Samtools v0.1.18 were used to sort, mark duplicated reads and reorder the bam alignment results of high-quality reads in order to identify SNPs and insertions and deletions (InDels) within the genes that showed differential expression between the 24 HAI dormant and non-dormant samples. The GATK2 software v3.2 (McKenna et al. 2010) was used for calling SNPs and InDels in the DEGs and gene variants within the previously identified QTL regions.

4.3.10. Validation of RNA-Seq data using RT-qPCR

Validation of the RNA-Seq data was conducted through expression analysis of selected DEGs using gene specific primers and qPCR. The primers of the genes analyzed for validation of the RNA-seq data [*TaNCED1*, *TaABI3*, *TaGA20ox1*, *TaGA2ox6*, *TaAux/IAA19*, *α-expansin 3* (*TaEXPA3*) and the reference gene *Taβactin*] are as described previously (Qiao et al. 2018; Izydorczyk et al. 2018; Sun et al. 2020; Tuan et al. 2021). Gene specific primers for *TaARF3* and *cyclin D1* (*TaCYCD1*) were designed from the conserved coding regions of the three homeologues of each target gene using Primer 3 software (Appendix 1). Synthesis of the cDNAs from the same total RNA samples used for RNA-Seq analysis and the subsequent qPCR assays were conducted as described previously (Izydorczyk et al. 2018). Gene relative expression levels were determined using the Livak and Schmittgen (2001) method, and the *Taβactin* gene was used as a reference gene for normalization.

4.4. Results

4.4.1. Phenotypic characterization

To verify variation in the level of dormancy between dormant and non-dormant DH lines, mature seeds were subjected to germination test over 7 days. Variation in germination percentage was evident between the dormant and non-dormant DH lines. Non-dormant seeds attained 98% germination after 24 h imbibition whereas no seeds of the dormant lines germinated within the same period, and 6% germinated after 7 days (Fig. 4.1).

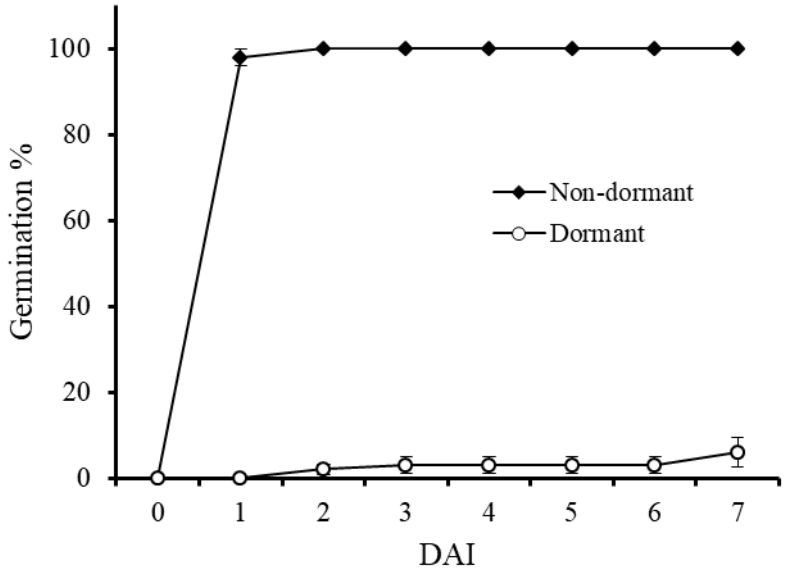


Fig. 4.1 Germination percentage of the seeds of dormant and non-dormant DH lines over 7 days of imbibition. DAI, days after imbibition.

4.4.2. Analysis of allelic frequency in DH lines

The allelic frequencies at previously identified QTLs of all tested DH lines were analyzed using the respective genotypic data from our previous genetic study (Liton et al. 2021). The alleles of all loci in dormant lines were contributed by resistance parent RL4137 (Table 4.1), except at *QPhs.umb-6D* of DH line #2 and Kukri_c10485_1346 locus (one of 2 SNPs from *QPhs.umb-1D*) of DH line #5. In the non-dormant lines, alleles at *QPhs.umb-4A* were contributed by 'Roblin' except BS00068243_51 locus (*QPhs.umb-4A*) of DH line #7. While allelic contribution at other

loci (*QPhs.umb-1D*, *QPhs.umb-6B* and *QPhs.umb-6D*) were contributed by both RL4137 and 'Roblin' and were found to vary with DH lines (Table 4.1).

Table 4.1 Allele frequency of selected dormant and non-dormant DH lines of 'Roblin'×RL4137

	Parent		Dormant (D)					Non-dormant (ND)				
Locus	'Roblin'	RL4137	DH1	DH2	DH3	DH4	DH5	DH6	DH7	DH8	DH9	DH10
					QPhs.umb-1D							
BS00022188_51	A	B	B	B	B	B	B	B	A	B	B	B
Kukri_c10485_1346	A	B	B	B	B	B	A	B	A	B	B	B
					QPhs.umb-4A							
BS00068243_51	A	B	B	B	B	B	B	A	B	A	A	A
BS00037019_51	A	B	B	B	B	B	B	A	A	A	A	A
					QPhs.umb-6B							
RAC875_c63209_154	A	B	B	B	B	B	B	A	B	A	B	B
BobWhite_c1509_622	A	B	B	B	B	B	B	A	B	A	B	B
BS00091262_51	A	B	B	B	B	B	B	A	B	A	B	B
RFL_Contig3609_1678	A	B	B	B	B	B	B	A	B	A	B	B
RFL_Contig2391_831	A	B	B	B	B	B	B	A	B	A	B	B
BS00048293_51	A	B	B	B	B	B	B	A	B	A	B	B
BobWhite_c1059_518	A	B	B	B	B	B	B	A	B	A	B	B
Tdurum_contig77500_369	A	B	B	B	B	B	B	A	B	A	B	B
Tdurum_contig47140_138	A	B	B	B	B	B	B	A	B	A	B	B
					QPhs.umb-6D							
Jagger_c211_207	A	B	B	A	B	B	B	B	A	B	B	A
IAAV7251	A	B	B	A	B	B	B	B	A	B	B	A
RFL_Contig4424_654	A	B	B	A	B	B	B	B	A	B	B	A

4.4.3. Quality of transcriptome assembly and gene expression quantification

A total of 173.20 Gb raw data covering 1.16 billion raw reads were generated from 24 libraries ranging from 42.22 million to 53.30 million raw reads per library. Of the 1.16 billion raw reads, 1.12 billion clean reads, which is equivalent to the 167.60 Gb of clean data, were acquired after removing the adapter and poly-N, and low quality reads (Appendix 2). Over 92% reads showed a quality score of Q30 (99.9% accuracy) indicating the higher accuracy of the sequencing data. The GC content of reads distributed within 53.49% to 55.42%. High quality reads were mapped

to bread wheat reference genome CS RefSeqv.1 (https://wheat-urgi.versailles.inra.fr/) and approximately 90% of reads were mapped to unique regions (Appendix 3). The percentage of reads mapped to multiple regions was below 6.5%, which was much lower than cut-off value of 10%. Our analysis showed the mapping of 120,744 genes to the reference genome.

Approximately 62.63% to 70.63% genes showed no to very low level of expression with an FPKM interval of 0~1, and 9.13% to 11.82% genes showed low level of expression with 1~3 FPKM (Appendix 4). The FPKM interval for 13.02% to 18.51% genes was 3~15, indicating that these genes exhibited an intermediate level of expression. High and very high level of expression was observed in approximately 5.5% and 1.5% genes in which the FPKM interval was 15~60 and >60, respectively, across the libraries. The R^2 value among the four replicates of each sample was ≥ 0.92 indicating consistency and reproducibility of the data (Appendix 5).

4.4.4. Cluster analysis of expressed genes

A total of 48,398 genes were expressed in our samples. Based on their FPKM, a hierarchical cluster analysis of the expressed genes between dry/imbibed dormant and non-dormant seeds was performed in order to determine their expression patterns (Fig. 4.2). Our data showed two distinct clusters of expressed genes between dry and imbibed seed samples, and the expression level of some genes increased while that of some other genes decreased with increased time of imbibition.

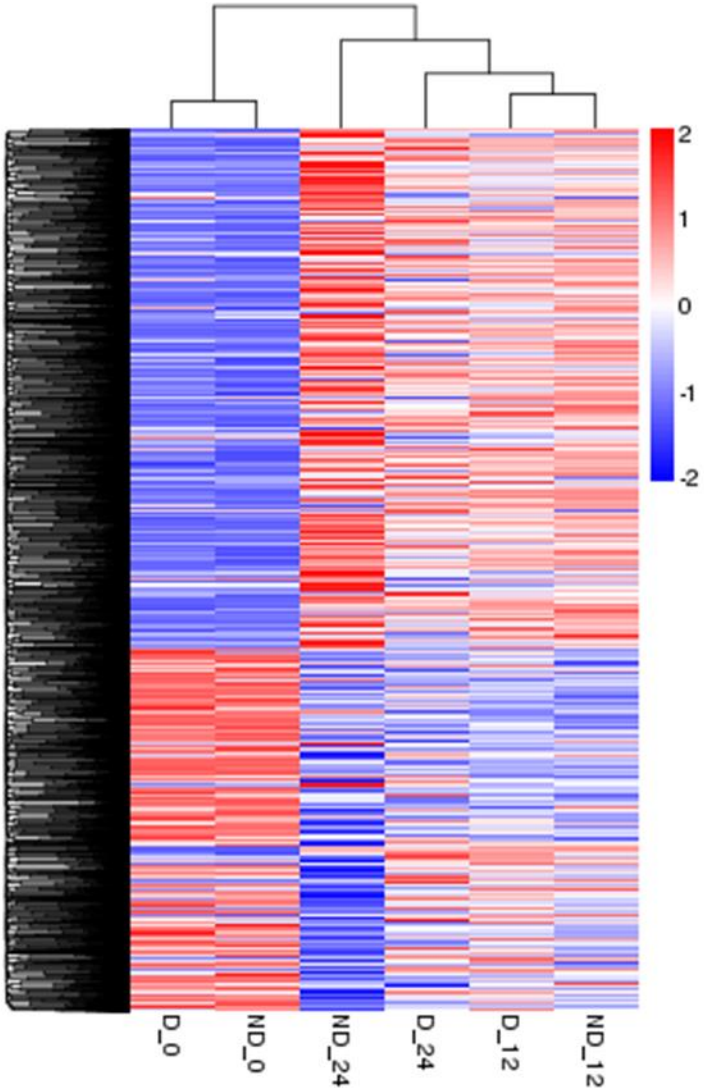


Fig. 4.2 Hierarchical clustering analysis based on similarities in expression patterns of genes expressed in the dry and imbibed seeds of dormant (D) and non-dormant (ND) DH lines at different time points of imbibition (0, 12 and 24 h). The negative and positive numbers on the bar show log10(FPKM+1) values while the colour scale indicates high (red) and low (blue) expression levels.

4.4.5. Analysis of differentially expressed genes (DEGs)

The expression patterns of the genes differentially expressed during imbibition and between the dry and imbibed seeds of the dormant and non-dormant samples was further investigated by pairwise comparison (Appendixes 6, 7). All the DEGs with $P \leq 0.05$ were considered for this analysis without considering a fold-change cut-off value. Our data showed 31,881 DEGs

between dry and 12 HAI dormant seeds with 15,168 genes upregulated and 16,713 genes downregulated in dry seeds while only 4,349 DEGs were observed between 12 and 24 HAI dormant seeds in which 2,134 genes were upregulated and 2,215 genes were downregulated in 12 HAI sample (Appendix 6). Similarly, 32,110 and 13,153 DEGs were detected between dry and 12 HAI, and 12 and 24 HAI non-dormant seeds, respectively. Of these, 14,990 and 5,574 genes were upregulated while 17,120 and 7,579 genes were downregulated in dry and 12 HAI non-dormant seed samples, respectively (Appendix 6).

Comparison between dry dormant and non-dormant seeds revealed 31 DEGs, of which 16 genes were upregulated and 15 genes were downregulated in the dormant seeds (Appendix 7). A total of 1,398 and 21,574 DEGs were observed between dormant and non-dormant seeds imbibed for 12 and 24 h, respectively, of which 616 and 9,383 genes were upregulated while 782 and 12,191 genes were downregulated in the dormant seeds in the respective sets of comparisons. Among the DEGs identified between the dormant and non-dormant seed samples, *TaMXK3.4* (Gene ID: TraesCS1A02G047400), which encodes transmembrane protein and located at 29,280,019 to 29,280,854 bp, showed the highest level of upregulation (1713-fold) in the dormant seeds imbibed for 24 h followed by *TaPM19-A1* (Gene ID: TraesCS4A02G313300; 818-fold upregulation), which encodes ABA induced plasma membrane protein and located at 604.055 to 604.056 Mbp on chromosome 4A (Appendix 8). While *beta-glucosidase 33* (*TaBGLU33*; Gene ID, TraesCS2B02G599900), which is located at 782.684 to 782.688 Mbp, was found to be the most downregulated gene (2671-fold) in dormant seed at 24 HAI followed by a gene encoding the Fn3_like domain-containing protein (1756-fold; Gene ID: TraesCS4D02G252100) and located at 420.939 to 420.944 Mbp. Using cut-off values of ≥ 300-fold change and P≤0.05, no downregulated gene was observed between of dry or 12 h imbibed dormant and non-dormant seed samples (Appendix 8). A very high cut-off value was applied here to narrow down the number of DEGs as a large number of genes showed differential expression between dormant and non-dormant seeds imbibed for 24.

Comparison of the DEGs between the dormant and non-dormant seeds (D_0/ND_0, D_12/ND_12 and D_24/ND_24) showed that the differential expression of 12, 110 and 20,281 genes between dry (D_0/ND_0), 12 h imbibed (D_12/ND_12) and 24 h imbibed (D_24/ND_24) dormant and non-dormant seeds, respectively (Fig. 4.3). Only 3 DEGS were shared between the D_0/ND_0 and D_12/ND_12 comparisons while 8 genes were common between the D_0/ND_0

and D_24/ND_24 comparisons. The highest number of shared DEGs (1,277 genes) was observed between the D_12/ND_12 and D_24/ND_24 comparisons. Only 8 DEGs were found to overlap among all the three sets of comparisons. Of these DEGs, only *TaPM19-A1* showed a consistently high level of upregulation across all the three sets of comparisons (Appendix 8). No significantly downregulated genes were shared among all the comparisons.

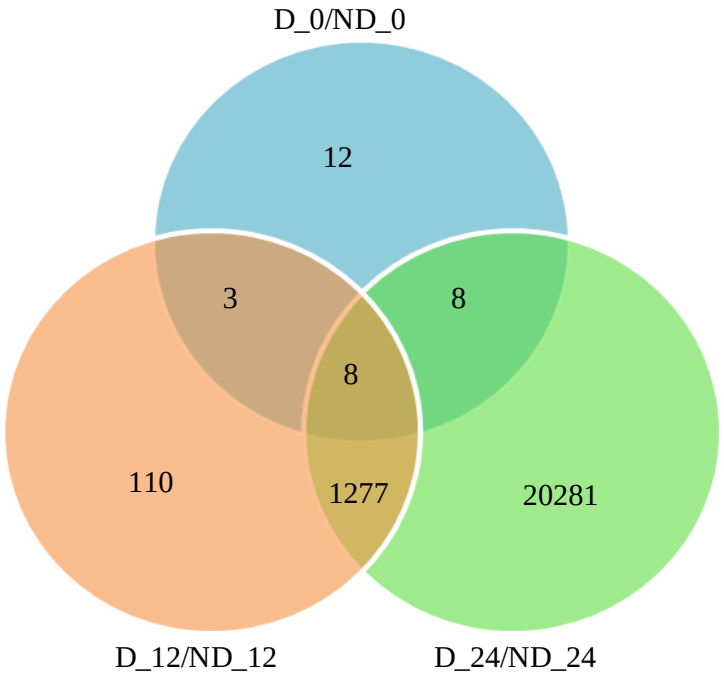


Fig. 4.3 Comparisons of the number of genes differentially expressed between seeds of dormant (D) and non-dormant (ND) DH lines at 0, 12 and 24 h after imbibition (D_0/ND_0, D_12/ND_12 and D_24/ND_24). The sum of the numbers in each circle represents the total number of genes expressed within each comparison and the numbers shown in overlapping areas represents the genes expressed in common between different comparisons.

4.4.6. GO enrichment analysis of DEGs associated with seed dormancy

Our analysis showed that the DEGs are enriched mainly in 30 GO terms (Figs. 4.4, 4.5). A total of 1,071 DEGs between 12 HAI dormant and non-dormant seeds (D_12/ND_12; 439 up- and 632 downregulated genes) were found to be enriched in GO terms representing several biological processes including response to stimulus (GO:0051716; 23 up- and 18 downregulated), response

to chemicals (GO:0042221; 45-up- and 48 downregulated) and response to abiotic stimulus (GO:0009628; 46 up and 9 downregulated) (Fig. 4.4).

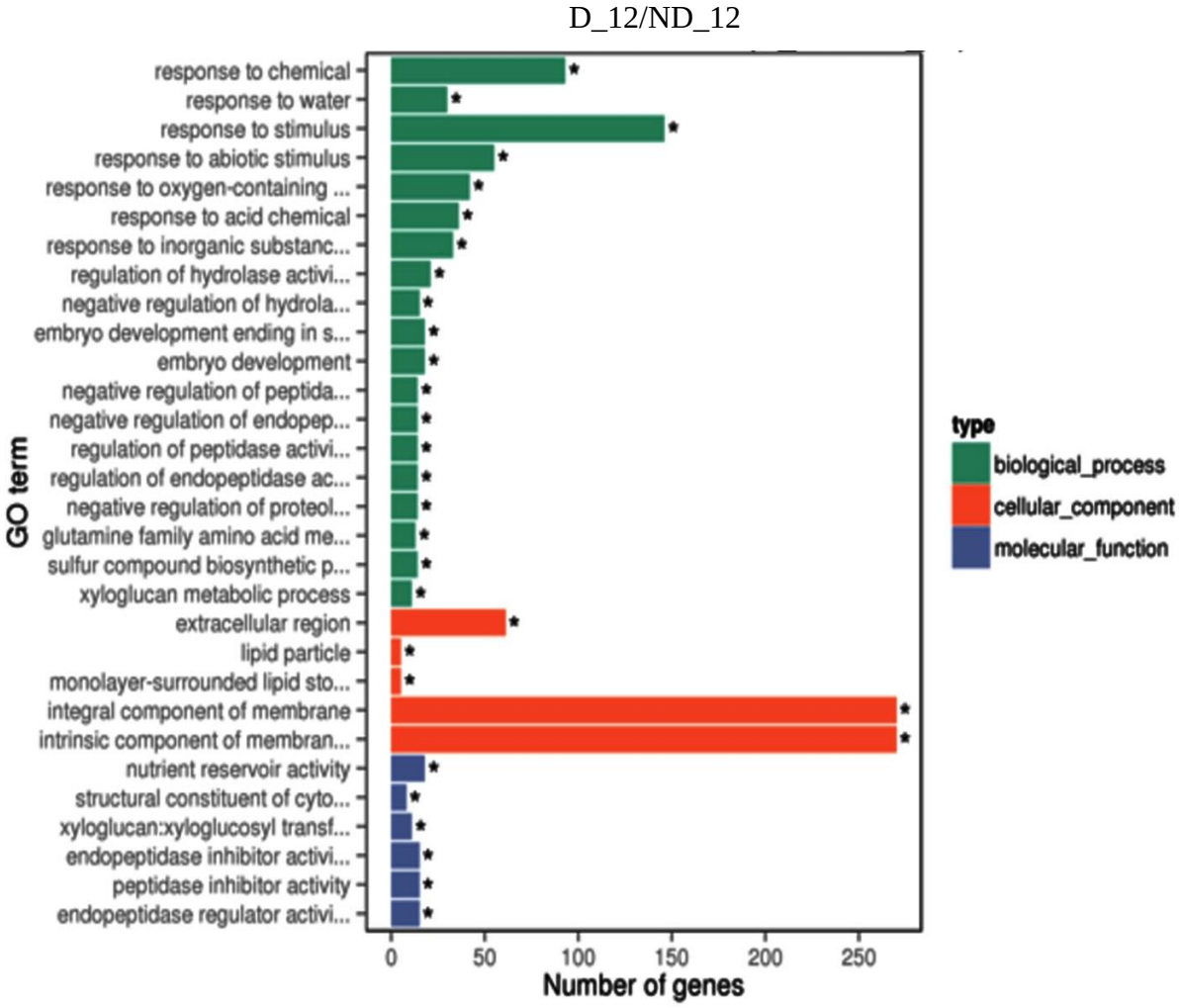


Fig. 4.4 Gene ontology (GO) classes enriched with genes differentially expressed between seeds of the dormant (D) and non-dormant (ND) DH lines at 12 h after imbibition. Bars with different colors represent the GO domains; biological process, cellular component and molecular function. GO terms significantly ($P \leq 0.05$) enriched with the DEGs are marked by asterisks.

With respect to the cellular component GO domain, intrinsic component of membrane (GO:0031224) or integral component of membrane (GO:0016021) GO terms were highly enriched with the DEGs; each GO term was enriched with 106 up- and 164 downregulated genes. Molecular function category was also significantly enriched with DEGs of the same

comparison; however, lower number of DEGs was represented in this domain of gene ontology. For example, the nutrient reservoir activity (GO:0045735) was enriched with 18 DEG (17 up- and 1 downregulated) while the endopeptidase inhibitor activity (GO:0004866)/peptidase inhibitor activity (GO:0030414)/endopeptidase regulator activity (GO:0061135) GO terms were enriched with 15 DEGs (13 up- and 2 downregulated genes).

Changes in GO enrichment patterns were observed with increased duration of imbibition. A total of 16,773 DEGs between 24 HAI dormant and non-dormant seeds (D-24/ND-24; 6,913 up- and 9,860 downregulated genes) were enriched in different GO terms (Appendix 9). Among the 30 top GO terms enriched with the DEGs, more GO terms representing the cellular component domain such as cell (GO:0005623; 1,995 up- and 3,128 downregulated), cell part (GO:0044464; 1,964 up- and 3,098 downregulated), intracellular (GO:0005622; 1,870 up- and 2,916 downregulated) and membrane-bounded organelle (GO:0043227; 1,459 up and 2,094 downregulated) were found to be enriched with DEGs, predominantly with those downregulated in 24 HAI dormant seeds. Three GO terms in biological process domain and two GO terms in the molecular function domain were among the 30 GO terms enriched with DEGs between 24 HAI dormant and non-dormant seeds, of which response to abiotic stimulus (GO:0009628; 283 up- and 284 downregulated) and protein dimerization activity (GO:0046983; 85 up- and 560 downregulated) under biological process and molecular function GO domains, respectively, were enriched with more DEGs (Fig. 4.5).

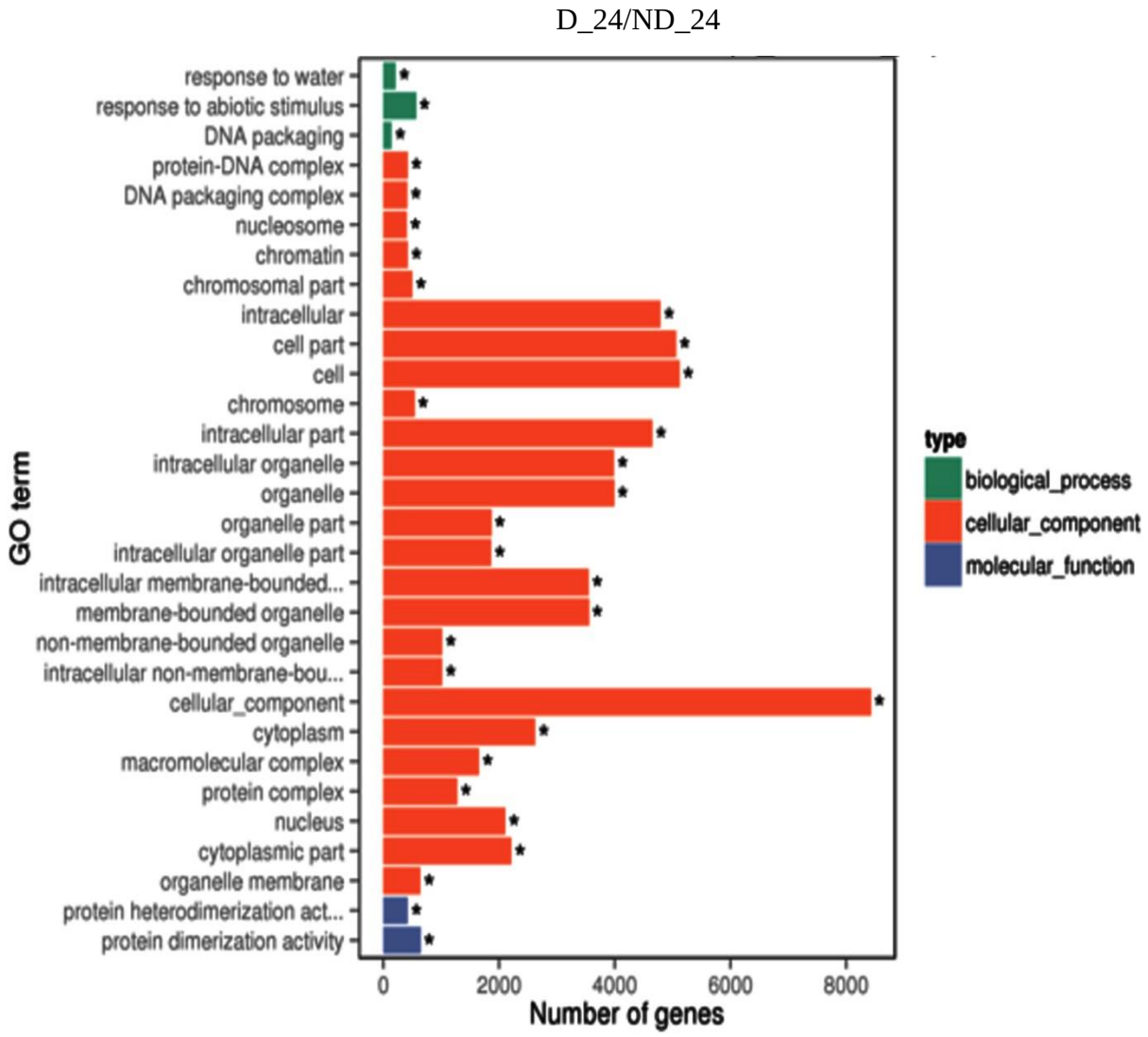


Fig. 4.5 Gene ontology (GO) classes enriched with genes differentially expressed between seeds of the dormant (D) and non-dormant (ND) DH lines at 24 h after imbibition. Bars with different colors represent the GO domains; biological process, cellular component and molecular function. GO terms significantly ($P \leq 0.05$) enriched with the DEGs are marked by asterisks.

4.4.7. KEGG pathway enrichment analysis of DEGs associated with seed dormancy

To identify biological pathways related to seed dormancy, KEGG pathway enrichment analysis was performed between dry/imbibed dormant and non-dormant wheat seeds. The analysis revealed that general metabolic, secondary metabolite biosynthesis, plant hormone signal transduction, starch and sucrose metabolism, phenylpropanoid biosynthesis and spliceosome

pathways are highly enriched with the DEGs between imbibed dormant and non-dormant seeds (Figs. 4.6, 4.7).

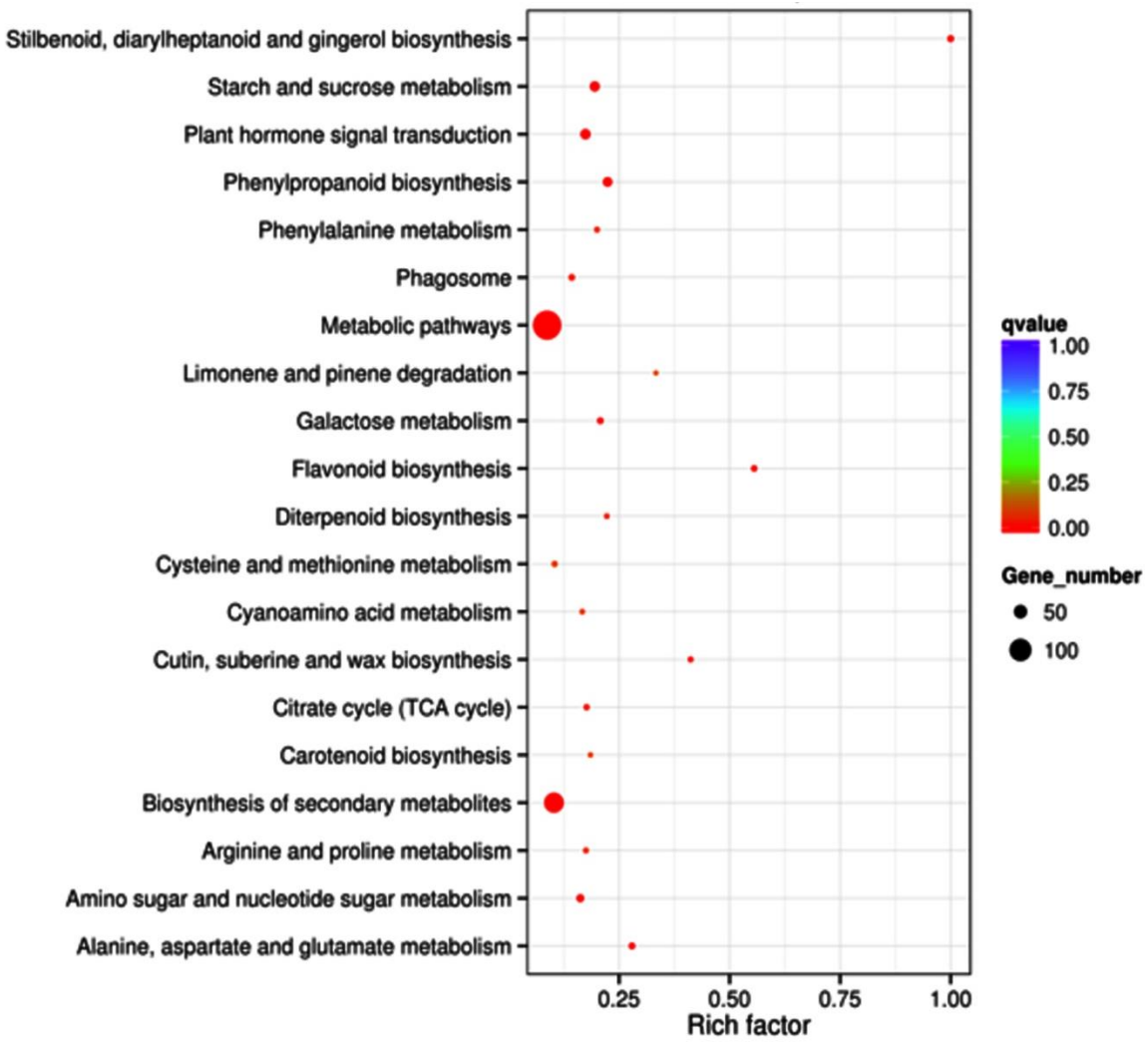


Fig. 4.6 KEGG pathway enrichment analysis of genes differentially expressed between seeds of dormant (D) and non-dormant (ND) DH lines at 12 h after imbibition (D_12/ND_12). The y-axis shows the pathways and the x-axis shows the rich factor. The size of the solid circles in the figure represents the number of genes and the color indicates the q-value.

All these pathways except the spliceosome pathway showed high degree of enrichment with genes differentially expressed between 12 HAI dormant and non-dormant seeds (Fig. 4.6) while the spliceosome pathway was enriched with genes differentially expressed between 24 HAI dormant and non-dormant seeds (Fig. 4.7). The plant hormone signal transduction and secondary metabolite biosynthesis pathways were analyzed in detail due to their role in the regulation of seed dormancy and germination.

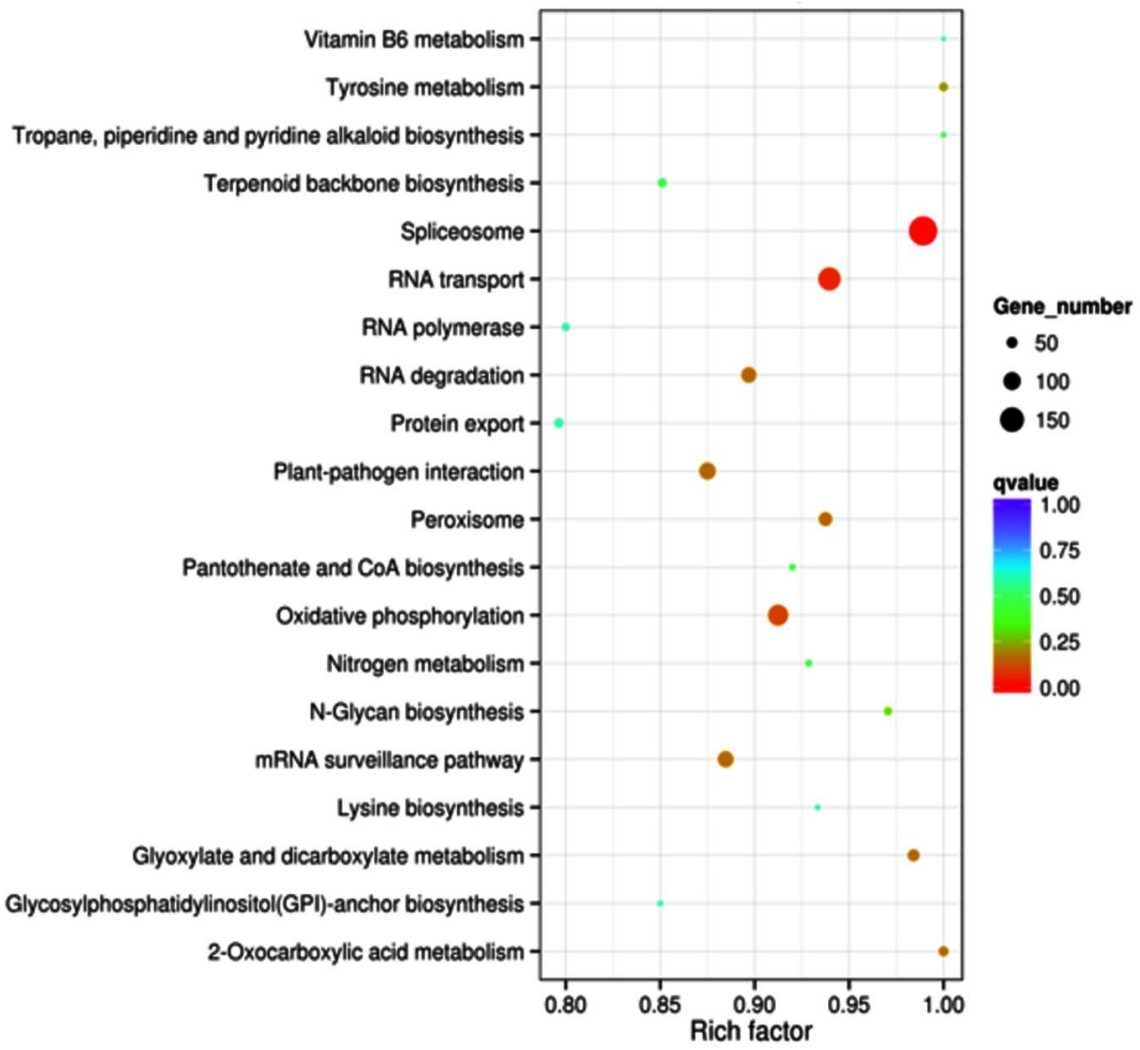


Fig. 4.7 KEGG pathway enrichment analysis of genes differentially expressed between seeds of dormant (D) and non-dormant (ND) DH lines at 24 h after imbibition (D_24/ND_24). The y-axis shows the pathways and the x-axis shows the rich factor. The size of the solid circles in the figure represents the number of genes and the color indicates the q-value.

4.4.8. Pathwayss involved in phytohormone signaling and secondary metabolites biosynthesis

Our data showed that hormone signal transduction pathways including ABA, auxin and ET signaling pathways are enriched with DEGs between dormant and non-dormant seeds imbibed for 24 h (Table 4.2, 4.3). While the secondary metabolite biosynthesis pathway including those of carotenoid, diterpenoid and flavonoid were enriched with genes differentially expressed between dormant and non-dormant seeds imbibed for both 12 and 24 h (Table 4.4). The DEGs with $\geq$ 3-fold change and $P \leq 0.05$ were selected for gene functional annotation. With respect to the plant hormone signal transduction pathway, our DEGs data showed genes involved in ABA and ET signaling pathways are upregulated in dormant seeds, while those involved in auxin signaling pathway are downregulated in the same sample (Table 4.2, 4.3; Fig. 4.8).

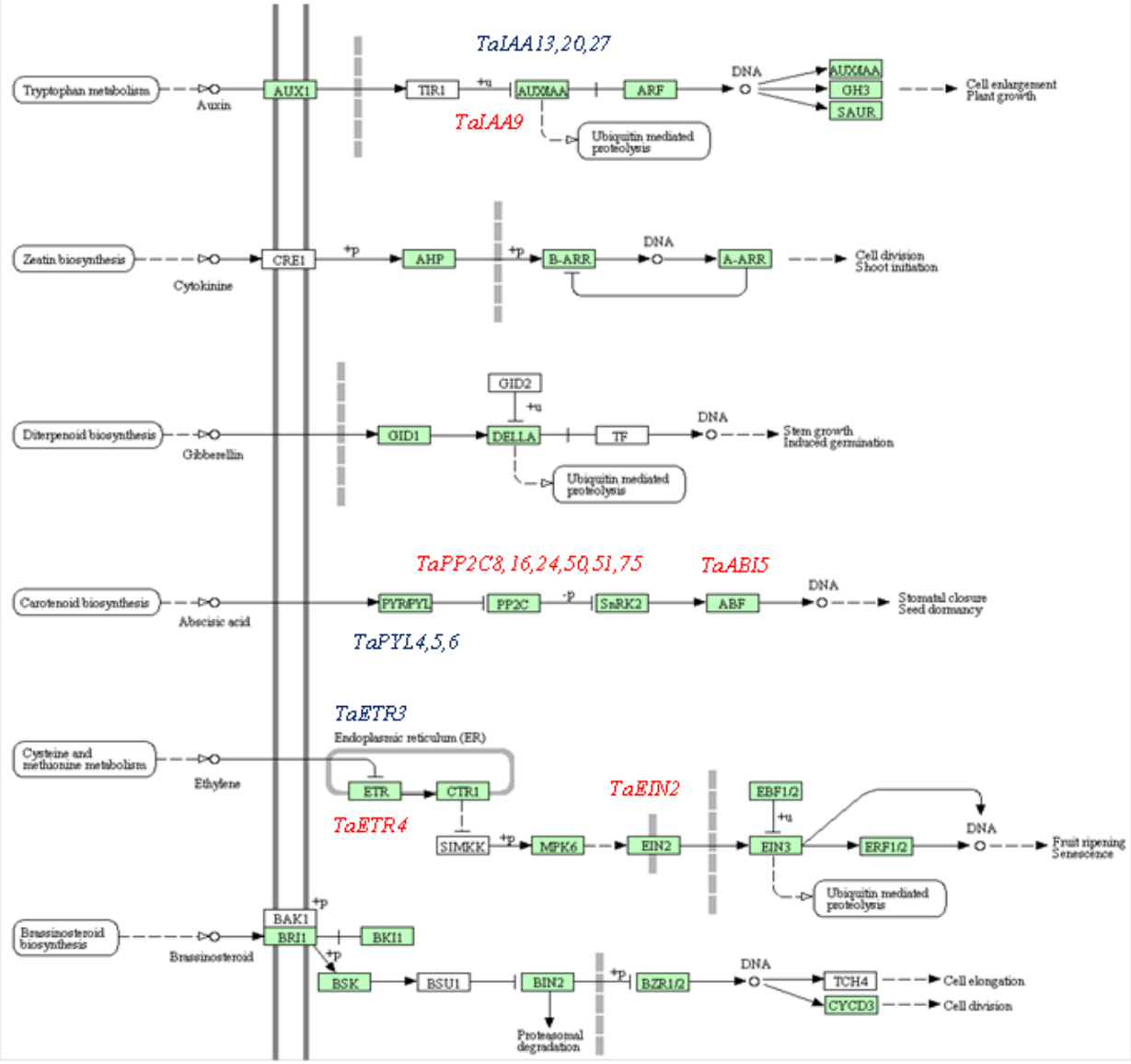


Fig. 4.8 Plant hormone signal transduction pathways enriched with genes differentially expressed between the seeds of dormant and non-dormant DH lines. Boxes highlighted in light green represent differentially expressed genes ($P \leq 0.05$). Genes with red and blue fonts represent those up- and downregulated in dormant seeds (≥ 3-fold), respectively.

Of the DEGs involved in the ABA signaling pathway, three *TaPYL*s (*TaPYL4*, *TaPYL5* and *TaPYL6*) were downregulated in imbibed dormant seeds, mainly at 24 HAI, while five *TaPP2C*s (*TaPP2C8, TaPP2C16, TaPP2C24, TaPP2C51* and *TaPP2C75*), three *TaSnRK2*s (*TaSnRK2.4*, *TaSnRK2.7* and *TaSnRK2.10*) and *TaABI5*s were upregulated in the dormant seeds (Table 4.2).

Table 4.2 Differentially expressed genes (≥ 3-fold, P ≤ 0.05) involved in ABA signaling

IWGSC gene ID	Gene annotation[1]	Abbreviated gene name	Comparisons	Physical position (bp) Start	Physical position (bp) End	Fold change
		Upregulated				
TraesCS2A02G000400	PROTEIN PHOSPHSTASE 2C 75-RELATED	*TaPP2C75-A*	D_24/ND_24	278821	289215	18.58
TraesCS3D02G238000	PROTEIN PHOSPHSTASE 2C 8-RELATED	*TaPP2C8-D*	D_24/ND_24	329494399	329499164	11.18
TraesCS3B02G266200	PROTEIN PHOSPHSTASE 2C 8-RELATED	*TaPP2C8-B*	D_24/ND_24	426994779	427001637	9.71
TraesCS2B02G023600	PROTEIN PHOSPHSTASE 2C 75-RELATED	*TaPP2C75-B*	D_24/ND_24	11006768	11011894	9.66
TraesCS2B02G521800	SNF1-RELATED KINASE 2.7	*TaSnRK2.7-B*	D_24/ND_24	716809651	716816243	9.04
TraesCS1A02G215900	SNF1-RELATED KINASE 2.4	*TaSnRK2.4-A*	D_24/ND_24	381819326	381822599	8.05
TraesCS1D02G218200	SNF1-RELATED KINASE 2.4	*TaSnRK2.4-D*	D_24/ND_24	304838300	304841343	7.88
TraesCS1B02G229400	SNF1-RELATED KINASE 2.4	*TaSnRK2.4-B*	D_24/ND_24	411987863	411990884	7.65
TraesCS5A02G183600	PROTEIN PHOSPHSTASE 2C 24-RELATED	*TaPP2C24-A*	D_24/ND_24	383418336	383419385	6.96
TraesCS5B02G182000	PROTEIN PHOSPHSTASE 2C 24-RELATED	*TaPP2C24-B*	D_24/ND_24	331373823	331375262	6.57
TraesCS5D02G188600	PROTEIN PHOSPHSTASE 2C 24-RELATED	*TaPP2C24-D*	D_24/ND_24	291084497	291086497	6.54
TraesCS4A02G235600	SNF1-RELATED KINASE 2.10	*TaSnRK2.10-A*	D_24/ND_24	544198215	544201187	5.52
TraesCS1B02G441400	PROTEIN PHOSPHSTASE 2C 51	*TaPP2C51-B*	D_24/ND_24	662520537	662522241	5.51
TraesCS3A02G237800	PROTEIN PHOSPHSTASE 2C 8-RELATED	*TaPP2C8-A*	D_24/ND_24	444534322	444538650	5.33
TraesCS3D02G249000	PROTEIN PHOSPHSTASE 2C 16	*TaPP2C16-D*	D_24/ND_24	348158072	348161269	5.03
TraesCS1A02G411200	PROTEIN PHOSPHSTASE 2C 51	*TaPP2C51-A*	D_24/ND_24	571986755	571989201	4.39
TraesCS3B02G404500	ABSCISIC ACID-INSENSITIVE 5	*TaABI5-B.1*	D_24/ND_24	639553616	639555147	4.31
TraesCS3A02G249000	PROTEIN PHOSPHSTASE 2C 16	*TaPP2C16-A*	D_24/ND_24	467153054	467156209	4.23
TraesCS1D02G419000	PROTEIN PHOSPHSTASE 2C 51	*TaPP2C51-D*	D_24/ND_24	476207922	476210244	4.18
TraesCS3B02G404600	ABSCISIC ACID-INSENSITIVE 5	*TaABI5-B.2*	D_24/ND_24	639752637	639756075	3.89
TraesCS3D02G364900	ABSCISIC ACID-INSENSITIVE 5	*TaABI5-D*	D_24/ND_24	478738607	478740450	3.77
TraesCS1A02G358600	PROTEIN PHOSPHSTASE 2C 50	*TaPP2C50*	D_24/ND_24	540336968	540339603	3.76
TraesCS3A02G371800	ABSCISIC ACID-INSENSITIVE 5	*TaABI5-A*	D_24/ND_24	622277558	622279313	3.68
TraesCS3B02G277900	PROTEIN PHOSPHSTASE 2C 16	*TaPP2C16-B*	D_24/ND_24	447981062	447984222	3.66
TraesCS4D02G078100	SNF1-RELATED KINASE 2.10	*TaSnRK2.10-D*	D_24/ND_24	52470433	52473315	3.52
TraesCS2A02G493800	SNF1-RELATED KINASE 2.7	*TaSnRK2.7-A*	D_24/ND_24	726029783	726034011	3.25
TraesCS5D02G188600	PROTEIN PHOSPHSTASE 2C 24-RELATED	*TaPP2C24-D*	D_12/ND_12	291084497	291086497	3.07
		Downregulated				
TraesCS1D02G293600	PYRABACTIN RESISTANCE 1-LIKE 4	*TaPYL4-D*	D_24/ND_24	392696780	392697872	-3.22
TraesCS1B02G306800	PYRABACTIN RESISTANCE 1-LIKE 4	*TaPYL4-B*	D_24/ND_24	529998098	529998682	-3.54
TraesCS1A02G297600	PYRABACTIN RESISTANCE 1-LIKE 4	*TaPYL4-A*	D_24/ND_24	492652071	492652999	-6.39
TraesCS3B02G380300	PYRABACTIN RESISTANCE 1-LIKE 5	*TaPYL5*	D_24/ND_24	597816092	597816980	-12.74
TraesCS4D02G191200	PYRABACTIN RESISTANCE 1-LIKE 6	*TaPYL6-D*	D_24/ND_24	334632132	334632806	-14.81
TraesCS4B02G189800	PYRABACTIN RESISTANCE 1-LIKE 6	*TaPYL6-B*	D_24/ND_24	412590620	412591294	-17.72
TraesCS4A02G114400	PYRABACTIN RESISTANCE 1-LIKE 6	*TaPYL6-A*	D_24/ND_24	138832030	138833256	-55.79

[1]Annotation based on PANTHER16.0, RAPDB and TAIR databases; bp, base pair

In the case of auxin signaling genes, *auxin responsive*/*indole-3-acetic acid inducible* (*TaAux*/*IAA*) genes (*TaIAA9*, *TaIAA13*, *TaIAA20* and *TaIAA27*) showed differential expression, of which *TaIAA13-B* showed the highest downregulation in dormant seeds (~7-fold) while *TaIAA9* was found to be the only upregulated (3-fold) gene in the same seed sample (Table 4.3). With respect to ET signaling genes, *ethylene receptor* (*TaETR4*) and *ethylene insensitive* genes (*TaEIN2*) were upregulated in 24 h imbibed dormant seeds while *TaETR3* was the only downregulated gene in the same sample (Table 4.3).

Table 4.3 Differentially expressed genes (≥ 3-fold, P ≤ 0.05) involved in auxin and ethylene signaling

IWGSC gene ID	Gene annotation[1]	Abbreviated gene name	Comparisons	Physical position (bp)		Fold change
				Start	End	
		Auxin signaling				
TraesCS6A02G373300	INDOLE-3-ACETIC ACID INDUCIBLE 9	*TaIAA9*	D_24/ND_24	596739621	596740759	3.01
TraesCS4B02G108600	INDOLE-3-ACETIC ACID INDUCIBLE 27	*TaIAA27*	D_24/ND_24	119339061	119346199	-3.30
TraesCS5A02G382600	INDOLE-3-ACETIC ACID INDUCIBLE 13	*TaIAA13-A*	D_24/ND_24	580464947	580467685	-4.26
TraesCS7B02G029000	INDOLE-3-ACETIC ACID INDUCIBLE 20	*TaIAA20*	D_24/ND_24	27486147	27487108	-5.01
TraesCS5D02G392000	INDOLE-3-ACETIC ACID INDUCIBLE 13	*TaIAA13-D*	D_24/ND_24	460393490	460396369	-5.62
TraesCS5B02G386800	INDOLE-3-ACETIC ACID INDUCIBLE 13	*TaIAA13-B*	D_24/ND_24	565928821	565931697	-6.62
		ET signaling				
TraesCS6B02G314600	ETHYLENE RECEPTOR 4	*TaETR4-B.1*	D_24/ND_24	562487341	562490051	12.39
TraesCS4B02G038100	ETHYLENE INSENSITIVE 2	*TaEIN2-B.1*	D_24/ND_24	27530536	27536178	12.21
TraesCS4B02G381200	ETHYLENE INSENSITIVE 2	*TaEIN2-B.2*	D_24/ND_24	660940620	660944513	11.59
TraesCS5A02G547500	ETHYLENE INSENSITIVE 2	*TaEIN2-A*	D_24/ND_24	701563703	701568791	9.93
TraesCS6D02G266200	ETHYLENE RECEPTOR 4	*TaETR4-D*	D_24/ND_24	375687004	375689647	8.25
TraesCSU02G131600	ETHYLENE INSENSITIVE 2	*TaEIN2-U*	D_24/ND_24	111040871	111045292	8.10
TraesCS6A02G329400	ETHYLENE RECEPTOR 4	*TaETR4-A.1*	D_24/ND_24	562685884	562688605	5.91
TraesCS6A02G285800	ETHYLENE RECEPTOR 4	*TaETR4-A.2*	D_24/ND_24	517557089	517559867	4.77
TraesCS6B02G360200	ETHYLENE RECEPTOR 4	*TaETR4-B.2*	D_24/ND_24	632179674	632182484	4.58
TraesCS4D02G036000	ETHYLENE INSENSITIVE 2	*TaEIN2-D*	D_24/ND_24	16098177	16105037	3.60
TraesCS6D02G383600	ETHYLENE RECEPTOR 3	*TaETR3*	D_24/ND_24	462483422	462487522	-3.02

[1]Annotation based on PANTHER16.0, RAPDB and TAIR databases; bp, base pair; the colour green represents upregulated genes and pink represents downregulated genes

Of the DEGs involved in carotenoid biosynthesis, the *NCEDs* such as *TaNCED6* and *TaNCED3*, *phytoene synthase*s (*PSY*s; *TaPSY1A1*, *TaPSY7B* and *TaPSY1D1*), *beta-carotene 3-hydroxylase 1* (*TaBCH1*) and *usually multiple acids move in and out transporter 6* (*TaUMAMIT6*) were upregulated in dormant seeds at both 12 and 24 HAI (Table 4.4, Fig. 4.9). The *TaUMAMIT6* was found to be the only carotenoid biosynthesis gene downregulated (5-fold) in dormant seeds at 24 HAI.

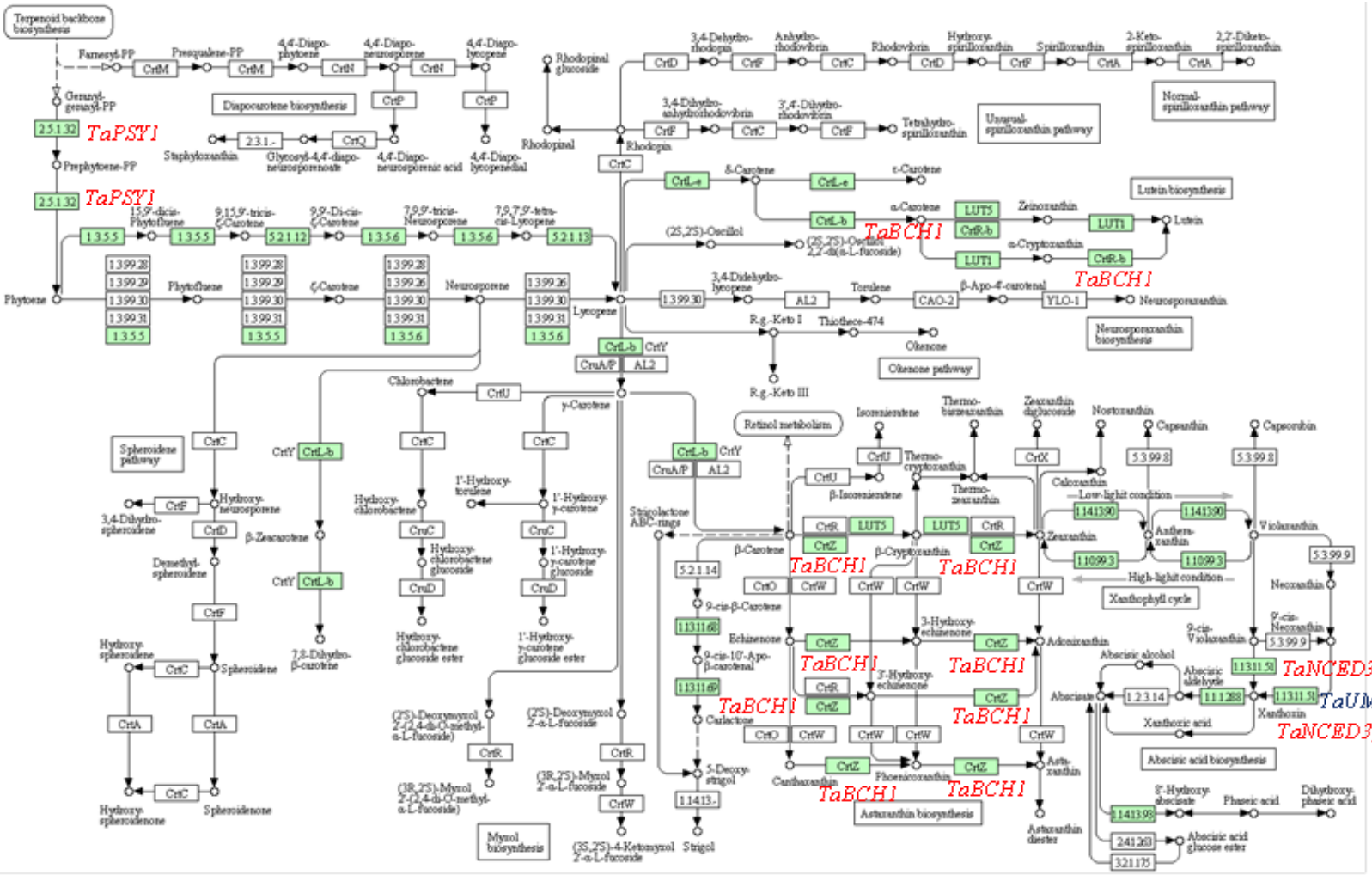


Fig. 4.9 Carotenoid biosynthesis pathway enriched with genes differentially expressed between the seeds of dormant and non-dormant double haploid lines. Boxes highlighted in light green represent differentially expressed genes (P≤0.05). Genes with red and blue fonts represent those up- and downregulated in dormant seeds (≥3-fold), respectively.

Genes involved in the diterpenoid biosynthesis pathways such as those involved in gibberellin metabolism and flavonoid biosynthesis also showed differential expression. Of the GA metabolism genes, the GA biosynthesis gene *TaGA20ox1* showed over 3.5-fold downregulation in dormant seeds at 12 HAI while the GA catabolism gene *TaGA2ox3* showed 45-fold upregulation in dormant seeds at 24 HAI (Table 4.4, Fig. 4.10).

Table 4.4 Differentially expressed genes (≥ 3-fold, P ≤ 0.05) involved in the biosynthesis of secondary metabolites

IWGSC gene ID	Gene annotation[1]	Abbreviated gene name	Comparisons	Physical position (bp)		Fold change
				Start	End	
Carotenoid						
TraesCS2A02G012300	9-CIS-EPOXYCAROTENOID DIOXYGENASE 6	*TaNCED6-A*	D_24/ND_24	4907004	4909365	194.01
TraesCS2D02G013000	9-CIS-EPOXYCAROTENOID DIOXYGENASE 6	*TaNCED6-D*	D_24/ND_24	6507849	6509790	40.19
TraesCS5B02G029300	9-CIS-EPOXYCAROTENOID DIOXYGENASE 3	*TaNCED3-B*	D_24/ND_24	29337289	29339976	21.46
TraesCS5D02G038800	9-CIS-EPOXYCAROTENOID DIOXYGENASE 3	*TaNCED3-D*	D_24/ND_24	38723664	38726233	9.15
TraesCS2D02G013000	9-CIS-EPOXYCAROTENOID DIOXYGENASE 6	*TaNCED6-D*	D_12/ND_12	6507849	6509790	8.18
TraesCS2A02G012300	9-CIS-EPOXYCAROTENOID DIOXYGENASE 6	*TaNCED6-A*	D_12/ND_12	4907004	4909365	6.36
TraesCS7B02G482000	PHYTOENE SYNTHASE 7B	*TaPSY7B*	D_12/ND_12	739442097	739446001	5.29
TraesCS7A02G557300	PHYTOENE SYNTHASE 1 A1	*TaPSY1A1*	D_24/ND_24	729397165	729401834	5.27
TraesCS7D02G553300	PHYTOENE SYNTHASE 1 D1	*TaPSY1D1*	D_24/ND_24	636765931	636771068	3.21
TraesCS2D02G393900	BETA CAROTENE 3-HYDROXYLASE 1	*TaBCH1*	D_24/ND_24	503582593	503584811	3.23
TraesCS6B02G298700	USUALLY MULTIPLE ACIDS MOVE IN AND OUT TRANSPORTERS 6	*TaUMAMIT6*	D_24/ND_24	536811349	536816690	-5.15
Diterpenoid						
TraesCS1B02G420800	GIBBERELLIN 2-OXIDASE 3	*TaGA2OX3*	D_24/ND_24	644525709	644527654	44.94
TraesCS7D02G360700	ENT-KAURENE OXIDASE-D	*TaKO-D*	D_24/ND_24	463698536	463704290	3.75
TraesCS4A02G319100	GIBBERELLIN 20 OXIDASE 1-A	*TaGA20OX1-A*	D_12/ND_12	608043459	608044983	-3.35
TraesCS5D02G566200	GIBBERELLIN 20 OXIDASE 1-D	*TaGA20OX1-D*	D_12/ND_12	564586783	564588466	-3.54
Flavonoid						
TraesCS2B02G507500	BANYULS	*TaBAN*	D_24/ND_24	704129187	704130949	9.91
TraesCS6D02G004300	ANTHOCYANIDIN SYNTHASE-D1	*TaANS-D1*	D_24/ND_24	2106676	2108293	3.10

[1]Annotation based on PANTHER16.0, RAPDB and TAIR databases; bp, base pair; genes in bold are present in both comparisons, D_12/ND_12 and D_24/ND_24; the colour green represents upregulated genes and pink represents downregulated genes

Similarly, genes involved in flavonoid biosynthesis such as *BANYULS* (*TaBAN)* and *anthocyanidin synthase-D1* (*TaANS-D1*) showed upregulation in dormant seed at 24 HAI (Table 4.4, Appendix 10). No gene involved in flavonoid biosynthesis was found to be downregulated in the dormant seeds.

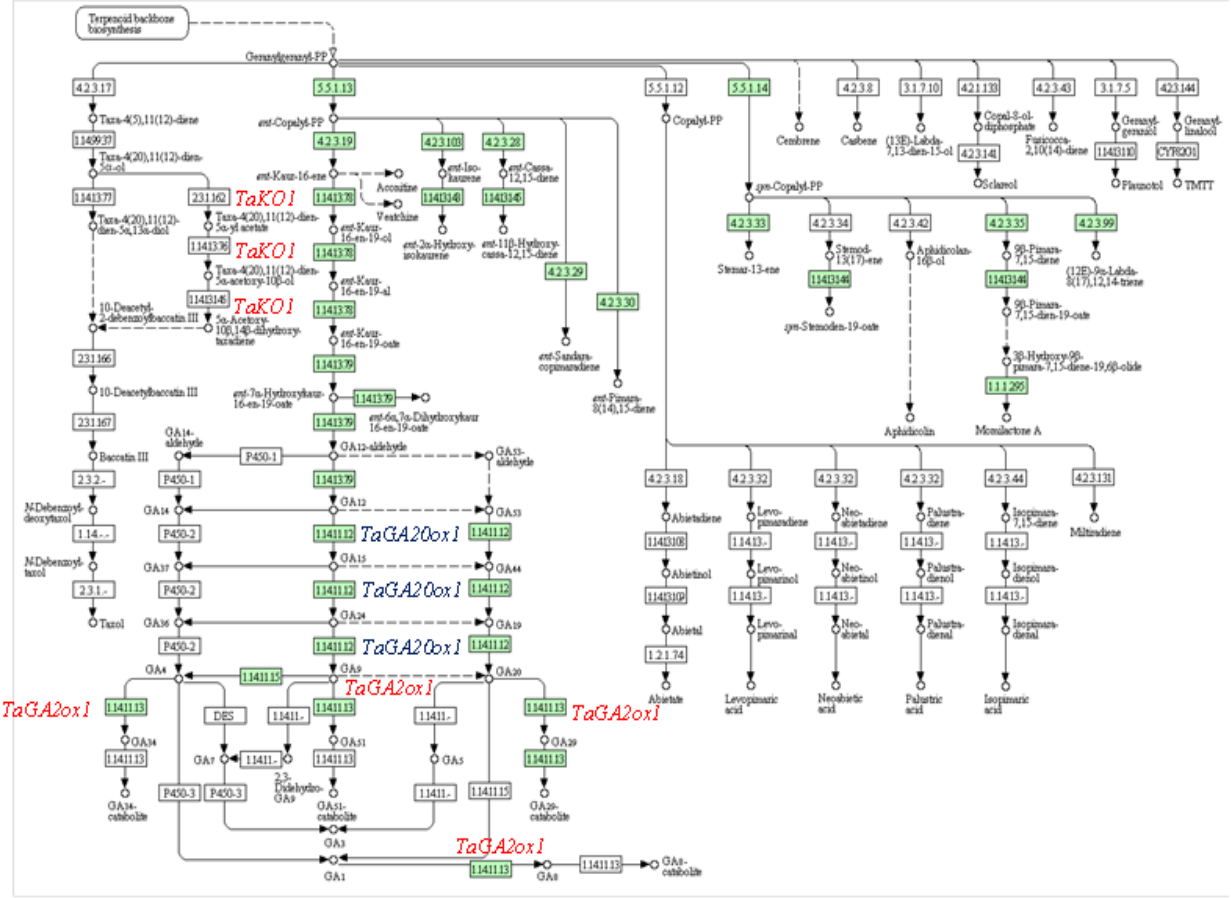


Fig. 4.10 Diterpenoids biosynthetic pathway enriched with genes differentially expressed between the seeds of dormant and non-dormant double haploid lines. Boxes highlighted in light green represent differentially expressed genes (P≤0.05). Genes with red and blue fonts represent those up- and downregulated in dormant seeds (≥3-fold), respectively.

4.4.9. Differentially expressed genes within QTLs

Our previous genetic study identified a major QTL on 4A and three minor QTLs on 1D, 6B and 6D chromosomes as loci regulating pre-harvest sprouting resistance (Liton et. al. 2021), and this study identified the DEGs located within these QTLs. Due to the major contribution of *QPhs.umb-4A* for PHS resistance, all the DEGs with a ≥2-fold and P≤0.05 were considered for further analysis (Table 4.5). Whereas only DEGs with ≥10-fold and P≤0.05 were considered for *QPhs.umb-1D*, *QPhs.umb-6B* and *QPhs.umb-6D* since a large number of DEGs are detected within these QTLs (Table 4.5, 4.6, 4.7). Using the criterion described above a total of 25 DEGs were located within the 4A QTL. Of these, *TaPM19-A1*, a gene encoding the ABA induced plasma membrane protein, showed the highest upregulation (818-fold) in 24h imbibed dormant seeds followed by *TaPM19-A2* (270-fold) (Table 4.5). While an uncharacterized gene (Gene ID: TraesCS4A02G317200), which is located at 606.608 to 606.609 Mbp of the 4A QTL, showed the highest downregulation (39.5-fold) in dormant seeds at 24 HAI followed by *GnK2 domain containing receptor-like kinase-8* (*TaGnk2RLK-8*; 23-fold). In addition, a well known GA biosynthesis gene *TaGA20ox1-A* that showed downregulation in imbibed dormant seeds at both 12 and 24 HAI was found to be localized at 608.043 to 608.044 Mbp of the 4A QTL (Table 4.5).

Table 4.5 Differentially expressed genes (≥ 2-fold, P ≤ 0.05) genes located in *QPhs.umb-4A*

			Physical position (bp)		D_0/ND_0		D_12/ND_12		D_24/ND_24	
IWGSC gene ID	Gene annotation[1]	Abbreviated gene name	Start	End	Fold change	Significant	Fold change	Significant	Fold change	Significant
TraesCS4A02G313300	ABA INDUCED PLASMA MEMBRANE 19-A1	*TaPM19-A1*	604055487	604056041	3.78	UP	N.D.	UP	818.14	UP
TraesCS4A02G313400	ABA INDUCED PLASMA MEMBRANE 19-A2	*TaPM19-A2*	604068519	604069061	-1.42	FALSE	23.12	UP	269.62	UP
TraesCS4A02G319700	11-BETA-HYDROXYSTEROID DEHYDROGENASE-LIKE 4A	*TaHSD4A*	608560764	608562892	-1.78	DOWN	1.09	FALSE	36.84	UP
TraesCS4A02G315800	MEDIATOR OF RNA POLYMERASE II TRANSCRIPTION SUBUNIT 13	*TaMED13*	605717570	605733559	1.37	FALSE	1.72	UP	2.85	UP
TraesCS4A02G311700	ACT DOMAIN REPEATS 10	*TaACR10*	603487335	603490592	2.37	FALSE	1.92	FALSE	2.70	UP
TraesCS4A02G320200	Ribosom_S30AE_C domain-containing protein	-	608866595	608869029	-1.25	FALSE	1.08	FALSE	2.58	UP
TraesCS4A02G312200	ATG8-INTERACTING PROTEIN 1	*TaATI1*	603530872	603533012	-1.48	FALSE	1.37	FALSE	2.44	UP
TraesCS4A02G313600	UBIQUITIN-CONJUGATING ENZYME 4	*TaUCB4*	604101159	604103952	-1.13	FALSE	1.39	FALSE	2.34	UP
TraesCS4A02G322600	E3 UBIQUITIN-PROTEIN LIGASE LISTERIN	*TaLTN1*	610618336	610629297	1.18	FALSE	1.39	FALSE	2.29	UP
TraesCS4A02G322700	F-box/LRR-repeat protein	*TaFBL75*	610639750	610642778	1.88	FALSE	1.53	FALSE	2.26	UP
TraesCS4A02G323300	F-box/LRR-repeat protein	*TaFBL75*	611200914	611204084	1.00	FALSE	1.26	FALSE	2.03	UP
TraesCS4A02G311300	ASCORBATE TRANSPORTER	*TaPHT4;4*	603458204	603463050	-1.13	FALSE	1.22	FALSE	-2.13	DOWN
TraesCS4A02G318900	SASA domain-containing protein	-	607432144	607434998	1.32	FALSE	-1.51	FALSE	-2.20	DOWN
TraesCS4A02G314400	RESISTANCE TO PSEUDOMONAS SYRINGAE 3	*TaRPM3*	604845889	604850471	-1.42	FALSE	-2.41	DOWN	-2.91	DOWN
TraesCS4A02G318800	SASA domain-containing protein	-	607427390	607428603	1.49	FALSE	-1.75	FALSE	-3.11	DOWN
TraesCS4A02G319100	GIBBERELLIN 20 OXIDASE 1-A	*TaGA20OX1A*	608043459	608044983	1.31	FALSE	-3.35	DOWN	-3.15	DOWN
TraesCS4A02G312600	CYCLOPHILIN 19-3	*TaCYP19-3*	603637101	603639737	1.22	FALSE	-1.33	FALSE	-3.37	DOWN
TraesCS4A02G323700	ADP-RIBOSYLATION FACTOR GTPASE-ACTIVATING PROTEIN	*TaAGD10*	611579524	611583216	1.52	FALSE	-1.84	DOWN	-4.14	DOWN
TraesCS4A02G323800	Gamma-glutamylcyclotransferase family protein	-	611589781	611592958	-1.24	FALSE	-1.06	FALSE	-4.66	DOWN
TraesCS4A02G314300	RECOGNITION OF PERONOSPORA PARASITICA 13	*TaRPP13*	604772917	604775106	-1.79	FALSE	-3.56	DOWN	-6.79	DOWN
TraesCS4A02G313200	INDOLE-3-PYRUVATE MONOOXYGENASE YUCCA3-RELATED	*TaYUC3*	603979784	603981997	1.07	FALSE	1.79	FALSE	-8.38	DOWN
TraesCS4A02G323500	GnK2 DOMAIN CONTAINING RECEPTOR-LIKE KINASE-8	*TaGnk2RLK-8*	611268434	611271719	N.D.	N.D.	-2.10	FALSE	-8.75	DOWN
TraesCS4A02G316700	Uncharacterized		606523442	606525449	-2.24	FALSE	-4.75	FALSE	-9.28	DOWN
TraesCS4A02G323400	GnK2 DOMAIN CONTAINING RECEPTOR-LIKE KINASE-7	*TaGnk2RLK-7*	611264184	611266257	N.D.	N.D.	-27.01	FALSE	-23.19	DOWN
TraesCS4A02G317200	Uncharacterized	-	606608730	606609562	N.D.	N.D.	1.02	FALSE	-39.48	DOWN

[1]Annotation based on PANTHER16.0, RAPDB and TAIR databases; N.D., not determined due to absence of read in one of two samples; FALSE, insignificant; bp, base pair

Of the DEGs located within the minor QTLs, genes downregulated in imbibed dormant seeds are predominant (Tables 4.6, 4.7). Within 6B QTL, an uncharacterized gene (Gene ID: TraesCS6B02G325000), located at 574,627,077 to 574,627,969 bp of the QTL showed the highest upregulation (68-fold) in dormant seeds at both imbibition time points. An ethylene receptor gene, *TaETR4*, which is also located at 562.487 to 562.490 Mbp of 6B QTL, showed upregulation (12-fold) in dormant seeds at 24 HAI. In contrast, a *RNA helicase 8* (*TaRH8*), which is located at 477.217 to 477.226 Mbp of 6B QTL showed downregulation (181-fold) in dormant seeds at 24 HAI (Table 4.6). In 6D QTL, *late embryogenesis abundant protein 4* (*TaLEA4*), which is located at 275.664 to 275.665 Mbp, showed the highest upregulation in dormant seeds at 12 and 24 HAI (11- and 28-fold, respectively) followed by *oxidative stress 3* (*TaOXS3*), which is located at 343.557 to 343.558 Mbp and exhibited 17-fold upregulation at 24 HAI (Table 4.7). Whereas *histone H4* (*TaHIS4*), located at 343,676,509 to 343,676,820 bp, showed the highest downregulation (81-fold) in dormant seeds at 24 HAI. No GO term was found to be significantly enriched by the DEGs located within the four QTLs (Appendixes 11, 12, 13).

Table 4.6 Differentially expressed genes (≥ 10-fold, P ≤ 0.05) genes located in *QPhs.umb-6B*

			Physical position (bp)		D_12/ND_12		D_24/ND_24	
IWGSC gene ID	Gene annotation[1]	Abbreviated gene name	Start	End	Fold change	Significant	Fold change	Significant
TraesCS6B02G325000	Uncharacterized	-	574627077	574627969	3.25	UP	67.95	UP
TraesCS6B02G296400	UBIQUINOL OXIDASE 1C	*TaAOX1C*	532150706	532152010	1.46	FALSE	14.22	UP
TraesCS6B02G326000	STYLE LENGTH 2	*TaSYL2*	575330987	575331220	1.68	FALSE	12.50	UP
TraesCS6B02G314600	ETHYLENE RECEPTOR 4	*TaETR4-B.1*	562487341	562490051	2.79	FALSE	12.39	UP
TraesCS6B02G282100	GLYCEROL-3-PHOSPHATE TRANSPORTER 1	*TaGLPT1*	508535186	508537822	2.43	UP	10.66	UP
TraesCS6B02G317100	DOF ZINC FINGER PROTEIN DOF1.2	*TaDOF1.2*	564992535	564993867	-1.30	FALSE	-10.02	DOWN
TraesCS6B02G291800	UBIQUITIN-SPECIFIC PROCESSING PROTEASE 7-2	*TaUBP7-2*	525105470	525110758	-1.15	FALSE	-11.34	DOWN
TraesCS6B02G331500	TYPE I INOSITOL POLYPHOSPHATE 5-PHOSPHATASE 5	*TaIP5P5*	582825538	582828306	-2.44	FALSE	-11.56	DOWN
TraesCS6B02G288800	GLYCOSYLTRANSFERASE	-	519847805	519853059	1.08	FALSE	-11.78	DOWN
TraesCS6B02G295400	Uncharacterized	-	530565991	530570039	1.42	FALSE	-11.98	DOWN
TraesCS6B02G276800	S-ADENOSYLMETHIONINE SYNTHETASE 2	*TaSAM2*	501061764	501064441	-2.80	FALSE	-12.02	DOWN
TraesCS6B02G303500	ROOT HAIR DEFECTIVE-SIX LIKE 2	*TaRSL2*	543825432	543826839	2.66	FALSE	-14.40	DOWN
TraesCS6B02G307400	EUKARYOTIC TRANSLATION INITIATION FACTOR 3 SUBUNIT B	*TaTIF3B*	551030907	551034918	-25.53	FALSE	-16.34	DOWN
TraesCS6B02G247400	SHORT-CHAIN DEHYDROGENASE/REDUCTASE 1	*TaSDR1*	440555383	440559379	N.D.	N.D.	-17.79	DOWN
TraesCS6B02G313200	EARLY NODULIN-LIKE PROTEIN 24	*TaENODL24*	561250167	561254138	-2.50	FALSE	-17.87	DOWN
TraesCS6B02G289300	FCS-LIKE ZINC FINGER 3	*TaFLZ3*	520545967	520546807	-18.97	FALSE	-19.95	DOWN
TraesCS6B02G272200	Uncharacterized	-	491604372	491604755	N.D.	N.D	-21.49	DOWN
TraesCS6B02G273400	DEHYDRIN 1	*TaDHN1*	493352704	493354073	1.31	FALSE	-22.89	DOWN
TraesCS6B02G275700	EXPANSIN-B3	*TaEXPB3*	498542519	498544160	N.D.	N.D.	-37.59	DOWN
TraesCS6B02G311300	Uncharacterized	-	558287373	558300494	-5.68	FALSE	-39.34	DOWN
TraesCS6B02G259000	PLASMA MEMBRANE INTRINSIC PROTEIN 2;2	*TaPIP2;2*	467447950	467450705	N.D.	N.D.	-43.23	DOWN
TraesCS6B02G288000	HISTONE H4	*TaHIS4*	518678603	518679295	1.03	FALSE	-56.29	DOWN
TraesCS6B02G284500	PROLINE-RICH RECEPTOR-LIKE PROTEIN KINASE 4	*TaPERK4*	513135335	513136893	N.D.	N.D.	-67.34	DOWN
TraesCS6B02G265300	RNA HELICASE 8	*TaRH8*	477217134	477226063	-31.65	FALSE	-181.37	DOWN

[1]Annotation based on PANTHER16.0, RAPDB and TAIR databases; N.D., not determined due to absence of read in one of two samples; FALSE, insignificant; bp, base pair

Within 1D QTL, a gene encoding cytochrome P450 protein, *cytochrome p450, family 71, subfamily b, polypeptide 11* (*TaCYP71B11*), which is located at 467.852 to 467.854 Mbp, showed the highest upregulation (17-fold) in dormant seeds at both 12 and 24 HAI followed by *delay of germination 1 like 2* (*TaDOG1L-2*) gene, which is located at 466.493 to 466.494 Mbp and showed 16-fold upregulation at 24 HAI. The highest downregulation (38.5-fold) was observed in an uncharacterized gene (Gene ID: TraesCS1D02G410700) located at 472.402 to 472.403 Mbp within this QTL (Table 4.7).

Table 4.7 Differentially expressed genes (≥ 10-fold, P ≤ 0.05) genes located in *QPhs.umb-1D* and *QPhs.umb-6D*

			Physical position (bp)		D_12/ND_12		D_24/ND_24	
IWGSC gene ID	Gene annotation[1]	Abbreviated gene name	Start	End	Fold change	Significant	Fold change	Significant
QPhs.umb-1D								
TraesCS1D02G403200	CYTOCHROME P450, FAMILY 71, SUBFAMILY B, POLYPEPTIDE 11	*TaCYP71B11*	467852770	467854500	2.28	UP	16.88	UP
TraesCS1D02G400300	DELAY OF GERMINATION 1 LIKE 2	*TaDOG1L-2*	466493735	466494723	2.13	FALSE	16.00	UP
TraesCS1D02G414700	IRON-SULFUR CLUSTER PROTEIN 5	TaISC5	474443065	474444523	1.44	FALSE	11.01	UP
TraesCS1D02G395300	SULFOTRANSFERASE 202B1	*TaSULT202B1*	464028942	464030155	-10.66	FALSE	-14.30	DOWN
TraesCS1D02G413600	SUGAR TRANSPORTER ERD6-LIKE 4	TaERDL4	474010191	474014673	-2.95	FALSE	-17.30	DOWN
TraesCS1D02G412000	ALPHA-AMYLASE-LIKE 2	*TaAMY2*	473524229	473525023	-9.54	DOWN	-20.31	DOWN
TraesCS1D02G396600	HISTONE H2B.4	*TaH2B.4*	464801853	464802597	-1.59	FALSE	-22.63	DOWN
TraesCS1D02G410700	Uncharacterized	-	472402084	472403142	-1.71	FALSE	-38.51	DOWN
QPhs.umb-6D								
TraesCS6D02G197600	LATE EMBRYOGENESIS ABUNDANT PROTEIN 4	*TaLEA4*	275664522	275665153	10.78	UP	27.78	UP
TraesCS6D02G241200	OXIDATIVE STRESS 3	*TaOXS3*	343557954	343558810	1.20	FALSE	16.54	UP
TraesCS6D02G224300	GLYCEROL-3-PHOSPHATE TRANSPORTER 1	*TaGLPT1*	316229057	316232206	2.38	UP	14.44	UP
TraesCS6D02G203800	TEMPERATURE STRESS-INDUCED LIPOCALIN	*TaTIL*	289023546	289024850	2.60	UP	14.20	UP
TraesCS6D02G211600	PHOSPHATE TRANSPORTER 4	*TaPT4*	298924441	298926287	1.55	FALSE	10.84	UP
TraesCS6D02G197500	BASIC HELIX-LOOP-HELIX 129	*TaBHLH129*	275152959	275163740	-1.84	FALSE	-10.93	DOWN
TraesCS6D02G222300	ALDEHYDE DEHYDROGENASE FAMILY 3 MEMBER E1	*TaALDH3E1*	313161968	313167297	-6.72	FALSE	-12.54	DOWN
TraesCS6D02G199500	CHITINASE 6	*TaCHI6*	277266155	277267379	-1.41	FALSE	-12.86	DOWN
TraesCS6D02G248800	OLIGOPEPTIDE TRANSPORTER 4	*TaOPT4*	352351278	352356843	-2.59	FALSE	-13.31	DOWN
TraesCS6D02G200400	SHORT-CHAIN DEHYDROGENASE/REDUCTASE 1	*TaSDR1*	280902346	280906489	-24.76	FALSE	-14.34	DOWN
TraesCS6D02G222400	PROLINE-RICH RECEPTOR-LIKE PROTEIN KINASE 4	*TaPERK4*	313186878	313188491	-9.85	FALSE	-15.40	DOWN
TraesCS6D02G226900	NUCLEOBASE CATION SYMPORTER 1	*TaNCS1*	318349000	318350940	-2.00	FALSE	-28.69	DOWN
TraesCS6D02G235300	RECEPTOR-LIKE CYTOPLASMIC KINASE 82	*TaRLCK82*	330156830	330158989	N.D.	N.D.	-29.89	DOWN
TraesCS6D02G207900	MINICHROMOSOME INSTABILITY 12 (MIS12)-LIKE	*TaMIS12L*	293314706	293318538	2.40	FALSE	-31.48	DOWN
TraesCS6D02G212900	AQUAPORIN 7	*TaAQP7*	302037059	302040674	-5.36	FALSE	-59.83	DOWN
TraesCS6D02G241500	HISTONE H4	*TaHIS4*	343676509	343676820	-1.65	FALSE	-80.82	DOWN

[1]Annotation based on PANTHER16.0, RAPDB and TAIR databases; N.D., not determined due to absence of read in one of two samples; FALSE, insignificant; bp, base pair;
No significant differential expression was observed between dry dormant and non-dormant seeds (D_0/ND_12)

4.4.10. Expression patterns of previously identified genes of PHS QTLs

The expression patterns of genes previously identified as causal and candidate genes for PHS QTLs including *TaMKK3-A, TaPHS1, TaMFT, TaVP1, TaDOG1, TaSdr4* and *TaMYB10* were examined in our samples (Table 4.8). To this effect, IWGSC gene IDs that showed higher homology with nucleotide sequences of these genes (E-value < $1.0E^{-10}$, sequence identity > 75%) were considered. Of these genes, three homologs of *TaVP1* (an ortholog of *TaABI3*) showed over 2-fold upregulation in 24 HAI dormant seeds while two homologs of *TaSdr4* showed upregulation (over 2-fold) in dormant seeds at both 12 and 24 HAI (Table 4.8). Of the three homologs of *TaABI3*, the one represented by Gene ID of TraesCS3B02G452200 exhibited the highest upregulation (3-fold) in dormant seeds at 24 HAI. Between the two homologs of *TaSdr4*, the homolog with Gene ID of TraesCS2B02G215300 showed upregulation in dormant seeds at both 12 and 24 HAI (3.5-fold and 32-fold, respectively). A homolog of *TaMYB10* (Gene ID: TraesCS4B02G323300) was the only gene that showed downregulation (20-fold) in 24 h imbibed dormant seeds (Table 4.8).

Table 4.8 Expression patterns of previously reported PHS resistance genes in our samples

Gene name	Gene abbreviation[1]	Orthologues	IWGSC gene ID	E-value[2]	%Identity	Physical position (bp)		D_12/ND_12		D_24/ND_24	
						Start	End	Fold change	Significant	Fold change	Significant
MITOGEN-ACTIVATED PROTEIN KINASE KINASE 3	*TaMKK3-A*	Os06g0473200	TraesCS5B02G565100	0.00	92	710226104	710234840	-1.02	FALSE	-1.02	FALSE
			TraesCS5D02G549600	0.00	89.9	556549608	556556744	-1.16	FALSE	1.01	FALSE
PROPYZAMIDE-HYPERSENSITIVE 1	*TaPHS1*	Os01g0311500	TraesCS1D02G265200	5.90E-177	90.3	361109557	361115773	-1.01	FALSE	-1.06	FALSE
			TraesCS1A02G265000	1.50E-176	90	460669777	460676972	1.25	FALSE	-1.29	FALSE
			TraesCS1B02G275700	1.20E-173	89	483367067	483372856	-1.10	FALSE	-1.76	DOWN
MOTHER OF FT AND TFL1	*TaMFT*	Os01g0111600	TraesCS3B02G007400	2.30E-61	76.4	3863692	3868727	5.21	FALSE	8.20	FALSE
VIVIPAROUS 1	*TaVP1*	Os01g0911700	TraesCS3D02G412800	1.40E-64	81.9	525469559	525473995	1.00	FALSE	2.72	UP
			TraesCS3B02G452200	1.90E-64	91.1	693338347	693342798	1.13	FALSE	3.05	UP
			TraesCS3A02G417300	1.70E-63	90.2	659552136	659556535	1.02	FALSE	2.89	UP
DELAY OF GERMINATION1	*TaDOG1*	Os01g0159000	TraesCS3D02G105800	1.40E-49	77.5	58107658	58108963	1.06	FALSE	1.27	FALSE
			TraesCS3A02G103500	5.10E-49	77.5	67123770	67124901	-1.07	FALSE	1.63	UP
			TraesCS3B02G120900	1.20E-47	75.8	91147416	91148571	-1.19	FALSE	-1.86	DOWN
			TraesCS2A02G010800	8.80E-34	78.9	4113715	4114311	N.D	N.D.	N.D.	N.D.
SEED DORMANCY 4	*TaSdr4*	Os07g0585700	TraesCS2B02G215300	1.00E-14	76.2	200572593	200573860	3.25	UP	31.86	UP
			TraesCS2D02G196200	1.00E-14	76.2	142666999	142668370	1.28	FALSE	-1.32	FALSE
			TraesCS2A02G191400	1.60E-11	93.3	158452365	158453649	2.97	FALSE	57.09	UP
MYB TRANSCRIPTION FACTOR 10	*TaMYB10*	AT5G35550	TraesCS3D02G468400	3.80E-39	87.3	570801163	570803376	N.D.	N.D.	N.D.	N.D.
			TraesCS3B02G515900	3.80E-39	87.3	757918264	757920082	N.D.	N.D.	N.D.	N.D.
			TraesCS4D02G319700	3.50E-38	84.7	483369177	483370210	N.D.	N.D.	N.D.	N.D.
			TraesCS5A02G495500	4.90E-38	84.5	663379037	663379971	N.D.	N.D.	N.D.	N.D.
			TraesCS4D02G319800	4.90E-38	84.5	483430537	483431661	N.D.	N.D.	N.D.	N.D.
			TraesCS7B02G070400	9.20E-38	83.1	77430167	77432792	N.D.	N.D.	N.D.	N.D.
			TraesCS7D02G166500	9.20E-38	83.1	117242813	117245836	N.D.	N.D.	N.D.	N.D.
			TraesCS7A02G165700	9.20E-38	83.1	121397856	121399015	N.D.	N.D.	N.D.	N.D.
			TraesCS4B02G323300	1.30E-37	83.1	613753689	613754765	N.D.	N.D.	-19.94	DOWN
			TraesCS4D02G224500	6.20E-37	80.6	381403684	381404442	N.D.	N.D.	N.D.	N.D.
			TraesCS3D02G257800	3.10E-36	80.6	360047967	360049197	1.50	FALSE	1.02	FALSE
			TraesCS3B02G290600	3.10E-36	80.6	466105785	466106991	-1.71	FALSE	2.30	FALSE
			TraesCS3A02G257800	3.10E-36	80.6	479481051	479482052	-1.79	FALSE	1.02	FALSE
			TraesCS5B02G330100	8.00E-36	76.4	514278252	514279918	-2.18	FALSE	-1.79	FALSE
			TraesCS5A02G329900	8.00E-36	76.4	539092242	539093667	-1.26	FALSE	1.00	FALSE
			TraesCS5D02G335700	8.00E-36	76.4	425076252	425077761	1.04	FALSE	1.11	FALSE

[1]Annotation based on PANTHER16.0, RAPDB and TAIR databases; [2]E-value: < $1.0E^{-10}$, %ID > 75; N.D., not determined due to absence of read in one of two samples; FALSE, insignificant; bp, base pair ; No significant expression was observed in dry dormant seed (D_0/ND_12).

4.4.11. Discovery of SNPs and InDels

Comparison between the mRNA/cDNA sequences from the dormant and non-dormant seed samples identified a total of 120,675 SNPs and 11,983 InDels (Appendix 14). The SNPs and InDels were not evenly distributed across the sub-genomes and chromosomes of the common wheat. The highest number of variants was identified on the B sub-genome followed by the A sub-genome, and then the D sub-genome. Consistently, the highest number of SNPs and InDels was detected on chromosome 2B (11,254 SNPs and 1,095 InDels) followed by chromosome 6B (10,784 SNPs and 961 InDels) (Appendix 14). The lowest number of SNPs and InDels was found on chromosome 4D (1,289 SNPs and 192 InDels). A total of 2,688 SNPs and 125 InDels were detected in sequences not assigned to any chromosome.

Table 4.9 SNPs in DEGs located at 1D, 4A, 6B and 6D QTLs along with their physical positions

Position (bp)	REF[1]	ALT[2]	D_24[3]	ND_24	IWGSC gene ID	Fold change	Significant
				QPhs.umb-1D			
469037464	T	C	0/0:3	1/1:4	TraesCS1D02G404200	3.32	UP
473999073	T	C	0/1:35	0/0:42	TraesCS1D02G413500	1.19	FALSE
				QPhs.umb-4A			
604102617	G	A	0/0:17	1/1:6	TraesCS4A02G313600	2.34	UP
605645857	A	T	0/0:78	1/1:99	TraesCS4A02G315200	1.50	FALSE
				QPhs.umb-6B			
446842663	C	T	1/1:7	0/0:5	TraesCS6B02G249100	-3.29	DOWN
461264948	A	C	0/0:6	1/1:4	TraesCS6B02G256600	2.22	UP
461264949	T	A	0/0:6	1/1:4	TraesCS6B02G256600	2.22	UP
461267887	G	A	1/1:7	0/0:5	TraesCS6B02G256700	-2.34	DOWN
461269769	A	T	1/1:4	0/0:3	TraesCS6B02G256700	-2.34	DOWN
461269770	T	A	1/1:4	0/0:3	TraesCS6B02G256700	-2.34	DOWN
462129188	C	T	1/1:7	0/1:3	TraesCS6B02G257600	3.10	UP
462129266	C	A	1/1:4	0/0:3	TraesCS6B02G257600	3.10	UP
462132461	T	C	1/1:38	0/1:4	TraesCS6B02G257600	3.10	UP
462133028	A	G	0/0:40	1/1:8	TraesCS6B02G257600	3.10	UP
471628650	A	G	0/0:3	0/1:35	TraesCS6B02G261900	-6.02	DOWN
487832727	A	T	0/1:3	1/1:7	TraesCS6B02G271000	-2.20	DOWN
515066066	T	G	1/1:7	0/1:3	TraesCS6B02G285900	8.23	UP
516522583	T	G	0/0:5	1/1:7	TraesCS6B02G286800	-2.49	DOWN
528085861	G	A	1/1:99	0/0:48	TraesCS6B02G293900	2.72	UP
528085892	C	T	1/1:99	0/0:51	TraesCS6B02G293900	2.72	UP
528086663	T	C	1/1:51	0/0:9	TraesCS6B02G293900	2.72	UP
528087159	C	T	1/1:99	0/0:99	TraesCS6B02G293900	2.72	UP
528087550	T	C	1/1:99	0/0:54	TraesCS6B02G293900	2.72	UP
528088293	T	G	1/1:36	0/0:15	TraesCS6B02G293900	2.72	UP
528089587	G	T	1/1:6	0/0:8	TraesCS6B02G293900	2.72	UP
541160650	C	T	0/1:42	0/0:99	TraesCS6B02G302000	-3.36	DOWN
556487266	T	A	1/1:6	0/0:8	TraesCS6B02G310600	-2.22	DOWN
556487324	T	A	1/1:7	0/0:5	TraesCS6B02G310600	-2.22	DOWN
560625327	A	G	0/0:93	1/1:27	TraesCS6B02G312900	2.53	UP
				QPhs.umb-6D			
276144093	T	C	0/1:35	0/0:11	TraesCS6D02G198300	-2.28	DOWN
323712394	C	T	1/1:4	0/0:3	TraesCS6D02G230500	-4.59	DOWN
340946686	T	A	0/1:53	0/0:5	TraesCS6D02G239800	2.14	UP

[1]REF, reference genome; [2]ATL, alternate genome (D or ND); [3]0/0: homozygous to REF; 0/1: heterozygous, heterozygous of REF and ALT; 1/1: homozygous to ALT; The number after ratio represents number of reads supporting that SNP; FALSE, insignificant.

The SNPs and InDels identified were further examined in order to determine the associated genes and their density in the QTLs detected by our previous genetic study. Our analysis detected 53 SNPs and 6 InDels in 1D QTL, 47 SNPs and 10 InDels in 4A QTL, 2237 SNPs and 190 InDels in 6B QTL, and 236 SNPs and 46 InDels in 6D QTL. Out of these, 2, 2, 25 and 3 SNPs are located in the genes identified to be located within the 1D, 4A, 6B and 6D QTL regions, respectively (Table 4.9). With respect to the InDels, 2 InDels in 1D, 9 InDels in 6B, and 2 InDels in 6D QTLs were found to correspond to the genes differentially expressed between our samples (Table 4.10). In terms of SNP density, no gene was found to contain more than one SNP in 1D, 4A and 6D QTL, whereas each of five genes in 6B QTL contained 2 or more SNPs in their respective chromosomal locations. The highest SNP density (7 SNPs) in 6B QTL was found in an uncharacterized gene (Gene ID: TraesCS6B02G293900) (Table 4.9) followed by cryptochrome 2 (TaCry2; Gene ID of TraesCS6B02G257600 (4 SNPs)). Regarding density of InDels, same gene (Gene ID: TraesCS6B02G293900) was found to contain the highest number of InDels (3 InDels) (Table 4.10).

The genes that contain SNPs and InDels and showed differential expression between the dormant and non-dormant seeds were further analyzed to determine the specific locations of the SNPs and InDels within the DEGs. Our analysis showed the detection of no SNPs and InDels in the protein coding sequences (CDS) of the corresponding genes, rather the SNPs and InDels are located in the untranslated regions (UTRs) of the genes. For instance, 100% homology was observed between the CDS of an uncharacterized gene (Gene ID: TraesCS6B02G293900) derived from dormant and non-dormant seeds (Appendix 15), however, the two genes were identified as heterozygous to each other because of the presence of SNPs (Table 4.9).

Table 4.10 InDels in DEGs located at 1D, 6B and 6D QTLs along with their physical positions

Position (bp)	REF	ALT	D_24	ND_24	IWGSC Gene Id	Fold change	Significant
		QPhs.umb-1D					
464820000	GTATATA	GTATA	0/0:21	1/1:3	TraesCS1D02G396700	5.13	UP
475728823	CTG	CTGTG	0/0:99	0/1:37	TraesCS1D02G417600	1.25	FALSE
		QPhs.umb-6B					
459804349	TAA	TAAA	1/1:6	0/0:99	TraesCS6B02G255800	-2.01	DOWN
459809789	ATTTTTTTT	ATTTTTTTTT	1/1:61	0/0:99	TraesCS6B02G255800	-2.01	DOWN
493324682	CAAAAAAAAAA	CAAAAAAAAAAAAAAAA,CAAAAAAAAAAAAA	1/1:4	0/1:3	TraesCS6B02G273200	-2.27	DOWN
528086686	ATT	AT	1/1:25	0/0:6	TraesCS6B02G293900	2.72	UP
528087401	ATTTTT	ATTTTTT	1/1:99	0/0:63	TraesCS6B02G293900	2.72	UP
528091707	CTTTTTT	CTTTTTTT	1/1:16	0/0:4	TraesCS6B02G293900	2.72	UP
561700129	TAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA	TAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG,TAGAGAGAGAGAGAGAGAGAGAGAGAGAGA GAGAGAG	1/1:4	0/1:3	TraesCS6B02G313900	-2.79	DOWN
573328174	CGGGG	CGGGGG	0/0:7	1/1:9	TraesCS6B02G324000	-2.55	DOWN
581868391	GAAAAAAA	GAAAAAAAA	0/0:18	1/1:21	TraesCS6B02G331000	-2.06	DOWN
		QPhs.umb-6D					
256316243	GGAGAGAGAGAGAGAGAGAG	GGAGAGAGAGAGAGAGAG	0/0:11	0/1:95	TraesCS6D02G188300	-2.34	DOWN

Data descriptions are as shown in Table 9

4.4.12. Analysis of transcription factors

A total of 3,030 transcription factors (TFs) and 773 transcriptional regulators (TRs) were identified as differentially expressed between the dormant and non-dormant samples, which are classified into 56 TF and 23 TR families (Appendix 16). The TF families with the highest number of DEGs include the AP2-EREBP family (250 DEGs) followed by the MYB (224 DEGs) and bHLH (204 DEGs) families. With respect to the TR families, the orphan family was found to have the highest number of DEGs (107 DEGs) followed by the PHD family (102 DEGs).

A total of 14 genes that belong to the AP2/EREBP, HSF, C/EBP and NAC families were found to exhibit high levels of differential expressions between our dormant and non-dormant seed samples (≥80-fold and P≤0.05). Of these, all the AP2/EREBP and HSF gene family members were upregulated in imbibed dormant seeds except a member of the AP2/EREBP family, *TaERF061*, which was downregulated in the same seed sample. All the C/EBP and NAC gene family members were, however, downregulated in the dormant seeds (Table 4.11). Among those upregulated in dormant seeds, *dehydration-response element-binding protein 2C-B* (*TaDREB2C-B*), a gene that belongs to the AP2/EREBP family and located at 451.134 to 451.135 Mbp on chromosome 1B, showed upregulation (187-fold) only at 24 HAI while *heat stress transcription factor C2A* (*TaHSFC2A*), a gene that belongs to the HSF family and located at 664.759 to 664.760 Mbp on chromosome 4B, showed upregulation (115-fold) at both 12 and 24 HAI (Table 12). In contrast, *histone H2A.2.1-A,* a gene that belongs to the C/EBP family and located at 20998069 to 20998918 bp on chromosome 1A, showed the highest downregulation (287-fold) in the 24 HAI dormant seed samples followed by *ethylene-responsive transcription factor 061* (*TaERF061*; 241-fold downregulation) (Table 4.11).

Gene regulatory network analysis of *TaDREB2C-B*, a TF gene that showed the highest upregulation in the dormant seeds, indicate that it regulates the expression of genes encoding abscisic acid responsive elements-binding factors including ABF2, ABF3 and ABI5-like protein in the ABA signaling pathway (Appendix 17). Consistently, *TaABI5* was found to be upregulated in the dormant seeds.

Table 4.11 Expression patterns of genes encoding transcription factors di(≥ 80-fold, P ≤ 0.05)

				Physical position (bp)		D_12/ND_12		D_24/ND_24	
IWGSC gene ID	Gene annotation	Gene abbreviation	TF Family	Start	End	Fold change	Significant	Fold change	Significant
TraesCS1B02G256000	DEHYDRATION-RESPONSIVE ELEMENT-BINDING PROTEIN 2C-B	*TaDREB2C-B*	AP2-EREBP	451134276	451135517	5.99	FALSE	186.68	UP
TraesCS1A02G244800	DEHYDRATION-RESPONSIVE ELEMENT-BINDING PROTEIN 2C-A	*TaDREB2C-A*	AP2-EREBP	432852085	432853063	21.26	FALSE	152.59	UP
TraesCS4B02G385800	HEAT STRESS TRANSCRIPTION FACTOR C2A	*TaHSFC2A*	HSF	664759151	664760300	13.87	UP	114.83	UP
TraesCS5A02G462100	HISTONE H2A.1-5A	-	C/EBP	642918526	642919457	-1.98	FALSE	-83.11	DOWN
TraesCS6B02G048600	HISTONE H2A.1-6B.1	-	C/EBP	28516616	28517411	-1.53	FALSE	-83.56	DOWN
TraesCS6B02G082500	HISTONE H2A.1-6B.2	-	C/EBP	59294863	59295674	N.D.	N.D	-85.33	DOWN
TraesCS6D02G040400	HISTONE H2A.1-6D	-	C/EBP	16502832	16503645	-1.44	FALSE	-88.51	DOWN
TraesCS3A02G102400	HISTONE H2A-3A	-	C/EBP	66249912	66250726	1.06	FALSE	-91.83	DOWN
TraesCS6B02G237900	NAC DOMAIN-CONTAINING PROTEIN 85	*TaNAC85*	NAC	413940233	413946455	N.D.	N.D.	-93.61	DOWN
TraesCS6A02G034900	HISTONE H2A-6A	-	C/EBP	16700898	16701729	N.D.	N.D.	-117.30	DOWN
TraesCS5D02G474800	HISTONE H2A.1-5D	-	C/EBP	514923266	514924170	N.D.	N.D.	-122.22	DOWN
TraesCS1D02G039800	HISTONE H2A.2.1-D	-	C/EBP	19150972	19151814	3.74	FALSE	-131.14	DOWN
TraesCS7B02G085900	ETHYLENE-RESPONSIVE TRANSCRIPTION FACTOR 061	*TaERF061*	AP2-EREBP	99088470	99090547	N.D.	N.D	-240.99	DOWN
TraesCS1A02G038100	HISTONE H2A.2.1-A	-	C/EBP	20998069	20998918	N.D.	N.D.	-286.46	DOWN

[1]Annotation based on PANTHER16.0, RAPDB and TAIR databases; TF, transcription factor; N.D., not determined due to absence of read in one of two samples; FALSE, insignificant; bp, base pair. No significant differential expression was observed between dry dormant and non-dormant seeds (D_0/ND_0);

4.4.13. Validation of differentially expressed genes using RT-qPCR

The RNA-seq data was validated by RT-qPCR analysis of selected DEGs related to ABA, GA and auxin metabolism and signaling, and cell expansion and cycle. Since the RT-qPCR primers were designed to amplify the three homeologues of each gene, the relative read count was calculated using the total read count of the three homeologues and that of the *TaActin*. The relative expression levels of the selected genes based on the RT-qPCR data exhibited high degrees of positive correlation ($R \geq 0.95$) with their relative read counts from the RNA-seq data (Fig. 4.14).

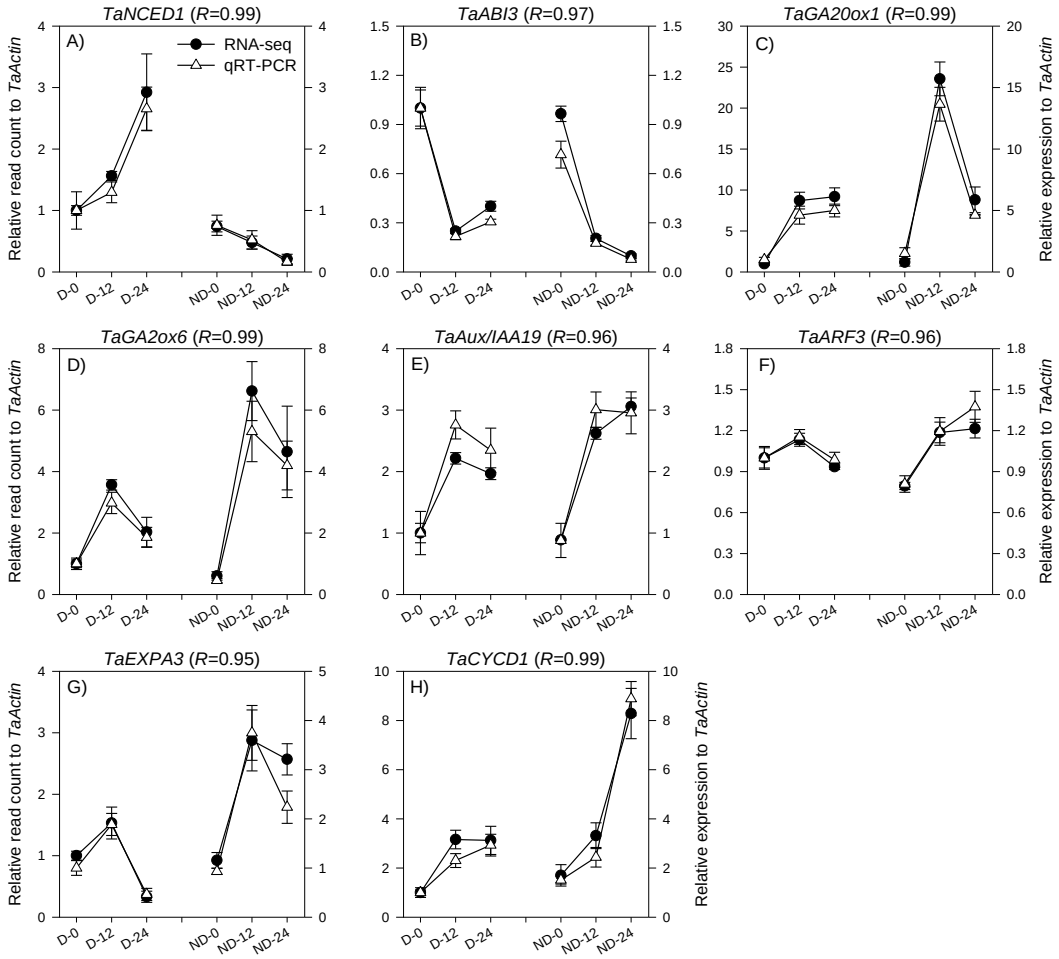


Fig. 4.14 Validation of the RNA-seq data using qPCR. The RNA-seq and qPCR-based expression data were compared for selected differentially expressed genes related to abscisic acid (ABA, a and b), gibberellin (GA, c and d) and auxin (e and f) metabolism and signaling, and those related to cell expansion and cycle (g and h). Curves with solid circles represent relative read count from RNA-seq data (left y-axis) while curves with open circles represent expression data based on qPCR (right y-axis). The read count and expression level of each gene were determined using *TaActin* as a reference gene and expressed relative to their respective levels detected in the dry dormant (D) seeds (D_0), which was set to a value of 1. D_0, D_12, and D_24 indicate dormant seeds imbibed for 0, 12, and 24 h; ND_0, ND_12, and ND_24 indicate non-dormant seeds imbibed for 0, 12, and 24 h. *NCED1*, *9-cis-epoxycarotenoid dioxygenase 1*; *ABI3*, *ABA insensitive 3*; *GA20ox1*, *GA 20-oxidase 1*; *GA2ox6*, *GA 2-oxidase 6*; *Aux/IAA19*, *auxin/indole-3-acetic acid 19*; *ARF3*, *auxin response factor 3*; *EXPA3*, *α-expansin 3*; *CYCD1*, *cyclin D1*.

4.5. Discussion

The present study undertaken transcriptomic analysis of seed dormancy in wheat to investigate the molecular mechanisms and to identify candidate genes regulating PHS resistance in RL4137. Analysis of the RNA-seq data revealed differential expression of 48,398 genes between dry and imbibed seeds of the dormant and non-dormant DH lines. Such a large number of DEGs between dry and imbibed dormant and non-dormant seeds has been reported in previous studies that involved large scale/genome-wide transcriptomic analysis in the model plant Arabidopsis (Rajjou et al. 2012), wheat (Gao et al. 2012), rice (Park et al. 2021) and barley (An and Lin 2011). Enrichment of the DEGs in different GO terms including cell, cell part, intracellular, integral component of membrane under the cellular component domain; response to stimulus and response to abiotic stimulus under the biological process domain, and protein dimerization activity and nutrient reservoir activity under the molecular function domain is consistent with previous reports (Wang et al. 2021; Gao et al. 2012), implying the roles of these GO terms in the regulation of seed dormancy in wheat. Our pathway analysis indicated enrichment of the DEGs in the carotenoid, diterpenoid and flavonoids biosynthesis pathways, and pathways involving signal transduction of plant hormones, mainly ABA, auxin and ET (Tables 4.2, 4.3, 4.4). The importance of hormone signal transduction and secondary metabolite biosynthetic pathways in regulation of seed dormancy and germination has been reported in wheat (Lee et al. 2021; Wang et al. 2021; Rikiishi et al. 2021; Tuan et al. 2021) and other species such as quinoa (Wu et al. 2020).

The plant hormone ABA, which is a known inducer of seed dormancy (Nambara et al. 2010; Tuan et al. 2018), is biosynthesized from carotenoids via several enzymatic reactions and the reaction catalyzed by the NCED is recognized as a rate limiting step (Schwartz et al. 2003; Nambara and Marion-Poll 2005). Thus, the upregulation of *TaNCED3* and *TaNCED6* in imbibed dormant seeds (Table 4.4) indicate the role of specific *NCED* genes in regulating ABA level, and dormancy induction and maintenance as reported previously (Zhang et al. 2014a; Son et al. 2016). It has recently been reported that the cutin layer, a waxy film located on the outer layer of endosperm and embryo tissues, enhances seed dormancy through regulating gaseous exchange between the seed and its outer environment and specific *NCED* genes positively regulate the biosynthesis/deposition of cutin in seed (Nonogaki 2019). The upregulation of the *TaNCED3* and *TaNCED6* genes in the dormant seeds may imply their importance in the regulation of cutin

deposition in the seed tissues and therefore dormancy. Gibberellin is the other major phytohormone that regulates seed dormancy and germination; high level of GA promotes germination whereas low level GA contributes to induction of dormancy (Bentsink and Koornneef 2008; Tuan et al. 2018). The observation of downregulation of the GA biosynthesis gene *TaGA20ox1* and upregulation of the GA catabolism gene *TaGA2ox3* in imbibed dormant seeds imply the role of these genes in limiting GA level and therefore in maintenance of dormancy (Liu et al. 2013a; Tuan et al. 2018, 2021). Interestingly, *TaGA20ox1* is located within 4A QTL, which is identified by our genetic study as a major QTL regulating PHS resistance (Liton et al. 2021). Consistently, another previous genetic study on PHS resistance in wheat has detected *TaGA20ox1* in the 4A QTL region (Cabral et al. 2014).

It has been shown previously that accumulations of catechin and proanthocyanidins promote seed dormancy through increasing testa thickness and mechanical strength (Debeaujon et al. 2007; Ali et al. 2019). Proanthocyanidins are synthesized through the flavonoid biosynthesis pathway, which is mediated by several enzymes including anthocyanidin synthase (ANS) and BANYULS (BAN) that catalyze the conversion of leucopelargonin to pelargonin and Flavan-3-ol to proanthocyanidin, respectively (Ali et al. 2019). Therefore, the upregulation of *TaBAN* and *TaANS* in imbibed dormant seeds might indicate the contribution of proanthocyanidins in enhancing dormancy (Debeaujon et al. 2007; Gianinetti et al. 2018; Ali et al. 2019).

ABA signaling plays important role in regulating seed ABA sensitivity and therefore dormancy (Nambara et al. 2010). Our analysis showed downregulation of the ABA receptor *PYL* genes but upregulation of *PP2C*s in imbibed dormant seed (Table 4.2). Although the expression of these genes has been reported to be associated with seed dormancy in wheat (Izydorczyk et al. 2018), there was no such apparent correlation between their expression patterns and dormancy phenotype in our samples, suggesting that these genes are subjected to post-transcriptional regulation in our samples. Previous studies in Arabidopsis and quinoa have also reported such a lack of correlation between the expression patterns of specific *PYL*s and *PP2C*s and seed ABA-sensitivity and dormancy (Maia et al. 2014; Wu et al. 2020). The *SnRK2* acts as positive regulators of ABA signaling and activates ABA-responsive element binding (AREB) proteins/AREB binding factors (ABF) that regulates ABA responsive genes (Ng et al. 2014), and therefore dormancy (Finkelstein et al. 2008; Kulik et al. 2011). In agreement with this,

upregulation of the *SnRK2* genes was observed in imbibed dormant seeds, indicating their involvement in maintaining seed dormancy in wheat. Among the AREB/ABF transcription factors, *ABI3*, *ABI4* and *ABI5* are crucial for transcriptional activation of ABA responsive genes and dormancy (Tuan et al. 2018; Ali et al. 2022). Thus, the upregulation of the *TaABI3/TaVP1* and *TaABI5* gene in imbibed dormant seeds, indicating their important roles in seed ABA sensitivity and dormancy as reported previously in wheat and barley (Mendiondo et al. 2010; Izydorczyk et al. 2018; Wang et al. 2021).

The plant hormone auxin also plays an important role in seed dormancy via increasing seed ABA sensitivity (Brady et al. 2003; Shu et al. 2016). For example, auxin controls the expression of *ABI3* through *ARF* (Liu et al. 2013c), and the activation of *ARF* and its role in enhancing seed dormancy has been shown to require repression of *Aux*/*IAA* by *IAA7*/*AXR2* and *IAA17*/*AXR3* (Leyser 2018; Hussain et al. 2020). However, the ARFs did not show any differential expression (at 3-fold cut-off) between the dormant and non-dormant seeds while the *TaAux*/*IAA*s exhibited downregulation in the dormant seeds except that *TaAux*/*IAA9* showed upregulation (Table 4.3). Given that, auxin signaling repressor *IAA8* showed higher expression in germinating seeds of Arabidopsis and promote seed germination through downregulating expression of *ABI3* in same sample (Hussain et al. 2020). This result indicates that the downregulation of *Aux*/*IAA*s including *TaIAA13, TaIAA20* and *TaIAA27* promotes seed dormancy in wheat while *TaIAA9* may regulate post-transcriptionally.

Previous studies have also shown the role of ET in seed dormancy and germination (KeÇpczyński and KeÇpczyńska 1997; Corbineau et al. 2014). The binding of ethylene to its receptor ETR1 inactivates the CTR1, which is a negative regulator of ET signaling, leading to transcriptional activation of the downstream ET signaling transcription factors such as *EIN2* and *EIN3*, and the ET responsive genes (Rzewuski and Sauter 2008; Stepanova and Alonso 2009; Yoo et al. 2009) resulting promotion of germination (Wang et al. 2002; Merchante et al. 2013; Corbineau et al. 2014). Our analysis, however, showed upregulation of *TaETR4* in imbibed dormant seeds, suggesting transcriptional inactivation of *CTR1* and therefore activation of that of the *EIN*s. In agreement with this, upregulation of *TaEIN2*s was observed in our imbibed dormant seed. Given that ET regulates dormancy negatively; our data might suggest that the *ETR* and *EIN*s genes in our seed samples are not regulated at transcriptional level. Previous studies have also reported upregulation of ET signaling genes such as *ethylene responsive factor 1* (*ERF1*),

ETR1, *ethylene response sensor 1* (*ERS1*), *EIN2* and *EIN3* in non-dormant/after-ripened seeds of wheat and other species (Chitnis et al. 2014; Corbineau et al. 2014; Zhang et al. 2019).

In order to identify more candidate genes for PHS resistance, the DEGs were also mapped to QTLs (4A, 6B, 1D and 6D QTLs) detected previously (Liton et al. 2021). Our analysis showed the mapping of 25 DEGs to 4A QTL and these genes include those upregulated (≥2-fold) in dormant seeds such as *TaPM19-A1*, *TaPM19-A2* and *11-beta-hydroxysteroid dehydrogenase-like 4A* (*TaHSD4A*), and those downregulated in same sample such as two genes encoding uncharacterized proteins and *TaGnk2RLK-7* (Table 4.5). Both the *TaPM19-A1* and *TaPM19-A2* genes, which show higher expression level in dormant seeds, have been reported as candidate PHS resistance genes (Barrero et al. 2015). Moreover, a study in Arabidopsis has shown *PM19* as regulons of *ABI3* and *ABI5* transcription factors that regulate ABA signaling and dormancy (Barrero et al. 2019). However, *TaMKK3-A*, which has been identified as a major causal gene of 4A QTL in wheat (Torada et al. 2016), didn't show differential expression between dormant and non-dormant seeds in our study.

A total of 24 DEGs (≥10-fold) were mapped to 6B QTL (Table 4.6); of which, a gene encoding an uncharacterized protein was highly upregulated in imbibed dormant seeds and *TaRH8* showed the highest downregulation in the same seed samples. RNA helicases, which belong to the large DEAD box protein family, play important roles in RNA metabolism and mRNA translation (Mingam et al. 2004) by disrupting misfolded mRNA structures and thus promoting proper mRNA folding (Lorsch 2002). Since translation of specific subsets of seed stored mRNAs is important for the release of seed dormancy and germination (Bazin et al. 2011; Meimoun et al. 2014), the downregulation *TaRH8* in our dormant seeds might imply its importance in restricting the translation of mRNAs required for germination and therefore maintaining dormancy.

Among the 8 DEGs (≥10-fold) located within 1D QTL, a gene encoding a member of the cytochrome P450 superfamily (*TaCYP71B11*) showed the highest upregulation (x-fold) in imbibed dormant seeds, suggesting it as a candidate gene influencing seed dormancy (Table 4.7). In agreement with this, *CYP71B11* of Arabidopsis has been shown to be regulated by ABA in which its expression in seeds is upregulated by ABA, which induces seed dormancy (Kim et al. 2019). In agreement with this, upregulation of *TaCYP71B11* in our dormant seed samples was shown to be associated with upregulation of the ABA biosynthetic *TaNCED* genes. A gene

encoding an uncharacterized protein (Gene ID: TraesCS1D02G410700) that showed the highest downregulation in dormant seeds is among the DEGs located in the same 1D QTL regulating PHS, indicating its potential role in influencing seed dormancy. Interestingly, the homolog of one of the most important seed dormancy genes, *TaDOG1L-2*, which showed 16-fold upregulation in imbibed dormant seed, was also located on 1D QTL, suggesting its importance in regulation of seed dormancy in our samples (Nishimura et al. 2018; Nonogaki 2019). No study in wheat has reported *DOG1* as a candidate/causal gene of seed dormancy/PHS QTL.

The 6D QTL was found to harbour 16 DEGs (≥10-fold) such as *TaLEA4*, which is highly upregulated in imbibed dormant seeds (Table 4.7). Since ectopic expression of the *LEA* gene of alfalfa in Arabidopsis induces the expression level of ABA metabolism and signaling genes and delayed germination (Jia et al. 2020), the upregulation of *TaLEA4* in dormant seeds implies its function in the maintenance of dormancy. Genes highly downregulated in dormant seeds such as *TaHIS4* was found to be located in this QTL, indicating the involvement of epigenetic mechanisms in the regulation of seed dormancy as reported previously (He et al. 2021). In addition to the genes described above, our analysis revealed the localization of many DEGs including those either characterized previously as regulators of dormancy/PHS resistance or encoding uncharacterized proteins within the QTLs identified by our genetic study. Further studies are needed to determine the function of the uncharacterized genes with respect to dormancy/PHS resistance.

SNP markers are used to identify QTLs and genes controlling a trait of interest in crops. Thus, genome-wide distribution of SNPs and SNP density are important parameters of consideration for identifying the target genes (Tanaka et al. 2019). Analysis of the distribution and density of SNPs in our previously identified QTLs revealed the presence of a total of 2 to 25 SNPs in DEGs located within each of the four QTLs (Table 4.9). The highest number of SNPs was located on 6B QTL, indicating their predominant distribution on the B genome. Our previous parental allelic contribution to the four detected QTLs showed that the dormant alleles in all the four QTLs are contributed by the PHS resistance parent RL4137 while the non-dormant alleles are contributed by both RL4137 (QTLs 1D, 6B and 6D) and 'Roblin' (QTL 4A) (Table 4.1) (Liton et al. 2021), indicating that the major portion of the SNPs in the DEGs located within the QTLs were contributed by RL4137. Furthermore, we determined that all the SNPs are located in the non-coding regions of the associated genes, and such a finding has been reported

previously (Wang et al. 2021; Nakamura et al. 2011). Previous studies have showed that SNPs located outside of the gene coding region can regulate the expression of a target gene (Ali et al. 2019; Nakamura et al. 2011). For example, a SNP located in the promoter region (-222 bp) of *TaMFT* has been shown to increase the expression the gene and therefore enhance seed dormancy in wheat. Our analysis detected high SNP density in genes located in 6B QTL, and the presence of such a high SNP density in a gene during marker-trait association analysis increases the probability of identifying a target gene regulating the trait of interest (Kujur et al. 2015). Thus, the SNPs associated with the DEGs located within the QTLs might cause the observed variation in dormancy levels, indicating their potential use to develop user friendly DNA marker for marker-assisted breeding.

Of the 3,803 DEGs encoding TFs, a large proportion represents the AP2-EREBP, MYB, bHLH, mTERF and bZIP TF families, which have been reported to play important roles in hormone signal transductions and therefore in controlling seed dormancy (Liu and Zhang 2017; Wu et al. 2020). For instance, gene regulatory network analysis and other studies revealed that *TaDREB2C-B*, a gene encoding a member of the AP2-EREBP family and exhibited the highest upregulation in dormant seeds, functions as a positive regulator of ABA metabolism and signaling genes (Je et al. 2014; Shu et al. 2018; Liu et al. 2020; Sano and Marion-Poll 2021). This is well supported by the upregulation of the *TaNCED* and *TaABI5* genes in our dormant seed samples. In contrast, *histone H2A*s, genes encoding a member of the C/EBP family, exhibited the highest downregulation in dormant seeds, suggesting their role as negative regulators of seed dormancy. Similarly, mutants (*hta9 hta11*) of two *histone H2A* (*HTA*) genes, *hta9* and *hta11*, encode a histone H2A variant histone H2A.Z in Arabidopsis showed delayed germination under normal conditions (Sura et al. 2017). In addition, *histone regulator A* (*HIRA*) has been shown to acts as a positive regulator of seed germination in Arabidopsis (Layat et al. 2021). Understanding the mechanism by which *histone H2A* regulates dormancy in wheat seeds requires further study.

In conclusion, the present study identified a large number of genes as differentially expressed between dormant and non-dormant seeds of wheat. The DEGs were found to be enriched in several gene ontology classes of biological processes including hormone signal transduction, secondary metabolite biosynthesis and general metabolism, suggesting the role of these specific physiological processes in seed dormancy in wheat. Specific genes that have been

shown to regulate seed dormancy positively including *TaNCED6* and *TaNCED3* of ABA biosynthesis and *TaSnRK2.7* and *TaABI5* of ABA signaling showed upregulation in dormant seeds while *TaGA2ox3* of GA catabolism showed downregulation. In addition, some flavonoid biosynthesis genes including *TaBAN* and *TaANS*, exhibited upregulation in dormant seed. Our study also showed upregulation of specific TFs such as *TaDREB2C* in dormant seeds, suggesting their role in regulating the expression of dormancy related genes. In addition, specific genes located within the previously identified seed dormancy QTLs showed differential expression between the dormant and non-dormant seeds, suggesting their potential role in regulation of seed dormancy. The study also identified SNPs that could potentially be used for gene targeting/tagging in the future. Consequently, the findings of this study will provide insight into the molecular mechanisms and some candidate genes of seed dormancy that potentially increase PHS resistance in wheat.

CHAPTER 5

GENERAL DISCUSSION AND CONCLUSION

Pre-harvest sprouting is a major concern in the production of many cereals including wheat, rice, barley, maize and sorghum since it causes significant yield and end-use quality losses and therefore resulting financial loss to the producers. Irregular weather patterns due to climate change may pronounce the effects of PHS in the years to come. Thus, it is highly desirable to improve PHS resistance in wheat through using advanced technologies and updated genomic information. Since PHS is a quantitative trait, in depth understanding of the genetic and molecular regulatory mechanisms of PHS resistance, which have been studied scarcely in wheat, is important. The present study was, therefore, aimed at identifying QTLs, genes and molecular mechanisms of PHS resistance in wheat.

The first research chapter of the thesis examined genetic mechanisms of PHS resistance through identifying some major and minor effect QTLs. This specific study involved the use of a mapping population of 330 DH lines derived from 'Roblin'/RL4137 and the population along with the parents was tested over five different environments. Phenotypic evaluation of the population revealed variations in their SGI value across the environments. Seeds of some genotypes completed their germination before those of the susceptible parent and seeds of other genotypes showed stronger dormancy than those of the resistant parent, indicating high transgressive segregation of population for the studied trait and this facilitated to the precise detection of the QTL location. Our data also showed a highly significant and positive phenotypic correlation among the trial environments, and the higher heritability observed across the trial environments represented a greater contribution of genetic factors in the regulation of PHS resistance. Frequency distribution of the DH lines based on their respective SGI values showed skewness towards PHS susceptibility while frequency distribution with intermediate SGI value was almost normal, indicating the involvement of multiple alleles in controlling PHS resistance.

Genetic linkage maps are important tools for QTL mapping and map-based cloning of genes controlling quantitative traits such as PHS resistance and marker-assisted breeding. The availability of several high-throughput marker systems such as SNP, DArT and GBS has facilitated the development of high density linkage maps in wheat. Of these, SNP based linkage maps are commonly used in genomic studies since SNPs are more abundant in the genome.

Having a shorter map length and higher marker density are critical to increase the precision of QTL and gene mapping. Thus, increasing the size of a mapping population is necessary as it provides higher genetic variation, which offers more markers for a greater genome coverage and construction of high density linkage map. The present study constructed a high-density linkage map of 2343.9 cM using the mapping population studied. The linkage map comprised a total of 8751 SNPs covering all the 21 chromosomes of common wheat where the marker density was 3.7 SNPs/cM. The SNPs were unevenly distributed across the genome and the longest map with the highest number of SNPs was found in the B sub-genome because the chromosomes are long. While the shortest map with the lowest number of SNPs was observed in the D sub-genome due to recent hybridization and the short nature of the chromosomes.

QTL mapping is a powerful approach to elucidate complex genetic architectures of quantitative traits like PHS resistance. Greater efforts have been made to identify QTLs for PHS resistance and their underlying genes in cereals including wheat, rice, and barley. In this study, a total of five QTLs on chromosome 1D, 4A, 6B, 6D and 7A were detected. Of these, the 4A QTL (*QPhs.umb-4A*) was detected consistently across all the environments and had large effect in phenotypic variation (approx. 50%) in studied population. This *QPhs.umb-4A* was located on chromosome 4A flanked by BS00068243_51 and BS00037019_51. BLAST search revealed the position of BS00068243_51 at 603.067 Mbp on chromosome 4A of Chinese Spring RefSeq v1.0. This QTL showed highest positive additive effect, indicating resistance allele was contributed by PHS resistant parent RL4137 at this locus. *QPhs.umb-1D* was the novel PHS resistance QTL identified by our study and it was consistently detected in more than two environments. This QTL was identified at 142.5 cM of chromosome 1D and had a minor effect in phenotypic variation (approx. 2.3%). The QTL *QPhs.umb-1D* showed negative additive effect, indicating the resistance alleles were contributed by susceptible parent ‘Roblin’. The flanking SNPs of 1D QTL, BS00022188_51 and Kukri_c10485_1346 are located at 462.203 Mbp and 475.724 Mbp of wheat reference genome, respectively. Other QTLs in this study were also detected in more than two environments except 7A QTL and had minor effects in phenotypic variation of the population. RL4137 contributed the resistant allele at those loci including 7A QTL. Combined effect analysis of QTLs provides information about the contribution of individual QTLs in phenotypic variation of a mapping population, which facilitates pyramiding of QTLs in breeding program. Combined effect analysis of 1D, 4A, 6B and 6D QTLs revealed that the genotypes

containing resistance allele at all four loci showed the highest PHS resistance in the studied population.

The first chapter of the research in general identified QTLs that may be vital for PHS resistance, however, further study is still needed for better understanding the genetic mechanisms of PHS resistance in wheat. Some important QTLs for PHS resistance such as 3A and 4B QTLs were not detected in this study. The possible reason for this could be lack of those alleles in parental genotypes or the alleles are common between two parents thus did not segregate in the mapping population. Further study is needed to clarify the reason for this. Study of PHS using new genetic resources/germplasms is necessary to validate our result and identify new QTLs. In addition, fine mapping of the chromosomal regions identified by this study as associated with PHS using various types of markers including SSR, DArT and GBS is critical to clone PHS resistance genes or identify markers that are closely linked with genes/QTLs for marker-assisted selection.

The second research chapter of the thesis was aimed to identify molecular mechanisms regulating PHS resistance in wheat using genome-wide transcriptomic analysis of selected dormant and non-dormant DH lines. The application of NGS technologies in gene expression analysis such as RNA-seq technology is enabling us to understand complex biological processes controlling quantitative trait through offering the opportunities to perform genome-wide gene expression analysis precisely. This approach can also identify the location of genes differentially expressed between two contrasting genotypes in the genome and detect variations in gene sequences using available wheat reference genome information. GO enrichment analysis of the genes differentially expressed between the dormant and non-dormant seeds indicated that DEGs mainly those downregulated in imbibed dormant seeds are enriched in the cell, cell part, intracellular and integral component of membrane GO categories under the cellular component domain, indicating the role these biological process in the regulation of seed dormancy in wheat. Pathway enrichment analysis reported that pathways related to hormone signal transduction and biosynthesis of secondary metabolites are enriched with genes differentially expressed between the imbibed dormant and non-dormant wheat seeds. The ABA and ET signal transduction pathways are enriched with genes upregulated in dormant seeds while auxin signaling pathway is enriched with genes downregulated in the same sample. Regarding the secondary metabolites biosynthesis pathway, the diterpenoid, carotenoid and flavonoid biosynthesis pathways are

enriched with genes mainly upregulated in dormant samples. These results suggest hormonal regulation of seed dormancy and provide important information on how hormone metabolism and signaling affect seed dormancy and PHS resistance in wheat.

Expression analysis of individual gene involved in the hormone signal transduction and biosynthesis of secondary metabolites revealed that *TaPP2C*s, *TaSnRK2*s, and *TaABI5*s involved in ABA signaling, *TaIAA9* in auxin signaling and *TaETR4*s, *TaEIN2*s in ET signaling pathways are upregulated in imbibed dormant seeds. While *TaPYL*s that encode the ABA receptor protein, *TaIAA13*s, *TaIAA*20 and *TaIAA*27 genes involved in auxin signaling, and *TaETR3* involved in ET signaling pathways were downregulated in same samples. However, upregulation of *TaPP2C*s of ABA, *TaIAA9* of auxin signaling, and *TaETR4*s and *TaEIN2*s of ET signaling and downregulation of *TaPYL*s of ABA signaling pathways lacked correlation with observed seed dormancy phenotype, indicating that they don't regulate seed dormancy in wheat at transcriptional level. In secondary metabolites biosynthesis, the ABA biosynthesis genes *TaNCED*s and *TaPSY*s that are designated by our pathway analysis under the carotenoid biosynthesis, the GA catabolism gene *GA2ox3* that is categorized under diterpenoid biosynthesis and the *TaBAN* and *TaANS-D* that are categorized under flavonoid biosynthesis pathways were upregulated while the GA metabolism gene *GA20ox1*, designated under diterpenoid biosynthesis pathways, was downregulated in imbibed dormant seeds. These results indicate the involvement of those genes in regulation of seed dormancy and PHS resistance in wheat through controlling hormone metabolism and signaling. In addition, our analysis identified transcription factors belonging to 55 families as differentially expressed between dormant and non-dormant seeds. Among the TF families, the highest number of differentially expressed TFs belongs to the AP2-EREBP family, and one of that family member *TaDREB2C*, showed the highest upregulation in imbibed dormant seed. In contrast, a member of the C/EBP TF family, *histoneH2A.2.1-A*, was highly downregulated TF in the same sample. These results indicate that the two TFs may play important roles in regulating dormancy/PHS resistance related genes.

Integration of RNA-Seq data with our previous genetic study identified potential candidate genes in QTL regions detected by the genetic study. Of these, the *TaCYP71B11* and *TaDOG1L-2* genes located on 1D, *TaPM19-A1* and *TaPM19-A2* located on 4A, an uncharacterized gene with gene ID: TraesCS6B02G325000 located on 6B and *TaLEA4* located on 6D QTL regions showed substantial upregulation in imbibed dormant seed; while

uncharacterized genes with gene IDs: TraesCS1D02G410700 on 1D and TraesCS4A02G317200 on 4A; *TaRH8* on 6B and *TaHIS4* on 6D QTL exhibited downregulation in the same sample. Among the upregulated genes, *TaPM19-A1* and *A2* were identified previously as candidate genes for PHS resistance and they can modulate ABA responsiveness to improve PHS resistance in wheat. The present study is the first to report the localization of *TaDOG1L-2* within seed dormancy QTL and validate the candidacy of *TaPM19-A1* and *A2* for PHS resistance. The present study also identified 32 SNPs in DEGs located within QTL regions identified by our previous study including 2 SNPs on 1D, 2 SNPs on 4A, 25 SNPs on 6B and 3 SNPs on 6D QTLs. Positional analysis of those SNPs revealed that the SNPs were present in the non-coding region of the DEGs. Taken together, the second chapter of the thesis identified molecular mechanisms and candidate genes for seed dormancy/PHS resistance. Thus, further investigation of the candidate genes is needed to uncover their function in dormancy/PHS through approaches that involve gene mutation such as gene editing approach. The SNPs identified in this study can be used to develop KASP markers for marker-assisted breeding via testing their specificity and efficacy of targeting the specific genes associated with seed dormancy.

In conclusion, the outcomes of this thesis contribute to further our understanding of the genetic and molecular mechanisms of PHS resistance in wheat. In addition, the results of this study can have important contribution in wheat breeding for PHS resistance. The breeders can stack the QTLs identified in their breeding program especially the major effect 4A QTL and novel 1D QTL. The markers identified have potential to identify the same chromosomal regions in other genomic studies and can be used to develop user friendly Kompetitive Allele Specific PCR (KASP) marker. The PHS resistant germplasms identified in our study have the potential to be used as source of PHS resistance in future breeding program. The candidate genes identified through the transcriptomic study have the potential to pyramid genes to improve PHS resistance in commercial cultivars.

LITERATURE CITED

Adams MD, Kelley JM, Gocayne JD, et al (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252:1651–1656. https://doi.org/10.1126/science.2047873

Akkaya MS, Bhagwat AA, Cregan PB (1992) Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131–1139. https://doi.org/10.1093/genetics/132.4.1131

Ali A, Cao J, Jiang H, et al (2019) Unraveling molecular and genetic studies of wheat (*Triticum aestivum* L.) resistance against factors causing pre-harvest sprouting. Agronomy 9:117

Ali F, Qanmber G, Li F, Wang Z (2022) Updated role of ABA in seed maturation, dormancy, and germination. J Adv Res 35:199–214. https://doi.org/https://doi.org/10.1016/j.jare.2021.03.011

Allen AM, Winfield MO, Burridge AJ, et al (2017) Characterization of a Wheat Breeders' Array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (*Triticum aestivum*). Plant Biotechnol J 15:390–401. https://doi.org/https://doi.org/10.1111/pbi.12635

Alvarado G, López M, Mateo V, et al (2015) META-R (Multi Environment Trial Analysis with R for Windows) Version 6.04. CIMMYT Research Data & Software Repository Network, V23. https://hdl.handle.net/11529/10201

Amiteye S (2021) Basic concepts and methodologies of DNA marker systems in plant molecular breeding. Heliyon 7:e08093

An YQ, Lin L (2011) Transcriptional regulatory programs underlying barley germination and regulatory functions of gibberellin and abscisic acid. BMC Plant Biol 11:105. https://doi.org/10.1186/1471-2229-11-105

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106. https://doi.org/10.1186/gb-2010-11-10-r106

Anders S, Huber W (2012) Differential expression of RNA-Seq data at the gene level–the DESeq package. Heidelberg, Ger Eur Mol Biol Lab 10:f1000research.

Anderson JA, Sorrells ME, Tanksley SD (1993) RFLP Analysis of genomic regions associated with resistance to pre-harvest sprouting in wheat. Crop Sci 33:453–459 https://doi.org/10.2135/cropsci1993.0011183X003300030008x

Appels R, Kellye E, Nils S, et al (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361:eaar7191. https://doi.org/10.1126/science.aar7191

Arc E, Sechet J, Corbineau F, et al (2013) ABA crosstalk with ethylene and nitric oxide in seed dormancy and germination. Front. Plant Sci. 4:63. https://doi.org/10.3389/fpls.2013.00063

Ashburner M, Ball CA, Blake JA, et al (2000) Gene Ontology: tool for the unification of

biology. Nat Genet 25:25–29. https://doi.org/10.1038/75556

Ates D, Aldemir S, Alsaleh A, et al (2018) A consensus linkage map of lentil based on DArT markers from three RIL mapping populations. PLoS One 13:e0191375. https://doi.org/10.1371/journal.pone.0191375

Augenlicht LH, Kobrin D (1982) Cloning and screening of sequences expressed in a mouse colon tumor. Cancer Res 42:1088–1093

Båga M, Chodaparambil S V, Limin AE, et al (2007) Identification of quantitative trait loci and associated candidate genes for low-temperature tolerance in cold-hardy winter wheat. Funct Integr Genomics 7:53–68. https://doi.org/10.1007/s10142-006-0030-7

Bailey PC, McKibbin RS, Lenton JR, et al (1999) Genetic map locations for orthologous *Vp1* genes in wheat and rice. Theor Appl Genet 98:281–284. https://doi.org/10.1007/s001220051069

Barbeau WE, Griffey CA, Yan Z (2006) Evidence that minor sprout damage can lead to significant reductions in gluten strength of winter wheats. Cereal Chem 83:306–310. https://doi.org/10.1094/CC-83-0306

Barrero JM, Cavanagh C, Verbyla KL, et al (2015) Transcriptomic analysis of wheat near-isogenic lines identifies *PM19-A1* and *A2* as candidates for a major dormancy QTL. Genome Biol 16:93. https://doi.org/10.1186/s13059-015-0665-6

Barrero JM, Downie AB, Xu Q, et al (2014) A role for barley *CRYPTOCHROME1* in light regulation of grain dormancy and germination. Plant Cell 26:1094–1104. https://doi.org/10.1105/tpc.113.121830

Barrero JM, Jacobsen J V, Talbot MJ, et al (2012) Grain dormancy and light quality effects on germination in the model grass *Brachypodium distachyon*. New Phytol 193:376–386. https://doi.org/10.1111/j.1469-8137.2011.03938.x

Bar-Yosef O (1998) The Natufian culture in the Levant, threshold to the origins of agriculture. Evol Anthropol 6:159–177. https://doi.org/10.1002/(SICI)1520-6505(1998)6:5<159::AID-EVAN4>3.0.CO;2-7

Baskin JM, Baskin CC (2004) A classification system for seed dormancy. Seed Sci Res 14:1–16. https://doi.org/10.1079/SSR2003150

Bazin J, Langlade N, Vincourt P, et al (2011) Targeted mRNA oxidation regulates sunflower seed dormancy alleviation during dry after-ripening. Plant Cell 23:2196–2208. https://doi.org/10.1105/tpc.111.086694

Begna T, Yesuf H (2021) Genetic mapping in crop plants. Open J Plant Sci 6:019–026. https://doi.org/10.17352/ojps.000028

Benech-Arnold RL, Gualano N, Leymarie J, et al (2006) Hypoxia interferes with ABA metabolism and increases ABA sensitivity in embryos of dormant barley grains. J Exp Bot 57:1423–1430. https://doi.org/10.1093/jxb/erj122

Bentsink L, Koornneef M (2008) Seed dormancy and germination. Arab B 6:e0119–e0119. https://doi.org/10.1199/tab.0119

Berger S, Bell E, Mullet JE (1996) Two methyl jasmonate-insensitive mutants show altered expression of *AtVsp* in response to methyl jasmonate and wounding. Plant Physiol 111:525–531. https://doi.org/10.1104/pp.111.2.525

Bewley JD (1997) Seed germination and dormancy. Plant Cell 9:1055–1066. https://doi.org/10.1105/tpc.9.7.1055

Bewley JD, Bradford KJ, Hilhorst HWM, et al (2013) Seeds: Physiology of development, germination and dormancy, 3rd edn. Springer Science & Business Media, Springer New York, NY

Boeven PHG, Longin CFH, Leiser WL, et al (2016) Genetic architecture of male floral traits required for hybrid wheat breeding. Theor Appl Genet 129:2343–2357. https://doi.org/10.1007/s00122-016-2771-6

Boopathi NM (2020) Mapping population development. In: Genetic mapping and marker assisted selection. Springer, Singapore, pp. 69-106.

Bowden WM (1959) The taxonomy and nomenclature of the wheats, barleys, and ryes and their wild relatives. Can J Bot 37:657–684. https://doi.org/10.1139/b59-053

Brady SM, Sarkar SF, Bonetta D, et al (2003) The *ABSCISIC ACID INSENSITIVE 3* (*ABI3*) gene is modulated by farnesylation and is involved in auxin signaling and lateral root development in Arabidopsis. Plant J 34:67–75. https://doi.org/10.1046/j.1365-313X.2003.01707.x

Braun HJ, Atlin G, Payne T (2010) Multi-location testing as a tool to identify plant response to global climate change. In: Reynlods MP (Ed) Climate change and crop production. CABI International, Oxfordshire, pp.115-138. https://doi.org/10.1080/01916122.2004.9989599

Buller AHR (1919) Essays on wheat, including the discovery and introduction of marquis wheat, the early history of wheat-growing in Manitoba, wheat in western Canada, the origin of red bobs and kitchener, and the wild wheat of Palestine. New York, The Macmillan Company. Image. https://www.loc.gov/item/20000838/

Cabral AL, Jordan MC, McCartney CA, et al (2014) Identification of candidate genes, regions and markers for pre-harvest sprouting resistance in wheat (*Triticum aestivum* L.). BMC Plant Biol 14:340. https://doi.org/10.1186/s12870-014-0340-1

Cadalen T, Boeuf C, Bernard S, et al (1997) An intervarietal molecular marker map in *Triticum aestivum* L. Em. Thell. and comparison with a map from a wide cross. Theor Appl Genet 94:367–377. https://doi.org/10.1007/s001220050425

Calderon-Villalobos LI, Tan X, Zheng N, et al (2010) Auxin perception--structural insights. Cold Spring Harb Perspect Biol 2:a005546. https://doi.org/10.1101/cshperspect.a005546

Campbell AB, Czarnecki E (1987) Roblin hard red spring wheat. Can J Plant Sci 67:803-804.

https://doi-org.uml.idm.oclc.org/10.4141/cjps87-108

Canadian wheat classes. https://grainscanada.gc.ca/en/grain-quality/grain-grading/wheat-classes.html. Accessed: November 4, 2021.

Carrillo-Barral N, Rodríguez-Gacio MD, Matilla AJ (2020) *Delay of Germination-1* (*DOG1*): A key to understanding seed dormancy. Plants 9:480. https://doi.org/10.3390/plants9040480

Cavanagh CR, Chao S, Wang S, et al (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc Natl Acad Sci 110:8057– 8062. https://doi.org/10.1073/pnas.1217133110

Chang C, Zhang HP, Feng JM, et al (2010) Identifying alleles of *Viviparous-1B* associated with pre-harvest sprouting in micro-core collections of Chinese wheat germplasm. Mol Breed 25:481–490. https://doi.org/10.1007/s11032-009-9346-z

Chao S, Sharp PJ, Worland AJ, et al (1989) RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. Theor Appl Genet 78:495–504. https://doi.org/10.1007/bf00290833

Chapman V, Miller TE, Riley R (1976) Equivalence of the A genome of bread wheat and that of *Triticum urartu*. Genet Res 27:69–76. https://doi.org/10.1017/S0016672300016244

Chen CX, Cai SB, Bai GH (2008) A major QTL controlling seed dormancy and pre-harvest sprouting resistance on chromosome 4A in a Chinese wheat landrace. Mol Breed 21:351–358. https://doi.org/10.1007/s11032-007-9135-5

Cheng X, Tian B, Gao C, et al (2021) Identification and expression analysis of candidate genes related to seed dormancy and germination in the wheat GATA family. Plant Physiol Biochem 169:343–359. https://doi.org/10.1016/j.plaphy.2021.11.012

Chitnis VR, Gao F, Yao Z, et al (2014) After-ripening induced transcriptional changes of hormonal genes in wheat seeds: the cases of brassinosteroids, ethylene, cytokinin and salicylic acid. PLoS One 9:e87543. https://doi.org/10.1371/journal.pone.0087543

Chiwocha SDS, Cutler AJ, Abrams SR, et al (2005) The *etr1-2* mutation in *Arabidopsis thaliana* affects the abscisic acid, auxin, cytokinin and gibberellin metabolic pathways during maintenance of seed dormancy, moist-chilling and germination. Plant J 42:35–48. https://doi.org/10.1111/j.1365-313X.2005.02359.x

Chono M, Honda I, Shinoda S, et al (2006) Field studies on the regulation of abscisic acid content and germinability during grain development of barley: molecular and chemical analysis of pre-harvest sprouting. J Exp Bot 57:2421–2434. https://doi.org/10.1093/jxb/erj215

Chono M, Matsunaka H, Seki M, et al (2013) Isolation of a wheat (*Triticum aestivum* L.) mutant in *ABA 8'-hydroxylase* gene: effect of reduced ABA catabolism on germination inhibition under field condition. Breed Sci 63:104–115. https://doi.org/10.1270/jsbbs.63.104

Clarke FR, Clarke JM, DePauw RM, et al (2005) Strategic approach to mitigating weather induced defects of wheat quality. Euphytica 143:285–290. https://doi.org/10.1007/s10681-

005-7883-z

Collard BC, Mackill DJ (2007) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Phil Trans R Soc B 363:557-572. https://doi.org/10.1098/rstb.2007.2170

Collard BCY, Jahufer MZZ, Brouwer JB, et al (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142:169–196. https://doi.org/10.1007/s10681-005-1681-5

Corbineau F, Xia Q, Bailly C, et al (2014) Ethylene, a key factor in the regulation of seed dormancy. Front. Plant Sci. 5:539. https://doi.org/10.3389/fpls.2014.00539

Creixell P, Reimand J, Haider S, et al (2015) Pathway and network analysis of cancer genomes. Nat Methods 12:615–621. https://doi.org/10.1038/nmeth.3440

Danquah A, de Zélicourt A, Boudsocq M, et al (2015) Identification and characterization of an ABA-activated MAP kinase cascade in *Arabidopsis thaliana*. Plant J 82:232–244. https://doi.org/10.1111/tpj.12808

De Giorgi J, Piskurewicz U, Loubery S, et al (2015) An endosperm-associated cuticle is required for Arabidopsis seed viability, dormancy and early control of germination. PLoS Genet 11:e1005708. https://doi.org/10.1371/journal.pgen.1005708

De Laethauwer S, De Riek J, Stals I, et al (2013) α-Amylase gene expression during kernel development in relation to pre-harvest sprouting in wheat and triticale. Acta Physiol Plant 35:2927–2938. https://doi.org/10.1007/s11738-013-1323-9

Debeaujon I, Lepiniec L, Pourcel L, et al (2007) Seed coat development and dormancy. In: Seed development, dormancy and germination. Wiley-Blackwell, pp 392

Dekkers BJW, Pearce SP, van Bolderen-Veldkamp RPM, et al (2016) Dormant and after-ripened *Arabidopsis thaliana* seeds are distinguished by early transcriptional differences in the imbibed state. Front. Plant Sci. 7:1323. https://doi.org/10.3389/fpls.2016.01323

DePauw R, Hunt T (2001) Canadian wheat pool. In: Bonjean AP, Angus WJ (ed) The world wheat book. Lavoisier Publishing, France, pp 479-515.

DePauw RM, Clarke FR, Fofana B, et al (2009) RL4137 contributes preharvest sprouting resistance to Canadian wheats. Euphytica 168:347–361. https://doi.org/10.1007/s10681-009-9933-4

DePauw RM, Knox RE, Singh AK, et al (2012) Developing standardized methods for breeding pre-harvest sprouting resistant wheat, challenges and successes in Canadian wheat. Euphytica 188:7–14. https://doi.org/10.1007/s10681-011-0611-y

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

Derera NF (1990) A perspective of sprouting research. In: Ringlund K, Mosleth E, Mares DJ

(eds) Fifth international symposium on pre-harvest sprouting in cereals. Westview Press, Colorado, USA, pp 3-11.

Derera NF (2018) Breeding for preharvest sprouting tolerance. In: Preharvest Field Sprouting in Cereals. CRC Press, pp 111–128.

Diouf I, Albert E, Duboscq R, et al (2020) Integration of QTL, transcriptome and polymorphism studies reveals candidate genes for water stress response in tomato. Genes 11:900. https://doi.org/10.3390/genes11080900

Doerge RW (2002) Mapping and analysis of quantitative trait loci in experimental populations. Nat Rev Genet 3:43–52. https://doi.org/10.1038/nrg703

Eksi G, Kurbanoglu S, Erdem SA (2020) Chapter 9 - Analysis of diterpenes and diterpenoids. In: Sanches Silva A, Nabavi SF, Saeedi M, Nabavi SMBT-RA in NPA (ed). Elsevier, pp 313–345. https://doi.org/10.1016/B978-0-12-816455-6.00009-3

FAO. 2021. World Food and Agriculture – Statistical Yearbook 2021. Rome. pp. 157

Fedak G, Burvill M, Voldeng H (1997) A comparison of anther culture and maize pollination for haploid production in wheat. J Appl Genet 38: 407–414.

Feldman M (1995) Wheats. In: Evolution of crop plants. The Weizmann Institute of Science, pp 184–192

Feldman M, Levy AA (2012) Genome evolution due to allopolyploidization in wheat. Genetics 192:763–774. https://doi.org/10.1534/genetics.112.146316

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. https://doi.org/10.1093/jxb/erw477

Finch-Savage WE, Leubner-Metzger G (2006) Seed dormancy and the control of germination. New Phytol 171:501–523. https://doi.org/10.1111/j.1469-8137.2006.01787.x

Finkelstein R, Reeves W, Ariizumi T, et al (2008) Molecular aspects of seed dormancy. Annu Rev Plant Biol 59:387–415. https://doi.org/10.1146/annurev.arplant.59.032607.092740

Finkelstein RR, Gampala SSL, Rock CD (2002) Abscisic acid signaling in seeds and seedlings. Plant Cell 14:15–45. https://doi.org/10.1105/tpc.010441

Flintham J, Adlam R, Bassoi M, et al (2002) Mapping genes for resistance to sprouting damage in wheat. Euphytica 126:39-45. https://doi.org/10.1023/A:1019632008244

Flintham JE (2000) Different genetic components control coat-imposed and embryo-imposed dormancy in wheat. Seed Sci Res 10:43–50. https://doi.org/10.1017/S0960258500000052

Fofana B, Humphreys DG, Rasul G, et al (2009) Mapping quantitative trait loci controlling pre-harvest sprouting resistance in a red × white seeded spring wheat cross. Euphytica 165:509–521. https://doi.org/10.1007/s10681-008-9766-6

Fraser JGC, Whiteside AGO (1956) Handbook of Canadian spring wheat varieties. pp. 45

Frey A, Effroy D, Lefebvre V, et al (2012) Epoxycarotenoid cleavage by *NCED5* fine-tunes ABA accumulation and affects seed dormancy and drought tolerance with other NCED family members. Plant J 70:501–512. https://doi.org/10.1111/j.1365-313X.2011.04887.x

Gao F, Ayele BT (2014) Functional genomics of seed dormancy in wheat: advances and prospects. Front. Plant Sci. 5:458. https://doi.org/10.3389/fpls.2014.00458

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. https://doi.org/10.1111/j.1467-7652.2012.00682.x

Gao X, Hu CH, Li HZ, et al (2013) Factors affecting pre-harvest sprouting resistance in wheat (*Triticum aestivum* L.): A review. J Anim Plant Sci 23:556–565.

Gardner KA, Wittern LM, Mackay IJ (2016) A highly recombined, high-density, eight-founder wheat MAGIC map reveals extensive segregation distortion and genomic locations of introgression segments. Plant Biotechnol J 14:1406–1417. https://doi.org/10.1111/pbi.12504

Gatford KT, Eastwood RF, Halloran GM (2002) Germination inhibitors in bracts surrounding the grain of *Triticum tauschii*. Funct Plant Biol 29:881–890. https://doi.org/10.1071/PP01011

Ghahremaninejad F, Hoseini E, Jalali S (2021) The cultivation and domestication of wheat and barley in Iran, brief review of a long history. Bot Rev 87:1–22. https://doi.org/10.1007/s12229-020-09244-w

Ghassemian M, Nambara E, Cutler S, et al (2000) Regulation of abscisic acid signaling by the ethylene response pathway in Arabidopsis. Plant Cell 12:1117–1126. https://doi.org/10.1105/tpc.12.7.1117

Ghorbani M, Jonckheere EA, Bogdan P (2018) Gene expression is not random: scaling, long-range cross-dependence, and fractal characteristics of gene regulatory networks. Front. Physiol. 9:1446. https://doi.org/10.3389/fphys.2018.01446

Gianinetti A, Finocchiaro F, Bagnaresi P, et al (2018) Seed dormancy involves a transcriptional program that supports early plastid functionality during imbibition. Plants 7:35. https://doi.org/10.3390/plants7020035

Gill BS, Friebe B (2002) Cytogenetics, phylogeny and evolution of cultivated wheats. Bread wheat: improvement and production. Food and Agriculture Organization of the United Nations, Rome, pp.71-88.

Goncharov NP (2011) Genus *Triticum* L. taxonomy: the present and the future. Plant Syst Evol 295:1–11. https://doi.org/10.1007/s00606-011-0480-9

Gordon IL, Balaam LN, Derera NF (1979) Selection against sprouting damage in wheat. II.* Harvest ripeness, grain maturity and germinability. Aust J Agric Res 30:1–17.

Gordon IL, Derera NF, Balaam LN (1977) Selection against sprouting damage in wheat I. Germination of unthreshed grain, with a standard wetting procedure. Aust J Agric Res 28:583–596.

Grain varieties by acreage insured. https://www.grainscanada.gc.ca/en/grain-research/statistics/varieties-by-acreage/. Accessed: November 4, 2021.

Grodzicker T, Williams J, Sharp P, et al (1975) Physical mapping of temperature-sensitive mutations of adenoviruses. Cold Spring Harb Symp Quant Biol 39:439–446. https://doi.org/10.1101/sqb.1974.039.01.056

Groos C, Gay G, Perretant MR, et al (2002) Study of the relationship between pre-harvest sprouting and grain color by quantitative trait loci analysis in a white×red grain bread-wheat cross. Theor Appl Genet 104:39–47. https://doi.org/10.1007/s001220200004

Gualano NA, Benech-Arnold RL (2009) Predicting pre-harvest sprouting susceptibility in barley: Looking for "sensitivity windows" to temperature throughout grain filling in various commercial cultivars. F Crop Res 114:35–44. https://doi.org/10.1016/j.fcr.2009.06.016

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. https://doi.org/10.1104/pp.107.115469

Gupta PK, Rustgi S, Mir RR (2008) Array-based high-throughput DNA markers for crop improvement. Heredity 101:5–18. https://doi.org/10.1038/hdy.2008.35

Gutierrez-Gonzalez JJ, Mascher M, Poland J, et al (2019) Dense genotyping-by-sequencing linkage maps of two Synthetic W7984×Opata reference populations provide insights into wheat structural diversity. Sci Rep 9:1793. https://doi.org/10.1038/s41598-018-38111-3

Hartl DL, Jones EW (2009) Genetics: analysis of genes and genomes, 7th edn. Jones & Bartlett Learning, Sudbury, MA, pp. 176-220.

Hatcher DW, Symons SJ (2000) Influence of sprout damage on oriental noodle appearance as assessed by image analysis. Cereal Chem 77:380–387. https://doi.org/10.1094/CCHEM.2000.77.3.380

Hawkesford MJ, Araus JL, Park R, et al (2013) Prospects of doubling global wheat yields. Food Energy Secur 2:34–48. https://doi.org/10.1002/fes3.15

He J, Zhang D, Chen X, et al (2021) Identification of QTLs and a Candidate Gene for Reducing Pre-Harvest Sprouting in *Aegilops tauschii–Triticum aestivum* Chromosome Segment Substitution Lines. Int. J. Mol. Sci. 22:3729. https://doi.org/10.3390/ijms22073729

Head L, Atchison J (2012) Ingrained : A human bio-geography of wheat. Routledge, Farnham, Surrey. pp.16-35.

Heffelfinger C, Fragoso CA, Lorieux M (2017) Constructing linkage maps in the genomics era with MapDisto 2.0. Bioinformatics 33:2224–2225. https://doi.org/10.1093/bioinformatics/btx177

Heun M, Ralf SP, Dieter K, et al (1997) Site of einkorn wheat domestication identified by DNA fingerprinting. Science 278:1312–1314. https://doi.org/10.1126/science.278.5341.1312

Heydecker W (1973) Glossary of terms. Seed Ecol (W Heydecker, ed) Butterworths, London pp.553–557

Hilhorst HWM, Karssen CM (1992) Seed dormancy and germination: the role of abscisic acid and gibberellins and the importance of hormone mutants. Plant Growth Regul 11:225–238. https://doi.org/10.1007/BF00024561

Himi E, Maekawa M, Miura H, et al (2011) Development of PCR markers for *Tamyb10* related to *R-1*, red grain color gene in wheat. Theor Appl Genet 122:1561–1576. https://doi.org/10.1007/s00122-011-1555-2

Himi E, Mares DJ, Yanagisawa A, et al (2002) Effect of grain colour gene (*R*) on grain dormancy and sensitivity of the embryo to abscisic acid (ABA) in wheat. J Exp Bot 53:1569–1574. https://doi.org/10.1093/jxb/erf005

Himi E, Noda K (2005) Red grain colour gene (*R*) of wheat is a Myb-type transcription factor. Euphytica 143:239–242. https://doi.org/10.1007/s10681-005-7854-4

Hoang HH, Bailly C, Corbineau F, et al (2013) Induction of secondary dormancy by hypoxia in barley grains and its hormonal regulation. J Exp Bot 64:2017–2025. https://doi.org/10.1093/jxb/ert062

Hoshino T, Tomooka N, Fukunaga K, et al (1989) Testing method of pre-harvest sprouting and genealogical pedigree of pre-harvest sprouting resistant cultivars in wheat. Japanese J Breed 39:365–372. https://doi.org/10.1270/jsbbs1951.39.365

Howell KA, Narsai R, Carroll A, et al (2009) Mapping metabolic and transcript temporal switches during germination in rice highlights specific transcription factors and the role of RNA instability in the germination process. Plant Physiol 149:961–980. https://doi.org/10.1104/pp.108.129874

Howitt CA, Pogson BJ (2006) Carotenoid accumulation and function in seeds and non-green tissues. Plant Cell Environ 29:435–445. https://doi.org/10.1111/j.1365-3040.2005.01492.x

Hu Z, Liu X, Tian M, et al (2021) Recent progress and new perspectives for diterpenoid biosynthesis in medicinal plants. Med Res Rev 41:2971–2997. https://doi.org/10.1002/med.21816

Hucl P, Chibbar RN (1996) Variation for starch concentration in spring wheat and its repeatability relative to protein concentration. Cereal Chem 73:756–758.

Humphreys DG, Noll J (2002) Methods for characterization of pre-harvest sprouting resistance in a wheat breeding program. Euphytica 126:61–65. https://doi.org/10.1023/A:1019671622356

Hussain S, Kim SH, Bahk S, et al (2020) The auxin signaling repressor *IAA8* promotes seed germination through down-regulation of *ABI3* transcription in Arabidopsis. Front. Plant Sci.

11:111. https://doi.org/10.3389/fpls.2020.00111

Hussain W, Baenziger PS, Belamkar V, et al (2017) Genotyping-by-Sequencing derived high-density linkage map and its application to QTL mapping of flag leaf traits in bread wheat. Sci Rep 7:16394. https://doi.org/10.1038/s41598-017-16006-z

Imtiaz M, Ogbonnaya FC, Oman J, et al (2008) Characterization of quantitative trait loci controlling genetic variation for preharvest sprouting in synthetic backcross-derived wheat lines. Genetics 178:1725-1736. https://doi.org/10.1534/genetics.107.084939

Izydorczyk C, Nguyen TN, Jo S, et al (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. https://doi.org/10.1111/pce.12949

Je J, Chen H, Song C, et al (2014) Arabidopsis *DREB2C* modulates ABA biosynthesis during germination. Biochem Biophys Res Commun 452:91–98. https://doi.org/10.1016/j.bbrc.2014.08.052

Jia H, Wang X, Shi Y, et al (2020) Overexpression of *Medicago sativa LEA4-4* can improve the salt, drought, and oxidation resistance of transgenic Arabidopsis. PLoS One 15:e0234085. https://doi.org/10.1371/journal.pone.0234085

Jia M, Guan J, Zhai Z, et al (2018) Wheat functional genomics in the era of next generation sequencing: An update. Crop J 6:7–14. https://doi.org/10.1016/j.cj.2017.09.003

Jian H, Zhang A, Ma J, et al (2019) Joint QTL mapping and transcriptome sequencing analysis reveal candidate flowering time genes in *Brassica napus* L. BMC Genomics 20:21. https://doi.org/10.1186/s12864-018-5356-8

Jiang GL (2013) Molecular markers and marker-assisted breeding in plants. In: Plant breeding from laboratories to fields, 3:45-83.

Jiang H, Zhao LX, Chen XJ, et al (2018) A novel 33-bp insertion in the promoter of *TaMFT-3A* is associated with pre-harvest sprouting resistance in common wheat. Mol Breed 38:69. https://doi.org/10.1007/s11032-018-0830-1

Jones N, Ougham H, Thomas H, et al (2009) Markers and mapping revisited: finding your gene. New Phytol 183:935–966. https://doi.org/10.1111/j.1469-8137.2009.02933.x

Kadirvel P, Senthilvel S, Geethanjali S, et al (2015) Genetic markers, trait mapping and marker-assisted selection in plant breeding. In: Plant biology and biotechnology. Springer, New Delhi, pp. 65-88.

Kanjana W, Suzuki T, Ishii K, et al (2016) Transcriptome analysis of seed dormancy after rinsing and chilling in ornamental peaches (*Prunus persica* (L.) Batsch). BMC Genomics 17:575. https://doi.org/10.1186/s12864-016-2973-y

Kao CH (2000) On the differences between maximum likelihood and regression interval mapping in the analysis of quantitative trait loci. Genetics 156:855–865.

https://doi.org/10.1093/genetics/156.2.855

Kashiwakura Y, Kobayashi D, Jikumaru Y, et al (2016) Highly sprouting-tolerant wheat grain exhibits extreme dormancy and cold imbibition-resistant accumulation of abscisic acid. Plant Cell Physiol 57:715–732. https://doi.org/10.1093/pcp/pcw051

Kato K, Nakamura W, Tabiki T, et al (2001) Detection of loci controlling seed dormancy on group 4 chromosomes of wheat and comparative mapping with rice and barley genomes. Theor Appl Genet 102:980–985. https://doi.org/10.1007/s001220000494

KeÇpczyński J, KeÇpczyńska E (1997) Ethylene in seed dormancy and germination. Physiol Plant 101:720–726. https://doi.org/10.1111/j.1399-3054.1997.tb01056.x

Kendall S, Penfield S (2012) Maternal and zygotic temperature signalling in the control of seed dormancy and germination. Seed Sci Res 22:23–29. https://doi.org/10.1017/S0960258511000390

Kepinski S, Leyser O (2005) The Arabidopsis F-box protein TIR1 is an auxin receptor. Nature 435:446–451. https://doi.org/10.1038/nature03542

Keswani C, Mishra S, Sarma BK, et al (2014) Unraveling the efficient applications of secondary metabolites of various *Trichoderma spp*. Appl Microbiol Biotechnol 98:533–544. https://doi.org/10.1007/s00253-013-5344-5

Khatri P, Sirota M, Butte AJ (2012) Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Comput Biol 8:e1002375. https://doi.org/10.1371/journal.pcbi.1002375

Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12:357–360. https://doi.org/10.1038/nmeth.3317

Kim JS, Lim JY, Shin H, et al (2019) *ROS1*-dependent DNA demethylation is required for ABA-inducible *NIC3* expression. Plant Physiol 179:1810–1821. https://doi.org/10.1104/pp.18.01471

Kim KH, Kim JY (2021) Understanding wheat starch metabolism in properties, environmental stress condition, and molecular approaches for value-added utilization. Plants 10:2282. https://doi.org/10.3390/plants10112282

Kim KH, Tsao R, Yang R, et al (2006) Phenolic acid profiles and antioxidant activities of wheat bran extracts and the effect of hydrolysis conditions. Food Chem 95:466–473. https://doi.org/10.1016/j.foodchem.2005.01.032

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. https://doi.org/10.1071/AR9840327

Klupczyńska EA, Pawłowski TA (2021) Regulation of seed dormancy and germination mechanisms in a changing environment. Int. J. Mol. Sci. 22:1357. https://doi.org/10.3390/ijms22031357

Knox RE, Clarke FR, Clarke JM, et al (2012) Enhancing the identification of genetic loci and transgressive segregants for pre-harvest sprouting resistance in a durum wheat population. Euphytica 186:193–206. https://doi.org/10.1007/s10681-011-0557-0

Knox RE, Clarke FR, Clarke JM, et al (2005) Genetic analysis of pre-harvest sprouting in a durum wheat cross. Euphytica 143:261–264. https://doi.org/10.1007/s10681-005-7874-0

Knox RE, Fox SL, Clarke FR, et al (1998) Genetic analysis of dormancy in *Triticum aestivum* and *Triticum turgidum*. In: Proceedings of the 9th International Wheat Genetics Symposium, University of Saskatchewan, Canada, pp 2-7.

Kosambi DD (2016) The estimation of map distances from recombination values. In: DD Kosambi. Springer, New Delhi, pp. 125-130.

Kottearachchi NS, Uchino N, Kato K, et al (2006) Increased grain dormancy in white-grained wheat by introgression of preharvest sprouting tolerance QTLs. Euphytica 152:421–428. https://doi.org/10.1007/s10681-006-9231-3

Kujur A, Upadhyaya HD, Shree T, et al (2015) Ultra-high density intra-specific genetic linkage maps accelerate identification of functionally relevant molecular tags governing important agronomic traits in chickpea. Sci Rep 5:9468. https://doi.org/10.1038/srep09468

Kulik A, Wawer I, Krzywińska E, et al (2011) SnRK2 protein kinases—key regulators of plant response to abiotic stresses. Omi A J Integr Biol 15:859–872. https://doi.org/10.1089/omi.2011.0091

Kulwal P, Ishikawa G, Benscher D, et al (2012) Association mapping for pre-harvest sprouting resistance in white winter wheat. Theor Appl Genet 125:793–805. https://doi.org/10.1007/s00122-012-1872-0

Kulwal PL, Kumar N, Gaur A, et al (2005). Mapping of a major QTL for pre-harvest sprouting tolerance on chromosome 3A in bread wheat. Theor Appl Genet 111:1052-1059. https://doi.org/10.1007/s00122-005-0021-4

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

Kulwal PL, Singh R, Balyan HS, et al (2004) Genetic basis of pre-harvest sprouting tolerance using single-locus and two-locus QTL analyses in bread wheat. Funct Integr Genomics 4:94-101. https://doi.org/10.1007/s10142-004-0105-2

Kumar A, Kumar J, Singh R, et al (2009) QTL analysis for grain colour and pre-harvest sprouting in bread wheat. Plant Sci 177:114–122. https://doi.org/10.1016/j.plantsci.2009.04.004

Kumar N, Kulwal PL, Gaur A, et al (2006) QTL analysis for grain weight in common wheat. Euphytica, 151:135-144. https://doi.org/10.1007/s10681-006-9133-4

Kumar S, Knox RE, Clarke FR, et al (2015) Maximizing the identification of QTL for pre-harvest sprouting resistance using seed dormancy measures in a white-grained hexaploid

wheat population. Euphytica 205:287–309. https://doi.org/10.1007/s10681-015-1460-x

Kushiro T, Okamoto M, Nakabayashi K, et al (2004) The Arabidopsis cytochrome P450 *CYP707A* encodes ABA 8′-hydroxylases: key enzymes in ABA catabolism. EMBO J 23:1647–1656. https://doi.org/10.1038/sj.emboj.7600121

Kwiatek MT, Kurasiak-Popowska D, Mikołajczyk S, et al (2019) Cytological markers used for identification and transfer of *Aegilops* spp. chromatin carrying valuable genes into cultivated forms of *Triticum*. Comp Cytogenet 13:41–59. https://doi.org/10.3897/CompCytogen.v13i1.30673

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199. https://doi.org/10.1093/genetics/121.1.185

Lander ES, Green P, Abrahamson J, et al (1987) MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181. https://doi.org/10.1016/0888-7543(87)90010-3

Lang J, Fu Y, Zhou Y, et al (2021) *Myb10-D* confers PHS-3D resistance to pre-harvest sprouting by regulating *NCED* in ABA biosynthesis pathway of wheat. New Phytol 230:1940–1952. https://doi.org/10.1111/nph.17312

Lawson WR, Godwin ID, Cooper M, et al (1997) Genetic analysis of pre-harvest sprouting tolerance in three wheat crosses. Aust J Agric Res 48:215-222. https://doi.org/10.1071/A96071

Layat E, Bourcy M, Cotterell S, et al (2021) The Histone Chaperone *HIRA* Is a Positive Regulator of Seed Germination. Int. J. Mol. Sci. 22:4031. https://doi.org/10.3390/ijms22084031

Lee YJ, Park SY, Kim DY, et al (2021) Differential expression analysis of phytohormone-related genes of Korean wheat (*Triticum aestivum*) in response to pre-harvest sprouting and abscisic acid (ABA). Appl. Sci. 11:3562. https://doi.org/10.3390/app11083562

Lei L, Zhu X, Wang S, et al (2013) *TaMFT-A1* is associated with seed germination sensitive to temperature in winter wheat. PLoS One 8:e73330. https://doi.org/10.1371/journal.pone.0073330

Leubner-Metzger G, Petruzzelli L, Waldvogel R, et al (1998) Ethylene-responsive element binding protein (EREBP) expression and the transcriptional regulation of class I β-1,3-glucanase during tobacco seed germination. Plant Mol Biol 38:785–795. https://doi.org/10.1023/A:1006040425383

Leyser O (2018) Auxin Signaling. Plant Physiol 176:465–479. https://doi.org/10.1104/pp.17.00765

Li C, Bai G, Chao S, et al (2015a) A high-density SNP and SSR consensus map reveals segregation distortion regions in wheat. Biomed Res Int 2015:830618. https://doi.org/10.1155/2015/830618

Li H, Vikram P, Singh RP, et al (2015b) A high density GBS map of bread wheat and its application for dissecting complex disease resistance traits. BMC Genomics 16:216. https://doi.org/10.1186/s12864-015-1424-5

Li LF, Zhang ZB, Wang ZH, et al (2022) Genome sequences of five *Sitopsis* species of *Aegilops* and the origin of polyploid wheat B subgenome. Mol Plant 15:488–503. https://doi.org/10.1016/j.molp.2021.12.019

Li S, Wang P, Yuan W, et al (2016) Endocidal regulation of secondary metabolites in the producing organisms. Sci Rep 6:29315. https://doi.org/10.1038/srep29315

Li T, Deng G, Tang Y, et al (2021) Identification and validation of a novel locus controlling spikelet number in bread wheat (*Triticum aestivum* L.). Front. Plant Sci. 12:611106. https://doi.org/10.3389/fpls.2021.611106

Lin M, Cai S, Wang S, et al (2015) Genotyping-by-sequencing (GBS) identified SNP tightly linked to QTL for pre-harvest sprouting resistance. Theor Appl Genet 128:1385–1395. https://doi.org/10.1007/s00122-015-2513-1

Lin M, Liu S, Zhang G, et al (2018) Effects of *TaPHS1* and *TaMKK3-A* genes on wheat pre-harvest sprouting resistance. Agronomy 8:210. https://doi.org/10.3390/agronomy8100210

Lin M, Zhang D, Liu S, et al (2016) Genome-wide association analysis on pre-harvest sprouting resistance and grain color in U.S. winter wheat. BMC Genomics 17:794. https://doi.org/10.1186/s12864-016-3148-6

Lincoln SE (1992) Mapping genes controlling quantitative traits with MAPMAKER/QTL 1.1. Whitehead Inst Tech Rep. pp. 1–46

Linkies A, Leubner-Metzger G (2012) Beyond gibberellins and abscisic acid: how ethylene and jasmonates control seed germination. Plant Cell Rep 31:253–270. https://doi.org/10.1007/s00299-011-1180-1

Linkies A, Müller K, Morris K, et al (2009) Ethylene interacts with abscisic acid to regulate endosperm rupture during germination: A comparative approach using *Lepidium sativum* and *Arabidopsis thaliana*. Plant Cell 21:3803–3822. https://doi.org/10.1105/tpc.109.070201

Liton MMUA, McCartney CA, Hiebert CW, et al (2021) Identification of loci for pre-harvest sprouting resistance in the highly dormant spring wheat RL4137. Theor Appl Genet 134:113–124. https://doi.org/10.1007/s00122-020-03685-y

Liu A, Gao F, Kanno Y, et al (2013a) Regulation of wheat seed dormancy by after-ripening is mediated by specific transcriptional switches that induce changes in seed hormone metabolism and signaling. PLoS One 8:e56570. https://doi.org/10.1371/journal.pone.0056570

Liu C, Zhang T (2017) Expansion and stress responses of the AP2/EREBP superfamily in cotton. BMC Genomics 18:118. https://doi.org/10.1186/s12864-017-3517-9

Liu F, Zhang H, Ding L, et al (2020) *REVERSAL OF RDO5 1*, a homolog of *rice seed*

dormancy4, interacts with *bHLH57* and controls ABA biosynthesis and seed dormancy in Arabidopsis. Plant Cell 32:1933–1948. https://doi.org/10.1105/tpc.20.00026

Liu J, Luo W, Qin N, et al (2018) A 55 K SNP array-based genetic map and its utilization in QTL mapping for productive tiller number in common wheat. Theor Appl Genet 131:2439–2450. https://doi.org/10.1007/s00122-018-3164-9

Liu S, Assanga SO, Dhakal S, et al (2016) Validation of chromosomal locations of 90K array single nucleotide polymorphisms in US Wheat. Crop Sci 56:364–373. https://doi.org/10.2135/cropsci2015.03.0194

Liu S, Bai G, Cai S, et al (2011) Dissection of genetic components of pre-harvest sprouting resistance in white wheat. Mol Breed 27:511–523. https://doi.org/10.1007/s11032-010-9448-7

Liu S, Sehgal SK, Li J, et al (2013b) Cloning and characterization of a critical regulator for pre-harvest sprouting in wheat. Genetics 195:263–273. https://doi.org/10.1534/genetics.113.152330

Liu X, Zhang H, Zhao Y, et al (2013c) Auxin controls seed dormancy through stimulation of abscisic acid signaling by inducing *ARF*-mediated *ABI3* activation in Arabidopsis. Proc Natl Acad Sci 110:15485–15490. https://doi.org/10.1073/PNAS.1304651110

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 25:402–408. https://doi.org/10.1006/meth.2001.1262

Lorieux M (2012) MapDisto: fast and efficient computation of genetic linkage maps. Mol Breed 30:1231-1235. http://doi.org/10.1007/s11032-012-9706-y

Lorsch JR (2002) RNA chaperones exist and DEAD box proteins get a life. Cell 109:797–800. https://doi.org/10.1016/S0092-8674(02)00804-8

Ma Z, Bykova NV, Igamberdiev AU (2017) Cell signaling mechanisms and metabolic regulation of germination and dormancy in barley seeds. Crop J 5:459–477. https://doi.org/10.1016/j.cj.2017.08.007

Maia J, Dekkers BJW, Dolle MJ, et al (2014) Abscisic acid (ABA) sensitivity regulates desiccation tolerance in germinated Arabidopsis seeds. New Phytol 203:81–93. https://doi.org/10.1111/nph.12785

Mao X, Cai T, Olyarchuk JG, et al (2005) Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. Bioinformatics 21:3787–3793. https://doi.org/10.1093/bioinformatics/bti430

Marcussen T, Sandve SR, Heier L, et al (2014) Ancient hybridizations among the ancestral genomes of bread wheat. Science 345:1250092. https://doi.org/10.1126/science.1250092

Mares D, Himi E (2021) The role of *TaMYB10-A1* of wheat (*Triticum aestivum* L.) in determining grain coat colour and dormancy phenotype. Euphytica 217:89.

https://doi.org/10.1007/s10681-021-02826-8

Mares D, Mrva K (2008) Late-maturity α-amylase: Low falling number in wheat in the absence of pre-harvest sprouting. J Cereal Sci 47:6–17. https://doi.org/10.1016/j.jcs.2007.01.005

Mares D, Mrva K, Cheong J, et al (2005) A QTL located on chromosome 4A associated with dormancy in white- and red-grained wheats of diverse origin. Theor Appl Genet 111:1357–1364. https://doi.org/10.1007/s00122-005-0065-5

Mares DJ (1993) Pre-harvest sprouting in wheat. I. Influence of cultivar, rainfall and temperature during grain ripening. Aust J Agric Res 44:1259–1272.

Mares DJ (1996) Dormancy in white wheat: mechanism and location of genes. In: Noda K, Mares DJ (ed) Preharvest sprouting in cereals 1995. Centre for Academic Societies, Osaka, pp 179–184.

Mares DJ, Mrva K (2001) Mapping quantitative trait loci associated with variation in grain dormancy in Australian wheat. Aust J Agric Res 52:1257-1265. https://doi.org/10.1071/AR01049

Mares DJ, Mrva K, Cheong J, et al (2021) Dormancy and dormancy release in white-grained wheat (*Triticum aestivum* L.). Planta 253:5. https://doi.org/10.1007/s00425-020-03518-8

Mariotti R, Fornasiero A, Mousavi S, et al (2020) Genetic mapping of the incompatibility locus in olive and development of a linked sequence-tagged site marker. Front. Plant Sci. 10:1760. https://doi.org/10.3389/fpls.2019.01760

Martinez SA, Godoy J, Huang M, et al (2018) Genome-wide association mapping for tolerance to pre-harvest sprouting and low falling numbers in wheat. Front. Plant Sci. 9:141. https://doi.org/10.3389/fpls.2018.00141

Martinez SA, Shorinola O, Conselman S, et al (2020) Exome sequencing of bulked segregants identified a novel *TaMKK3-A* allele linked to the wheat *ERA8* ABA-hypersensitive germination phenotype. Theor Appl Genet 133:719–736. https://doi.org/10.1007/s00122-019-03503-0

Martinez SA, Tuttle KM, Takebayashi Y, et al (2016) The wheat ABA hypersensitive *ERA8* mutant is associated with increased pre-harvest sprouting tolerance and altered hormone accumulation. Euphytica 212:229–245. https://doi.org/10.1007/s10681-016-1763-6

Matsuoka Y (2011) Evolution of polyploid *Triticum* wheats under cultivation: The role of domestication, natural hybridization and allopolyploid speciation in their diversification. Plant Cell Physiol 52:750–764. https://doi.org/10.1093/pcp/pcr018

Matus-Cádiz MA, Hucl P (2004) Maximizing pre-harvest sprouting in white-and red-seeded wheat. In: Soils and crops workshop. pp. 1–4. http://hdl.handle.net/10388/9621

McCaig TN, Gan YT, Clarke P, et al (2006) Kernel colour changes associated with field weathering of spring wheat. Can J Plant Sci 86:371–377. https://doi.org/10.4141/P05-033

Mccallum JA, Walker JRL (1990) Proanthocyanidins in wheat bran. Cereal Chem 67:282–285.

McCarty DR, Hattori T, Carson CB, et al (1991) The *Viviparous-1* developmental gene of maize encodes a novel transcriptional activator. Cell 66:895–905. https://doi.org/10.1016/0092-8674(91)90436-3

McKenna A, Hanna M, Banks E, et al (2010) The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303. https://doi.org/10.1101/gr.107524.110

McKibbin RS, Wilkinson MD, Bailey PC, et al (2002) Transcripts of *Vp-1* homeologues are misspliced in modern wheat and ancestral species. Proc Natl Acad Sci 99:10203–10208. https://doi.org/10.1073/pnas.152318599

McMaster GJ, Derera NF (1976) Methodology and sample preparation when screening for sprouting damage in cereals. Cereal Res Commun 4:251–254.

Meimoun P, Mordret E, Langlade NB, et al (2014) Is gene transcription involved in seed dry after-ripening? PLoS One 9:e86442. https://doi.org/10.1371/journal.pone.0086442

Melini F, Carcea M (2016) Grains for feeding the world: The ICC/AISTEC conference on the occasion of the World EXPO Milan 2015. Quality Assurance and Safety of Crops & Foods, 8(1), pp.3-9.

Mendiondo GM, Leymarie J, Farrant JM, et al (2010) Differential expression of abscisic acid metabolism and signalling genes induced by seed-covering structures or hypoxia in barley (*Hordeum vulgare* L.) grains. Seed Sci Res 20:69–77. https://doi.org/10.1017/S0960258509990262

Meng L, Li H, Zhang L, et al (2015) QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in bi-parental populations. Crop J 3:269-283. http://dx.doi.org/10.1016/j.cj.2015.01.001 2214-5141

Merchante C, Alonso JM, Stepanova AN (2013) Ethylene signaling: simple ligand, complex regulation. Curr Opin Plant Biol 16:554–560. https://doi.org/10.1016/j.pbi.2013.08.001

Mi H, Ebert D, Muruganujan A, et al (2021) PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Res 49:394–403. https://doi.org/10.1093/nar/gkaa1106

Miles C, Wayne M (2008) Quantitative trait locus (QTL) analysis. Nature Education 1:208.

Millar AA, Jacobsen J V, Ross JJ, et al (2006) Seed dormancy and ABA metabolism in Arabidopsis and barley: the role of ABA 8′-hydroxylase. Plant J 45:942–954. https://doi.org/10.1111/j.1365-313X.2006.02659.x

Mingam A, Toffano-Nioche C, Brunaud V, et al (2004) DEAD-box RNA helicases in *Arabidopsis thaliana*: establishing a link between quantitative expression, gene structure and evolution of a family of genes. Plant Biotechnol J 2:401–415. https://doi.org/10.1111/j.1467-7652.2004.00084.x

Miyamoto T, Everson EH (1958) Biochemical and physiological studies of wheat seed pigmentation. Agron J 50:733–734. https://doi.org/10.2134/agronj1958.00021962005000120005x

Mohan A, Kulwal P, Singh R, et al (2009) Genome-wide QTL analysis for pre-harvest sprouting tolerance in bread wheat. Euphytica 168:319-329. https://doi.org/10.1007/s10681-009-9935-2

Molina I, Ohlrogge JB, Pollard M (2008) Deposition and localization of lipid polyester in developing seeds of *Brassica napus* and *Arabidopsis thaliana*. Plant J 53:437–449. https://doi.org/10.1111/j.1365-313X.2007.03348.x

Moose SP, Mumm RH (2008) Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiol 147:969–977. https://doi.org/10.1104/pp.108.118232

Mori M, Uchino N, Chono M, et al (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. https://doi.org/10.1007/s00122-005-1972-1

Mortazavi A, Williams BA, McCue K, et al (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5:621–628. https://doi.org/10.1038/nmeth.1226

Mourad AMI, Belamkar V, Baenziger PS (2020) Molecular genetic analysis of spring wheat core collection using genetic diversity, population structure, and linkage disequilibrium. BMC Genomics 21:434. https://doi.org/10.1186/s12864-020-06835-0

Murphy LD, Herzog CE, Rudick JB, et al (1990) Use of the polymerase chain reaction in the quantitation of *mdr-1* gene expression. Biochemistry 29:10351–10356.

Nakamura S (2018) Grain dormancy genes responsible for preventing pre-harvest sprouting in barley and wheat. Breed Sci 68:295–304. https://doi.org/10.1270/jsbbs.17138

Nakamura S, Abe F, Kawahigashi H, et al (2011) A wheat homolog of *MOTHER OF FT AND TFL1* acts in the regulation of germination. Plant Cell 23:3215–3229. https://doi.org/10.1105/tpc.111.088492

Nakamura S, Toyama T (2001) Isolation of a *VP1* homologue from wheat and analysis of its expression in embryos of dormant and non-dormant cultivars. J Exp Bot 52:875–876. https://doi.org/10.1093/jexbot/52.357.875

Nakashima K, Fujita Y, Kanamori N, et al (2009) Three Arabidopsis SnRK2 Protein Kinases, SRK2D/SnRK2.2, SRK2E/SnRK2.6/OST1 and SRK2I/SnRK2.3, involved in ABA signaling are essential for the control of seed development and dormancy. Plant Cell Physiol 50:1345–1363. https://doi.org/10.1093/pcp/pcp083

Nambara E, Marion-Poll A (2003) ABA action and interactions in seeds. Trends Plant Sci 8:213–217. https://doi.org/10.1016/S1360-1385(03)00060-8

Nambara E, Marion-Poll A (2005) Abscisic acid biosynthesis and catabolism. Annu Rev Plant Biol 56:165–185. https://doi.org/10.1146/annurev.arplant.56.032604.144046

Nambara E, Okamoto M, Tatematsu K, et al (2010) Abscisic acid and the control of seed dormancy and germination. Seed Sci Res 20:55-67. https://doi.org/10.1017/S0960258510000012

Née G, Xiang Y, Soppe WJJ (2017) The release of dormancy, a wake-up call for seeds to germinate. Curr Opin Plant Biol 35:8–14. https://doi.org/10.1016/j.pbi.2016.09.002

Nelson JC (1997) QGENE: software for marker-based genomic analysis and breeding. Mol Breed 3:239–245. https://doi.org/10.1023/A:1009604312050

Nelson SK, Steber CM (2017) Transcriptional mechanisms associated with seed dormancy and dormancy loss in the gibberellin-insensitive *sly1-2* mutant of *Arabidopsis thaliana*. PLoS One 12:e0179143. https://doi.org/10.1371/journal.pone.0179143

Ng LM, Melcher K, Teh BT, et al (2014) Abscisic acid perception and signaling: structural mechanisms and applications. Acta Pharmacol Sin 35:567–584. https://doi.org/10.1038/aps.2014.5

Nikolaeva MG (2004) On criteria to use in studies of seed evolution. Seed Sci Res 14:315–320. https://doi.org/10.1079/SSR2004185

Nishimura N, Tsuchiya W, Moresco JJ, et al (2018) Control of seed dormancy and germination by *DOG1-AHG1* PP2C phosphatase complex via binding to heme. Nat Commun 9:2132. https://doi.org/10.1038/s41467-018-04437-9

Noda T, Ichinose Y, Takigawa S, et al (2003) The pasting properties of flour and starch in wheat grain damaged by alpha-amylase. Food Sci Technol Res 9:387–391. https://doi.org/10.3136/fstr.9.387

Nonogaki H (2019) Seed germination and dormancy: The classic story, new puzzles, and evolution. J Integr Plant Biol 61:541–563. https://doi.org/10.1111/jipb.12762

Nyachiro JM, Clarke FR, DePauw RM, et al (2002) Temperature effects on seed germination and expression of seed dormancy in wheat. Euphytica 126:123–127. https://doi.org/10.1023/A:1019694800066

O'Connor CM, Adams JU, Fairman J (2010) Essentials of cell biology. Cambridge, MA NPG Educ 1:54.

Official Grain Grading Guide. https://www.grainscanada.gc.ca/en/grain-quality/official-grain-grading-guide/. Accessed: November 5, 2021

Ogbonnaya FC, Imtiaz M, DePauw RM (2007) Haplotype diversity of pre-harvest sprouting QTLs in wheat. Genome 50:107–118. https://doi.org/10.1139/g06-142

Ogbonnaya FC, Imtiaz M, Ye G, et al (2008) Genetic and QTL analyses of seed dormancy and preharvest sprouting resistance in the wheat germplasm CN10955. Theor Appl Genet 116:891-902. https://doi.org/10.1007/s00122-008-0712-8

Okamoto M, Kuwahara A, Seo M, et al (2006) *CYP707A1* and *CYP707A2*, which encode

abscisic acid 8′-Hydroxylases, are indispensable for proper control of seed dormancy and germination in Arabidopsis. Plant Physiol 141:97–107. https://doi.org/10.1104/pp.106.079475

Olaerts H, Courtin CM (2018) Impact of pre-harvest sprouting on endogenous hydrolases and technological quality of wheat and bread: A review. Compr Rev Food Sci Food Saf 17:698–713. https://doi.org/10.1111/1541-4337.12347

Olaerts H, De Bondt Y, Courtin CM (2017) The heterogeneous distribution of α-amylase and endoxylanase activity over a population of pre-harvest sprouted wheat kernels and their localization in individual kernels. J Cereal Sci 74:200–209. https://doi.org/10.1016/j.jcs.2017.02.010

Olaerts H, Roye C, Derde LJ, et al (2016) Impact of pre-harvest sprouting of wheat (*Triticum aestivum*) in the field on starch, protein, and arabinoxylan properties. J Agric Food Chem 64:8324–8332. https://doi.org/10.1021/acs.jafc.6b03140

Olaerts H, Vandekerckhove L, Courtin CM (2018) A closer look at the bread making process and the quality of bread as a function of the degree of pre-harvest sprouting of wheat (*Triticum aestivum*). J Cereal Sci 80:188–197. https://doi.org/10.1016/j.jcs.2018.03.004

Olmstead AL, Rhode PW (2011) Adapting North American wheat production to climatic challenges, 1839–2009. Proc Natl Acad Sci 108:480–485. https://doi.org/10.1073/pnas.1008279108

Olson M, Leroy H, Charles C, David B (1989) A common language for physical mapping of the human genome. Science 245:1434–1435. https://doi.org/10.1126/science.2781285

Omoarelojie LO, Kulkarni MG, Finnie JF, et al (2021) Smoke-derived cues in the regulation of seed germination: are Ca2+-dependent signals involved. Plant Growth Regul 97:343–355. https://doi.org/10.1007/s10725-021-00745-1

Onishi K, Yamane M, Yamaji N, et al (2017) Sequence differences in the seed dormancy gene *Qsd1* among various wheat genomes. BMC Genomics 18:497. https://doi.org/10.1186/s12864-017-3880-6

Osa M, Kato K, Mori M, et al (2003) Mapping QTLs for seed dormancy and the *Vp1* homologue on chromosome 3A in wheat. Theor Appl Genet 106:1491–1496. https://doi.org/10.1007/s00122-003-1208-1

Osanai SI, Amano Y, Mares D (2005) Development of highly sprouting tolerant wheat germplasm with reduced germination at low temperature. Euphytica 143:301–307. https://doi.org/10.1007/s10681-005-7887-8

Park M, Choi W, Shin SY, et al (2021) Identification of genes and microRNAs affecting pre-harvest sprouting in rice (*Oryza sativa* L.) by transcriptome and small RNAome Analyses. Front. Plant Sci. 12:727302. https://doi.org/10.3389/fpls.2021.727302

Paterson AH (2002) What has QTL mapping taught us about plant domestication? New Phytol 154:591–608. https://doi.org/10.1046/j.1469-8137.2002.00420.x

Paterson AH, Sorrells ME, Obendorf RL (1989) Methods of evaluation for pre-harvest sprouting resistance in wheat breeding programs. Can J Plant Sci 69:681–689. https://doi.org/10.4141/cjps89-084

Penfield S (2017) Seed dormancy and germination. Curr Biol 27:874–878. https://doi.org/10.1016/j.cub.2017.05.050

Perten H (1964) Application of the falling number method for evaluating alpha-amylase activity. Cereal Chem 41:127–140.

Peterson RF (1965) Origin and history of the wheat species. Wheat: Botany, cultivation, and utilization. Interscience Publishers Inc., New York, pp.88-102.

Pfeifer M, Kugler KG, Sandve SR, et al (2014) Genome interplay in the grain transcriptome of hexaploid bread wheat. Science 345:1250091. https://doi.org/10.1126/science.1250091

Phillips C, Ballard D, Gill P, et al (2012) The recombination landscape around forensic STRs: Accurate measurement of genetic distances between syntenic STR pairs using HapMap high density SNP data. Forensic Sci Int Genet 6:354–365. https://doi.org/10.1016/j.fsigen.2011.07.012

Powell JJ, Carere J, Sablok G, et al (2017) Transcriptome analysis of *Brachypodium* during fungal pathogen infection reveals both shared and distinct defense responses with wheat. Sci Rep 7:17212. https://doi.org/10.1038/s41598-017-17454-3

Qiao L, Zhang W, Li X, et al (2018) Characterization and expression patterns of auxin response factors in wheat. Front. Plant Sci. 9:1395. https://doi.org/10.3389/fpls.2018.01395

Rahman S, Islam S, Yu Z, et al (2020) Current progress in understanding and recovering the wheat genes lost in evolution and domestication. Int. J. Mol. Sci. 21:5836. https://doi.org/10.3390/ijms21165836

Rajjou L, Duval M, Gallardo K, et al (2012) Seed germination and vigor. Annu Rev Plant Biol 63:507–533. https://doi.org/10.1146/annurev-arplant-042811-105550

Ramaih S, Guedira M, Paulsen GM (2003) Relationship of indoleacetic acid and tryptophan to dormancy and pre-harvest sprouting of wheat. Funct Plant Biol 30:939–945.

Ranathunge K, Shao S, Qutob D, et al (2010) Properties of the soybean seed coat cuticle change during development. Planta 231:1171–1188. https://doi.org/10.1007/s00425-010-1118-9

Rasheed A, Xia X (2019) From markers to genome-based breeding in wheat. Theor Appl Genet 132:767–784. https://doi.org/10.1007/s00122-019-03286-4

Rasul G, Humphreys DG, Brûlé-Babel A, et al (2009) Mapping QTLs for pre-harvest sprouting traits in the spring wheat cross 'RL4452/AC Domain.' Euphytica 168:363–378. https://doi.org/10.1007/s10681-009-9934-3

Rasul G, Humphreys GD, Wu J, et al (2012) Evaluation of pre-harvest sprouting traits in a collection of spring wheat germplasm using genotype and genotype×environment

interaction model. Plant Breed 131:244–251. https://doi.org/10.1111/j.1439-0523.2011.01931.x

Rathjen JR, Strounina E V, Mares DJ (2009) Water movement into dormant and non-dormant wheat (*Triticum aestivum* L.) grains. J Exp Bot 60:1619–1631. https://doi.org/10.1093/jxb/erp037

Rikiishi K, Maekawa M (2014) Seed maturation regulators are related to the control of seed dormancy in wheat (*Triticum aestivum* L.). PLoS One 9:e107618. https://doi.org/10.1371/journal.pone.0107618

Rikiishi K, Sugimoto M, Maekawa M (2021) Transcriptomic analysis of developing seeds in a wheat (*Triticum aestivum* L.) mutant RSD32 with reduced seed dormancy. Breed Sci 71:155–166. https://doi.org/10.1270/jsbbs.20016

Risch N (1992) Genetic linkage: interpreting lod scores. Science (80) 255:803–804. https://doi.org/10.1126/science.1536004

Röder MS, Korzun V, Wendehake K, et al (1998) A Microsatellite Map of Wheat. Genetics 149:2007–2023. https://doi.org/10.1093/genetics/149.4.2007

Rodríguez MV, Barrero JM, Corbineau F, et al (2015) Dormancy in cereals (not too much, not so little): about the mechanisms behind this trait. Seed Sci Res 25:99–119. https://doi.org/10.1017/S0960258515000021

Roy JK, Prasad M, Varshney RK, et al (1999) Identification of a microsatellite on chromosomes 6B and a STS on 7D of bread wheat showing an association with pre-harvest sprouting tolerance. Theor Appl Genet 99:336-340.

Rzewuski G, Sauter M (2008) Ethylene biosynthesis and signaling in rice. Plant Sci 175:32–42. https://doi.org/10.1016/j.plantsci.2008.01.012

Salse J, Chagué V, Bolot S, et al (2008) New insights into the origin of the B genome of hexaploid wheat: Evolutionary relationships at the SPA genomic region with the S genome of the diploid relative Aegilops speltoides. BMC Genomics 9:555. https://doi.org/10.1186/1471-2164-9-555

Sano N, Marion-Poll A (2021) ABA metabolism and homeostasis in seed dormancy and germination. Int. J. Mol. Sci. 22:5069. https://doi.org/10.3390/ijms22105069

Santos AG, Livingston DP, Jellen EN, et al (2006) A cytological marker associated with winterhardiness in oat. Crop Sci 46:203–208. https://doi.org/10.2135/cropsci2005.0152

Santos LT, Barth Pinto RJ, Franco FDA, et al (2010) Inheritance and potential use of grain color in the identification of genotypes resistant to pre-harvest sprouting in wheat. Crop Breed Appl Biotechnol 10:218–224. https://doi.org/10.1590/S1984-70332010000300006

Schwartz SH, Qin X, Zeevaart JAD (2003) Elucidation of the indirect pathway of abscisic acid biosynthesis by mutants, genes, and enzymes. Plant Physiol 131:1591–1601. https://doi.org/10.1104/pp.102.017921

Scott MF, Ladejobi O, Amer S, et al (2020) Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. Heredity 125:396–416. https://doi.org/10.1038/s41437-020-0336-6

Semagn K, Bjørnstad Å, Ndjiondjop MN (2006) An overview of molecular marker methods for plants. Afr J Biotechnol 5:2540-2568. https://doi.org/10.5897/AJB2006.000-5110

Semagn K, Bjørnstad Å, Skinnes H, et al (2006) Distribution of DArT, AFLP, and SSR markers in a genetic linkage map of a doubled-haploid hexaploid wheat population. Genome 49:545–555. https://doi.org/10.1139/g06-002

Shao M, Bai G, Rife TW, et al (2018) QTL mapping of pre-harvest sprouting resistance in a white wheat cultivar Danby. Theor Appl Genet 131:1683–1697. https://doi.org/10.1007/s00122-018-3107-5

Shewry PR (2009) Wheat. J Exp Bot 60:1537–1553. https://doi.org/10.1093/jxb/erp058

Shorinola O, Balcárková B, Hyles J, et al (2017) Haplotype analysis of the pre-harvest sprouting resistance locus *Phs-A1* reveals a causal role of *TaMKK3-A* in global germplasm. Front. Plant Sci. 8:1555. https://doi.org/10.3389/fpls.2017.01555

Shorinola O, Bird N, Simmonds J, et al (2016) The wheat *Phs-A1* pre-harvest sprouting resistance locus delays the rate of seed dormancy loss and maps 0.3 cM distal to the *PM19* genes in UK germplasm. J Exp Bot 67:4169–4178. https://doi.org/10.1093/jxb/erw194

Shu K, Liu X, Xie Q, He Z (2016) Two faces of one seed: Hormonal regulation of dormancy and germination. Mol Plant 9:34–45. https://doi.org/10.1016/j.molp.2015.08.010

Shu K, Zhou W, Yang W (2018) APETALA 2-domain-containing transcription factors: focusing on abscisic acid and gibberellins antagonism. New Phytol 217:977–983. https://doi.org/10.1111/nph.14880

Singh R, Matus-Cádiz M, Båga M, et al (2008) Comparison of different methods for phenotyping pre-harvest sprouting in white-grained wheat. Cereal Chem 85:238–242. https://doi.org/10.1094/CCHEM-85-2-0238

Skubacz A, Daszkowska-Golec A (2017) Seed dormancy: the complex process regulated by abscisic acid, gibberellins, and other phytohormones that makes seed germination work, In: A.D.E.M. El-Esawi (Ed.), Phytohormones: Signaling mechanisms and crosstalk in plant development and stress responses. Rijeka Croatia: In Tech. pp. 77–100. https://doi.org/10.5772/intechopen.68735

Söderman EM, Brocard IM, Lynch TJ, et al (2000) Regulation and function of the Arabidopsis *ABA-insensitive 4* gene in seed and abscisic acid response signaling networks. Plant Physiol 124:1752–1765. https://doi.org/10.1104/pp.124.4.1752

Sohindji FS, Sogbohossou DEO, Zohoungbogbo HPF, et al (2020) Understanding molecular mechanisms of seed dormancy for improved germination in traditional leafy vegetables: An overview. Agron. 10:57. https://doi.org/10.3390/agronomy10010057

Sohn SI, Pandian S, Kumar TS, et al (2021) Seed dormancy and pre-harvest sprouting in rice—An updated overview. Int. J. Mol. Sci. 22:11804. https://doi.org/10.3390/ijms222111804

Somyong S, Ishikawa G, Munkvold JD, et al (2014) Fine mapping of a pre-harvest sprouting QTL interval on chromosome 2B in white wheat. Theor Appl Genet 127:1843–1855. https://doi.org/10.1007/s00122-014-2345-4

Son S, Chitnis VR, Liu A, et al (2016) Abscisic acid metabolic genes of wheat (*Triticum aestivum* L.): Identification and insights into their functionality in seed dormancy and dehydration tolerance. Planta 244:429–447. https://doi.org/10.1007/s00425-016-2518-2

Song Q, Cheng S, Chen Z, et al (2019) Comparative transcriptome analysis revealing the potential mechanism of seed germination stimulated by exogenous gibberellin in *Fraxinus hupehensis*. BMC Plant Biol 19:199. https://doi.org/10.1186/s12870-019-1801-3

Sorrells ME, Anderson JA (1996) Quantitative trait loci associated with pre-harvest sprouting in white wheat. In: Proceedings of the 1995 Pre-harvest Sprouting Symp. in Japan. Osaka, Japan: Center for Academic Societies, Japan, pp 137-142.

SPSS Inc. (2007) SPSS for windows, Version 16.0. Chicago, SPSS Inc.

Stanley L, Yuan YW (2019) Transcriptional regulation of carotenoid biosynthesis in plants: So many regulators, so little consensus. Front. Plant Sci. 10:1017. https://doi.org/10.3389/fpls.2019.01017

Stepanova AN, Alonso JM (2009) Ethylene signaling and response: where different regulatory modules meet. Curr Opin Plant Biol 12:548–555. https://doi.org/10.1016/j.pbi.2009.07.009

Stone JM, Collinge MA, Smith RD, et al (1994) Interaction of a protein phosphatase with an Arabidopsis serine-threonine receptor kinase. Science 266:793–795. https://doi.org/10.1126/science.7973632

Su Q, Li J (2018) QTL detection for kernel size and weight in bread wheat (*Triticum aestivum* L.) using a high-density SNP and SSR-based linkage map. Front. Plant Sci. 9:1484. https://doi.org/10.3389/fpls.2018.01484

Subbiah V, Reddy KJ (2010) Interactions between ethylene, abscisic acid and cytokinin during germination and seedling establishment in Arabidopsis. J Biosci 35:451–458. https://doi.org/10.1007/s12038-010-0050-2

Sugimoto K, Takeuchi Y, Ebana K, et al (2010) Molecular cloning of *Sdr4*, a regulator involved in seed dormancy and domestication of rice. Proc Natl Acad Sci 107:5792–5797. https://doi.org/10.1073/pnas.0911965107

Sun C, Dong Z, Zhao L, et al (2020) The Wheat 660K SNP array demonstrates great potential for marker-assisted selection in polyploid wheat. Plant Biotechnol J 18:1354–1360. https://doi.org/10.1111/pbi.13361

Sun M, Tuan PA, Izydorczyk MS, et al (2020) Ethylene regulates post-germination seedling growth in wheat through spatial and temporal modulation of ABA/GA balance. J Exp Bot

71:1985–2004. https://doi.org/10.1093/jxb/erz566

Sun T, Gubler F (2004) Molecular mechanism of gibberellin signaling in plants. Annu Rev Plant Biol 55:197–223. https://doi.org/10.1146/annurev.arplant.55.031903.141753

Sura W, Kabza M, Karlowski WM, et al (2017) Dual role of the histone variant H2A.Z in transcriptional regulation of stress-response genes. Plant Cell 29:791–807. https://doi.org/10.1105/tpc.16.00573

Szklarczyk D, Gable AL, Lyon D, et al (2019) STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 47:607–613. https://doi.org/10.1093/nar/gky1131

Tan MK, Sharp PJ, Lu MQ, et al (2006) Genetics of grain dormancy in a white wheat. Aust J Agric Res 57:1157–1165. https://doi.org/10.1071/AR06101

Tanaka T, Ishikawa G, Ogiso-Tanaka E, et al (2019) Development of genome-wide SNP markers for barley via reference-based RNA-Seq analysis. Front. Plant Sci. 10:577. https://doi.org/10.3389/fpls.2019.00577

Tanksley SD (1993) Mapping polygenes. Annu Rev Genet 27:205–233. https://doi.org/10.1146/annurev.ge.27.120193.001225

The Gene Ontology Consortium (2021) The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res 49:325–334. https://doi.org/10.1093/nar/gkaa1113

Thomason WE, Hughes KR, Griffey CA, et al (2019) "Understanding pre-harvest sprouting of wheat," in Virginia Cooperative Extension, (Blacksburg, VA: Virginia Tech). pp. 1–4

Thudi M, Li Y, Jackson SA, et al (2012) Current state-of-art of sequencing technologies for plant genomics research. Brief Funct Genomics 11:3–11. https://doi.org/10.1093/bfgp/elr045

Tian F, Yang DC, Meng YQ, et al (2020) PlantRegMap: charting functional regulatory maps in plants. Nucleic Acids Res 48:1104–1113. https://doi.org/10.1093/nar/gkz1020

Toh S, Imamura A, Watanabe A, et al (2008) High temperature-induced abscisic acid biosynthesis and its role in the inhibition of gibberellin action in Arabidopsis seeds. Plant Physiol 146:1368–1385. https://doi.org/10.1104/pp.107.113738

Torada A, Ikeguchi S, Koike M (2005) Mapping and validation of PCR-based markers associated with a major QTL for seed dormancy in wheat. Euphytica, 143:251-255. https://doi.org/10.1007/s10681-005-7872-2

Torada A, Koike M, Mochida K, et al (2006) SSR-based linkage map with new markers using an intraspecific population of common wheat. Theor Appl Genet 112:1042–1051. https://doi.org/10.1007/s00122-006-0206-5

Torada A, Koike M, Ogawa T, et al (2016) A causal gene for seed dormancy on wheat chromosome 4A encodes a MAP kinase kinase. Curr Biol 26:782–787. https://doi.org/10.1016/j.cub.2016.01.063

Townley-Smith TF, Czarnecki EM (2008) AC Domain hard red spring wheat. Can J Plant Sci 88:347-350. https://doi.org/10.4141/CJPS07004

Trapnell C, Williams BA, Pertea G, et al (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511–515. https://doi.org/10.1038/nbt.1621

Trethowan RM (1995) Evaluation and selection of bread wheat (*Triticum aestivum* L.) for pre-harvest sprouting tolerance. Aust J Agric Res 46:463–474. https://doi.org/10.1071/AR9950463

Trethowan RM, Rajaram S, Ellison FW (1996) Pre-harvest sprouting tolerance in wheat in the field and under rain simulation. Aust J Agric Res 47:705–716. https://doi.org/10.1071/AR9960705

Tuan PA, Kumar R, Rehal PK, et al (2018) Molecular mechanisms underlying abscisic acid/gibberellin balance in the control of seed dormancy and germination in cereals. Fron Plant Sci 9:668. https://doi.org/10.3389/fpls.2018.00668

Tuan PA, Nguyen TN, Jordan MC, et al (2021) A shift in abscisic acid/gibberellin balance underlies retention of dormancy induced by seed development temperature. Plant Cell Environ 44:2230–2244. https://doi.org/10.1111/pce.13963

Tuan PA, Yamasaki Y, Kanno Y, et al (2019) Transcriptomics of cytokinin and auxin metabolism and signaling genes during seed maturation in dormant and non-dormant wheat genotypes. Sci Rep 9:3983. https://doi.org/10.1038/s41598-019-40657-9

Tyagi S, Gupta PK (2012) Meta-analysis of QTLs involved in pre-harvest sprouting tolerance and dormancy in bread wheat. Triticeae Genomics Genet 3:9-24. https://doi.org/10.5376/tgg.2012.03.0002

Tyrka M, Oleszczuk S, Rabiza-Swider J, et al (2018) Populations of doubled haploids for genetic mapping in hexaploid winter triticale. Mol Breed 38:46. https://doi.org/10.1007/s11032-018-0804-3

Ueguchi-Tanaka M, Ashikari M, Nakajima M, et al (2005) *GIBBERELLIN INSENSITIVE DWARF1* encodes a soluble receptor for gibberellin. Nature 437:693–698. https://doi.org/10.1038/nature04028

van Ooijen JW (2004) MapQTL®5, software for the mapping of quantitative trait loci in experimental populations. Kyazma BV, Wageningen, The Netherlands.

van Ooijen JW, Voorrips RE (2001) Joinmap®3.0, software for the calculation of genetic linkage maps. Plant Research International, Wageningen, The Netherlands.

Venske E, dos Santos RS, Busanello C, et al (2019) Bread wheat: a role model for plant domestication and breeding. Hereditas 156:16. https://doi.org/10.1186/s41065-019-0093-9

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

https://doi.org/10.1016/j.plantsci.2019.01.004

Vos P, Hogers R, Bleeker M, et al (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414. https://doi.org/10.1093/nar/23.21.4407

Waiho K, Shi X, Fazhan H, et al (2019) High-density genetic linkage maps provide novel insights into ZW/ZZ sex determination system and growth performance in mud crab (*Scylla paramamosain*). Front. Genet. 10:298. https://doi.org/10.3389/fgene.2019.00298

Walker-Simmons M (1988) Enhancement of ABA responsiveness in wheat embryos by high temperature. Plant Cell Environ 11:769–775. https://doi.org/10.1111/j.1365-3040.1988.tb01161.x

Walkowiak S, Gao L, Monat C, et al (2020) Multiple wheat genomes reveal global variation in modern breeding. Nature 588:277–283. https://doi.org/10.1038/s41586-020-2961-x

Wan Y, Poole RL, Huttly AK, et al (2008) Transcriptome analysis of grain development in hexaploid wheat. BMC Genomics 9:121. https://doi.org/10.1186/1471-2164-9-121

Wang JK (2009) Inclusive composite interval mapping of quantitative trait genes. ACTA Agron Sin 35:239–245. https://doi.org/10.3724/sp.j.1006.2009.00239

Wang KLC, Li H, Ecker JR (2002) Ethylene biosynthesis and signaling networks. Plant Cell 14:131–151. https://doi.org/10.1105/tpc.001768

Wang S, Wong D, Forrest K, et al (2014) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnol J 12:787–796. https://doi.org/10.1111/pbi.12183

Wang X, Liu H, Liu G, et al (2019) Phenotypic and genotypic characterization of near-isogenic lines targeting a major 4BL QTL responsible for pre-harvest sprouting in wheat. BMC Plant Biol 19:348. https://doi.org/10.1186/s12870-019-1961-1

Wang X, Liu H, Siddique KHM, Yan G (2021) Transcriptomic profiling of wheat near-isogenic lines reveals candidate genes on chromosome 3A for pre-harvest sprouting resistance. BMC Plant Biol 21:53. https://doi.org/10.1186/s12870-021-02824-x

Wang Y, Wang XL, Meng JY, et al (2016) Characterization of *Tamyb10* allelic variants and development of STS marker for pre-harvest sprouting resistance in Chinese bread wheat. Mol Breeding, 36:148. https://doi.org/10.1007/s11032-016-0573-9

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63. https://doi.org/10.1038/nrg2484

Warner RL, Kudrna DA, Spaeth SC, et al (2000) Dormancy in white-grain mutants of Chinese spring wheat (*Triticum aestivum* L.). Seed Sci Res 10:51–60. https://doi.org/10.1017/S0960258500000064

Wei W, Min X, Shan S, et al (2019) Isolation and characterization of *TaQsd1* genes for period of dormancy in common wheat (*Triticum aestivum* L.). Mol Breed 39:150.

https://doi.org/10.1007/s11032-019-1060-x

Weide A (2015) On the Identification of domesticated emmer wheat, *Triticum turgidum* subsp. *dicoccum* (Poaceae), in the Aceramic Neolithic of the Fertile Crescent. Archäologische Informationen 38:381-424. https://doi.org/10.11588/ai.2015.1.26205

Weitbrecht K, Müller K, Leubner-Metzger G (2011) First off the mark: early seed germination. J Exp Bot 62:3289–3309. https://doi.org/10.1093/jxb/err030

Wen W, He Z, Gao F, et al (2017) A high-density consensus map of common wheat integrating four mapping populations scanned by the 90K SNP array. Front. Plant Sci. 8:1389. https://doi.org/10.3389/fpls.2017.0138

Wheat in the world. https://wheat.org. Accessed: November 01, 2021

Wijayanto H, Hirai Y, Kamanaka Y, et al (2005) Patterns of C-heterochromatin and telomeric DNA in two representative groups of small apes, the genera *Hylobates* and *Symphalangus*. Chromosom Res 13:715–722. https://doi.org/10.1007/s10577-005-1007-4

Willenbockel D (2014) Scenarios for global agriculture and food security towards 2050: A review of recent studies. In: New Challenges to Food Security, 1st edn, Routledge. pp. 55-76.

Williams JGK, Kubelik AR, Livak KJ, et al (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535. https://doi.org/10.1093/nar/18.22.6531

Winfield MO, Allen AM, Burridge AJ, et al (2016) High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. Plant Biotechnol J 14:1195–1206. https://doi.org/10.1111/pbi.12485

Wu Q, Bai X, Wu X, et al (2020) Transcriptome profiling identifies transcription factors and key homologs involved in seed dormancy and germination regulation of *Chenopodium quinoa*. Plant Physiol Biochem 151:443–456. https://doi.org/10.1016/j.plaphy.2020.03.050

Wu QH, Chen YX, Zhou SH, et al (2015) High-density genetic linkage map construction and QTL mapping of grain shape and size in the wheat population Yanda1817× Beinong6. PLoS ONE 10:e0118144. https://doi.org/10.1371/journal.pone.0118144

Xiao SH, Zhang XY, Yan CS, et al (2002) Germplasm improvement for pre-harvest sprouting resistance in Chinese white-grained wheat: An overview of the current strategy. Euphytica 126:35–38. https://doi.org/10.1023/A:1019679924173

Xiu-Jin L, Deng-Cai L, Zhi-Rong W (1997) Inheritance in synthetic hexaploid wheat 'RSP' of sprouting tolerance derived from *Aegilops tauschii* Cosson. Euphytica 95:321–323. https://doi.org/10.1023/A:1003078801358

Xu E, Vaahtera L, Horak H, et al (2015) Quantitative trait loci mapping and transcriptome analysis reveal candidate genes regulating the response to ozone in *Arabidopsis thaliana*. Plant Cell Environ 38:1418–1433. https://doi.org/10.1111/pce.12499

Xu S (1998) Mapping quantitative trait loci using multiple families of line crosses. Genetics 148:517–524. https://doi.org/10.1093/genetics/148.1.517

Yang Y, Zhang CL, Liu SX, et al (2014) Characterization of the rich haplotypes of *Viviparous-1A* in Chinese wheats and development of a novel sequence-tagged site marker for pre-harvest sprouting resistance. Mol Breed 33:75–88. https://doi.org/10.1007/s11032-013-9935-8

Yang Y, Zhao XL, Xia LQ, et al (2007) Development and validation of a *Viviparous-1* STS marker for pre-harvest sprouting tolerance in Chinese wheats. Theor Appl Genet 115:971-980. https://doi.org/10.1007/s00122-007-0624-z

Yiwen H, Xuran D, Hongwei L, et al (2022) Identification of effective alleles and haplotypes conferring pre-harvest sprouting resistance in winter wheat cultivars. BMC Plant Biol 22:326. https://doi.org/10.1186/s12870-022-03710-w

Yoo SD, Cho Y, Sheen J (2009) Emerging connections in the ethylene signaling network. Trends Plant Sci 14:270–279. https://doi.org/10.1016/j.tplants.2009.02.007

Young MD, Wakefield MJ, Smyth GK, et al (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol 11:R14. https://doi.org/10.1186/gb-2010-11-2-r14

Yu Y, Zhen S, Wang S, et al (2016) Comparative transcriptome analysis of wheat embryo and endosperm responses to ABA and H_2O_2 stresses during seed germination. BMC Genomics 17:97. https://doi.org/10.1186/s12864-016-2416-9

Zanetti S, Winzeler M, Keller M, et al (2000) Genetic analysis of pre-harvest sprouting resistance in a wheat x spelt cross. Crop Sci 40:1406-1417.

Zhang C, Luo W, Li Y, et al (2019) Transcriptomic analysis of seed germination under salt stress in two desert sister species (*Populus euphratica* and *P. pruinosa*). Front. Genet 10:231. https://doi.org/10.3389/fgene.2019.00231

Zhang L, Li H, Wang J (2012) The Statistical power of inclusive composite interval mapping in detecting digenic epistasis showing common F_2 segregation ratios. J Integr Plant Biol 54:270–279. https://doi.org/10.1111/j.1744-7909.2012.01110.x

Zhang LY, Bai MY, Wu J, et al (2009) Antagonistic HLH/bHLH transcription factors mediate brassinosteroid regulation of cell elongation and plant development in rice and Arabidopsis. Plant Cell 21:3767–3780. https://doi.org/10.1105/tpc.109.070441

Zhang SJ, Song GQ, Li YL, et al (2014) Cloning of *9-cis-epoxycarotenoid dioxygenase* gene (*TaNCED1*) from wheat and its heterologous expression in tobacco. Biol Plant 58:89–98. https://doi.org/10.1007/s10535-013-0373-6

Zhang XQ, Li C, Tay A, et al (2008) A new PCR-based marker on chromosome 4AL for resistance to pre-harvest sprouting in wheat (*Triticum aestivum* L.). Mol Breed 22:227–236. https://doi.org/10.1007/s11032-008-9169-3

Zhang Y, Miao X, Xia X, et al (2014a) Cloning of seed dormancy genes (*TaSdr*) associated with tolerance to pre-harvest sprouting in common wheat and development of a functional marker. Theor Appl Genet 127:855-866. https://doi.org/10.1007/s00122-014-2262-6

Zhang YC, Liao JY, Li ZY, et al (2014b) Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. Genome Biol 15:512. https://doi.org/10.1186/s13059-014-0512-1

Zhang Y, Xia X, He Z (2017) The seed dormancy allele *TaSdr-A1*a associated with pre-harvest sprouting tolerance is mainly present in Chinese wheat landraces. Theor Appl Genet. 130:81-89 https://doi.org/10.1007/s00122-016-2793-0

Zhao Y, Li MC, Konaté MM, et al (2021) TPM, FPKM, or Normalized Counts: A comparative study of quantification measures for the analysis of RNA-seq data from the NCI patient-derived models repository. J Transl Med 19:269. https://doi.org/10.1186/s12967-021-02936-w

Zi J, Mafu S, Peters RJ (2014) To gibberellins and beyond surveying the evolution of (Di)terpenoid metabolism. Annu Rev Plant Biol 65:259–286. https://doi.org/10.1146/annurev-arplant-050213-035705

Zobaer A, Alam MJ, Hasan MN, et al (2019) A Comparison on some interval mapping approaches for QTL detection. Bioinformation 15:90–94. https://doi.org/10.6026/97320630015090

2020 Crop Summary. https://canadiancereals.ca/2020-wheat-crop/. Accessed: November 4, 2021

APPENDIX

Appendix 1 Primer sequences used for qRT-PCR analysis

Gene	Type	Sequence (5'—3')
TaARF3	F	TGGTTCAGAAGTGATCGATAGTGG
	R	AACCTTTTCAGCCAGAGAAATGC
TaCYCD1	F	GCTAGCTGTGACCTGCCTCT
	R	TGAGGACGAGAAGCTCCATT

F, forward primer; R, reverse primer

Appendix 2 Details of the RNA-seq data from all dry and imbibed dormant and non-dormant seed samples

Sample	Raw reads	Clean reads	Raw bases	Clean bases	Error rate (%)	Q20 (%)	Q30 (%)	GC content (%)
D_0_1	49922212	47934238	7.5G	7.2G	0.03	97.45	93.43	55.31
D_0_2	48775712	46747708	7.3G	7.0G	0.03	97.39	93.34	55.40
D_0_3	42218166	40349098	6.3G	6.1G	0.03	97.50	93.60	55.28
D_0_4	53300716	50982082	8.0G	7.6G	0.03	97.39	93.33	55.30
ND_0_1	45369500	43460908	6.8G	6.5G	0.03	97.45	93.48	54.94
ND_0_2	46031320	43911592	6.9G	6.6G	0.03	97.38	93.37	55.27
ND_0_3	45587152	43684032	6.8G	6.6G	0.03	97.31	93.24	55.42
ND_0_4	49580878	46884298	7.4G	7.0G	0.03	97.60	93.81	54.87
D_12_1	48271806	46841150	7.2G	7.0G	0.03	97.50	93.26	53.83
D_12_2	49213910	48284932	7.4G	7.2G	0.03	97.39	93.04	53.49
D_12_3	52523878	51383146	7.9G	7.7G	0.03	97.23	92.74	53.79
D_12_4	49001576	47460252	7.4G	7.1G	0.03	97.49	93.28	53.77
ND_12_1	50063060	48922296	7.5G	7.3G	0.03	97.43	93.13	53.78
ND_12_2	49283956	47823270	7.4G	7.2G	0.03	97.36	93.05	53.99
ND_12_3	45117888	44179382	6.8G	6.6G	0.03	97.29	92.89	53.99
ND_12_4	52791100	51521526	7.9G	7.7G	0.03	97.36	92.96	54.00
D_24_1	47742474	46607968	7.2G	7.0G	0.03	97.22	92.77	54.07
D_24_2	44084592	42847854	6.6G	6.4G	0.03	97.41	93.16	54.04
D_24_3	50801792	49244370	7.6G	7.4G	0.03	97.35	93.02	54.12
D_24_4	50130744	48830288	7.5G	7.3G	0.03	97.15	92.62	53.89
ND_24_1	50032546	48815178	7.5G	7.3G	0.03	97.42	93.16	54.09
ND_24_2	45233506	44004732	6.8G	6.6G	0.03	97.42	93.17	53.83
ND_24_3	42928038	41765708	6.4G	6.3G	0.03	97.37	93.09	54.01
ND_24_4	47319880	46028090	7.1G	6.9G	0.03	97.51	93.34	54.23
Total	1155326402	1118514098	173.2G	167.6G	-	-	-	-

D, dormant; ND, non-dormant; 0, 12 and 24 represent 0, 12 and 24 HAI, respectively; numbers 1, 2, 3 and 4 represent replication number.

Appendix 3 Overview of mapping of the transcriptomes of dry and imbibed seeds of dormant and non-dormant DH lines to the reference genome lines

Sample	Total reads	Total mapped	Multiply mapped	Uniquely mapped
D_0_1	47934238	44296388 (92.41%)	2565488 (5.35%)	41730900 (87.06%)
D_0_2	46747708	43024617 (92.04%)	2501336 (5.35%)	40523281 (86.69%)
D_0_3	40349098	36976936 (91.64%)	2221783 (5.51%)	34755153 (86.14%)
D_0_4	50982082	46844345 (91.88%)	2761202 (5.42%)	44083143 (86.47%)
ND_0_1	43460908	38895018 (89.49%)	2348637 (5.4%)	36546381 (84.09%)
ND_0_2	43911592	40152772 (91.44%)	2442113 (5.56%)	37710659 (85.88%)
ND_0_3	43684032	40011727 (91.59%)	2197947 (5.03%)	37813780 (86.56%)
ND_0_4	46884298	43081148 (91.89%)	2543875 (5.43%)	40537273 (86.46%)
D_12_1	46841150	41598909 (88.81%)	2931174 (6.26%)	38667735 (82.55%)
D_12_2	48284932	43359849 (89.8%)	2734085 (5.66%)	40625764 (84.14%)
D_12_3	51383146	45450431 (88.45%)	3136560 (6.1%)	42313871 (82.35%)
D_12_4	47460252	44737538 (94.26%)	2456677 (5.18%)	42280861 (89.09%)
ND_12_1	48922296	43498865 (88.91%)	2819435 (5.76%)	40679430 (83.15%)
ND_12_2	47823270	42630507 (89.14%)	2927975 (6.12%)	39702532 (83.02%)
ND_12_3	44179382	38818315 (87.87%)	2570044 (5.82%)	36248271 (82.05%)
ND_12_4	51521526	46062733 (89.4%)	2913379 (5.65%)	43149354 (83.75%)
D_24_1	46607968	43661952 (93.68%)	2197201 (4.71%)	41464751 (88.96%)
D_24_2	42847854	38829675 (90.62%)	2455846 (5.73%)	36373829 (84.89%)
D_24_3	49244370	44226317 (89.81%)	3030190 (6.15%)	41196127 (83.66%)
D_24_4	48830288	43259218 (88.59%)	2912348 (5.96%)	40346870 (82.63%)
ND_24_1	48815178	44014963 (90.17%)	2757239 (5.65%)	41257724 (84.52%)
ND_24_2	44004732	39107251 (88.87%)	2628785 (5.97%)	36478466 (82.9%)
ND_24_3	41765708	37511887 (89.82%)	2410112 (5.77%)	35101775 (84.04%)
ND_24_4	46028090	42107044 (91.48%)	2548878 (5.54%)	39558166 (85.94%)

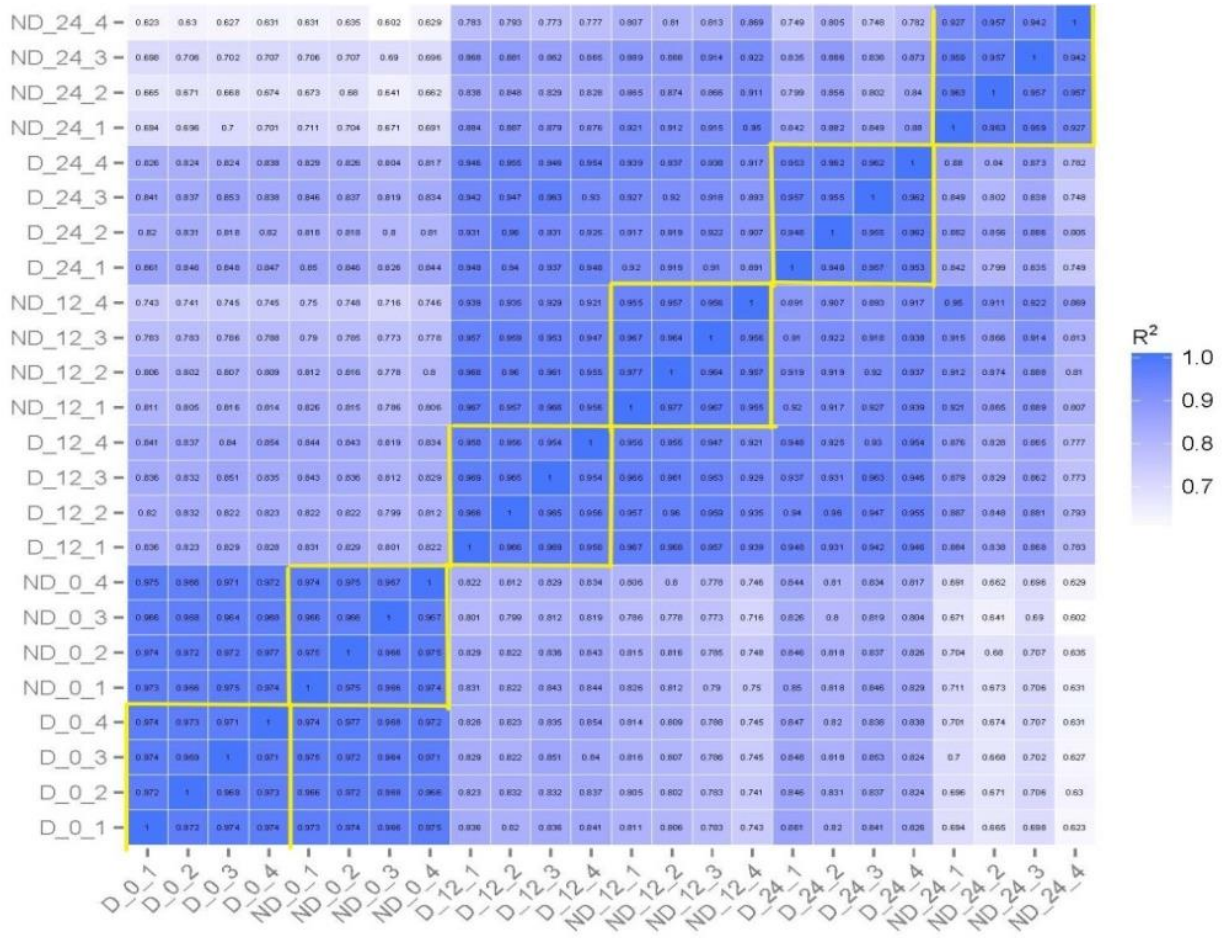
Sample descriptions are as shown in Table S2

Appendix 4 Number of genes expressed in each sample of the dormant and non-dormant seeds

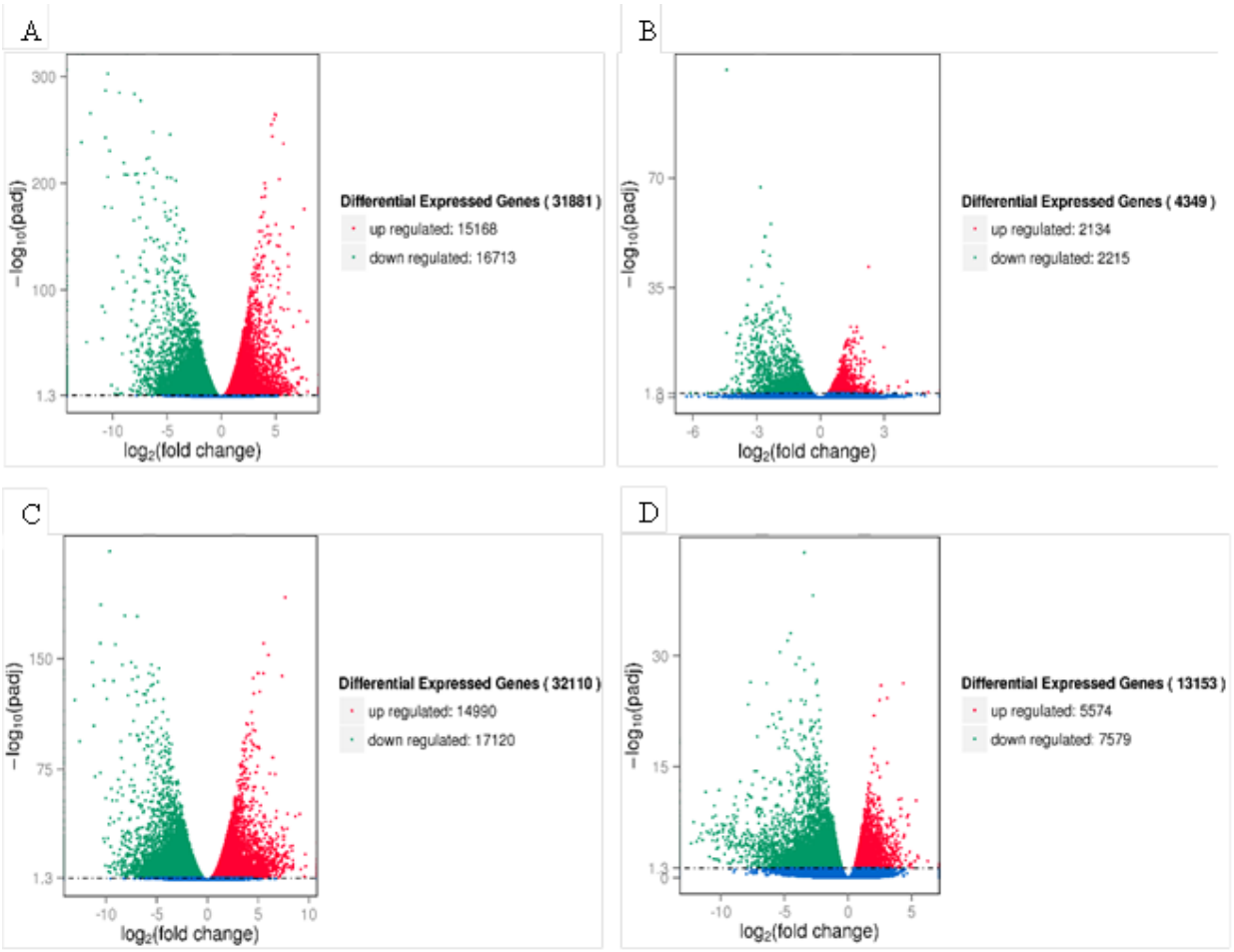
	FPKM Interval				
Sample *	0~1	1~3	3~15	15~60	>60
D_0_1	84352(69.86%)	11120(9.21%)	16633(13.78%)	6552(5.43%)	2087(1.73%)
D_0_2	83878(69.47%)	11442(9.48%)	16732(13.86%)	6588(5.46%)	2104(1.74%)
D_0_3	84406(69.90%)	11087(9.18%)	16507(13.67%)	6654(5.51%)	2090(1.73%)
D_0_4	84214(69.75%)	11136(9.22%)	16584(13.73%)	6666(5.52%)	2144(1.78%)
ND_0_1	84350(69.86%)	11020(9.13%)	16530(13.69%)	6766(5.60%)	2078(1.72%)
ND_0_2	84230(69.76%)	11455(9.49%)	16402(13.58%)	6549(5.42%)	2108(1.75%)
ND_0_3	85281(70.63%)	11244(9.31%)	15718(13.02%)	6317(5.23%)	2184(1.81%)
ND_0_4	84481(69.97%)	11259(9.32%)	16297(13.50%)	6578(5.45%)	2129(1.76%)
D_12_1	78134(64.71%)	12483(10.34%)	21522(17.82%)	6933(5.74%)	1672(1.38%)
D_12_2	77345(64.06%)	12482(10.34%)	22180(18.37%)	7060(5.85%)	1677(1.39%)
D_12_3	78085(64.67%)	12387(10.26%)	21400(17.72%)	7148(5.92%)	1724(1.43%)
D_12_4	78309(64.86%)	12326(10.21%)	21238(17.59%)	7161(5.93%)	1710(1.42%)
ND_12_1	78304(64.85%)	12251(10.15%)	21289(17.63%)	7184(5.95%)	1716(1.42%)
ND_12_2	78056(64.65%)	12410(10.28%)	21391(17.72%)	7153(5.92%)	1734(1.44%)
ND_12_3	78042(64.63%)	12265(10.16%)	21446(17.76%)	7228(5.99%)	1763(1.46%)
ND_12_4	76911(63.70%)	12638(10.47%)	22116(18.32%)	7352(6.09%)	1727(1.43%)
D_24_1	78299(64.85%)	12585(10.42%)	21321(17.66%)	6867(5.69%)	1672(1.38%)
D_24_2	76790(63.60%)	13096(10.85%)	22193(18.38%)	7003(5.80%)	1662(1.38%)
D_24_3	78636(65.13%)	12356(10.23%)	21095(17.47%)	6945(5.75%)	1712(1.42%)
D_24_4	77472(64.16%)	12597(10.43%)	21860(18.10%)	7163(5.93%)	1652(1.37%)
ND_24_1	76585(63.43%)	13060(10.82%)	22019(18.24%)	7298(6.04%)	1782(1.48%)
ND_24_2	76265(63.16%)	13687(11.34%)	21677(17.95%)	7237(5.99%)	1878(1.56%)
ND_24_3	75617(62.63%)	13593(11.26%)	22355(18.51%)	7340(6.08%)	1839(1.52%)
ND_24_4	76477(63.34%)	14274(11.82%)	21296(17.64%)	6683(5.53%)	2014(1.67%)

Sample descriptions are as shown in Table S2

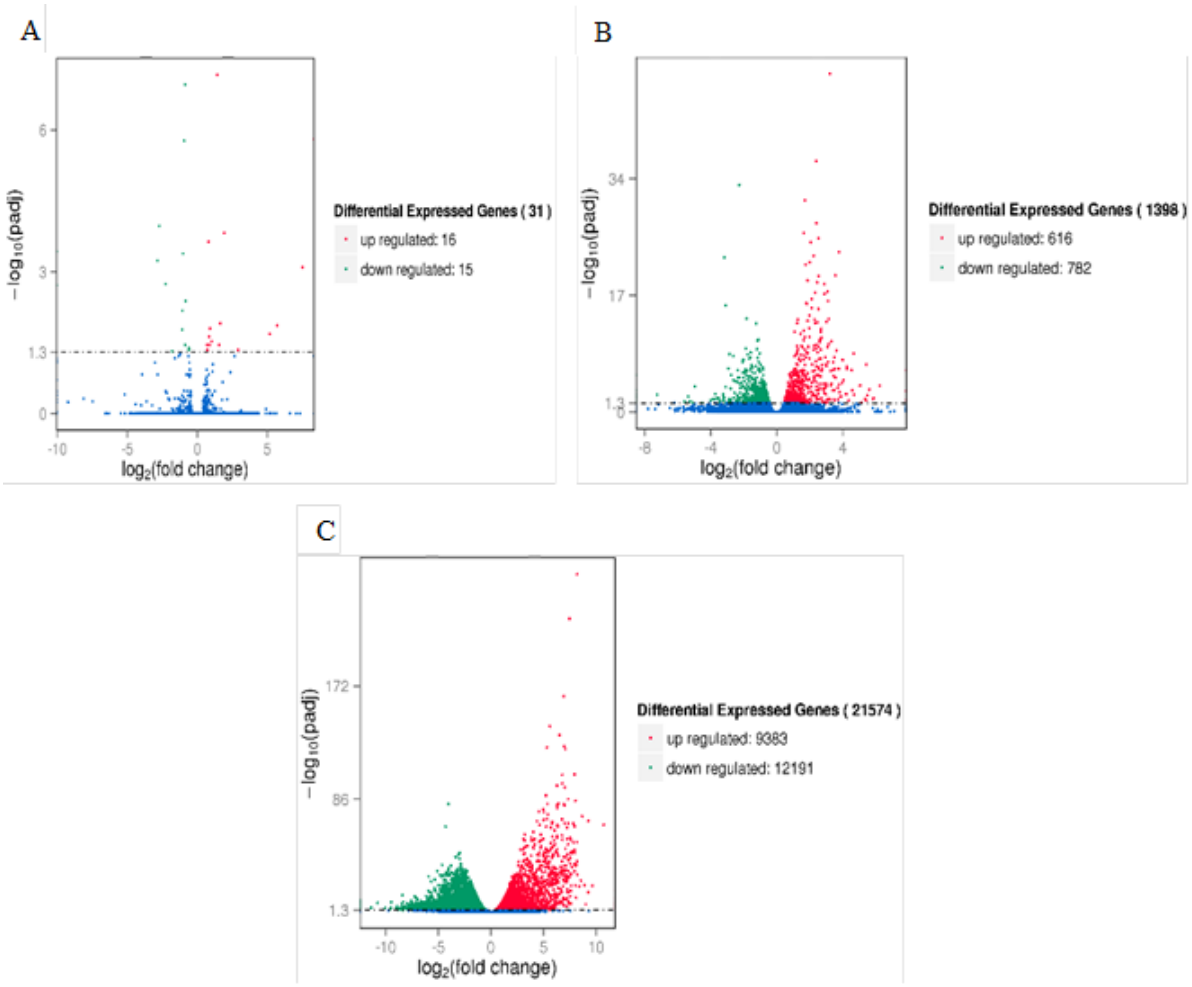
Appendix 5 Pearson correlation coefficients among the four biological replicates of each dormant and non-dormant seed samples used for RNA-seq analysis. Replications of the samples with highest correlation coefficients are highlighted in yellow boxes. D, dormant; ND, non-dormant genotypes; 0, 12 and 24 represent 0, 12 and 24 hours after imbibition (HAI), respectively.



Appendix 6 Genes differentially expressed between seeds of the dormant (D) and non-dormant (ND) double haploid lines. Comparison of expression levels between D seeds at 0 and 12 HAI (D_0/D_12) (A) and 12 and 24 HAI (D_12/D_24) (B), and between ND seeds at 0 and 12 HAI (ND_0/ND_12) (C) and 12 and 24 HAI (ND_12/ND_24) (D). The x-axis shows the log2 fold change in gene expression and the y-axis shows the adjusted P values. Genes significantly upregulated and downregulated are shown by red and green, respectively. Genes that did not show differential expression between the respective samples are shown in blue, separated from the red and green colours by a broken line.



Appendix 7 Genes differentially expressed between seeds of the dormant (D) and non-dormant (ND) double haploid lines. Comparisons of expression levels between D and ND seeds at 0 (D_0/ND_0) (A), 12 HAI (D_12/ND_12) (B) and 24 HAI (D_24/ND_24) (C). The x-axis shows the log2 fold change in gene expression and the y-axis shows the adjusted P values. Genes significantly upregulated and downregulated are shown by red and green, respectively. Genes that did not show differential expression between the respective samples are shown in blue, separated from the red and green colours by a broken line.



Appendix 8 List of genes differentially expressed (≥300-fold, P ≤ 0.05) between dry and imbibed dormant and non-dormant seed samples

	Physical position (bp)		D_0/ND_0		D_12/ND_12		D_24/ND_24	
IWGSC gene ID	Start	End	Fold change	Expression	Fold change	Expression	Fold change	Expression
TraesCS1A02G047400	29280019	29280854	-1.25	FALSE	6.76	UP	1712.63	UP
TraesCS4A02G313300	604055487	604056041	3.78	UP	N.D.	UP	818.14	UP
TraesCS4B02G017600	13077784	13080425	1.28	FALSE	13.67	UP	621.41	UP
TraesCS4A02G098800	110876869	110877138	-1.01	FALSE	2.34	FALSE	617.80	UP
TraesCS1B02G047600	27354337	27355181	1.56	FALSE	26.67	FALSE	527.82	UP
TraesCS7A02G179200	133667919	133668350	1.26	FALSE	9.79	FALSE	489.07	UP
TraesCS6D02G332900	435351033	435352218	1.35	FALSE	18.92	UP	415.59	UP
TraesCS2B02G471500	668029716	668030048	-1.01	FALSE	5.81	FALSE	388.78	UP
TraesCS1D02G293000	392130728	392133462	1.56	FALSE	2.60	UP	342.91	UP
TraesCS1D02G379300	455121874	455124594	1.33	FALSE	4.66	UP	303.28	UP
TraesCS7B02G367100	630662177	630662818	N.D.	FALSE	N.D.	N.D.	-317.41	DOWN
TraesCS2A02G572500	768626152	768627892	N.D.	N.D.	N.D.	N.D.	-319.28	DOWN
TraesCS2D02G108200	60077807	60079509	N.D.	N.D.	N.D.	N.D.	-319.66	DOWN
TraesCS6B02G048800	28722401	28723122	N.D.	N.D.	-2.66	FALSE	-319.75	DOWN
TraesCS5D02G388800	458179759	458183316	N.D.	N.D.	1.06	FALSE	-324.17	DOWN
TraesCS5D02G064100	59244121	59247328	N.D.	N.D.	N.D.	N.D.	-328.67	DOWN
TraesCS7A02G477100	670945813	670947380	N.D.	N.D.	N.D.	N.D.	-334.83	DOWN
TraesCS2D02G107500	59723856	59725690	N.D.	N.D.	N.D.	N.D.	-335.97	DOWN
TraesCS5A02G535600	692179351	692179782	N.D.	N.D.	N.D.	N.D.	-338.85	DOWN
TraesCS3D02G183400	168190550	168193371	2.07	FALSE	-2.45	FALSE	-341.68	DOWN
TraesCS4D02G207900	357529483	357531474	N.D.	N.D.	-1.09	FALSE	-359.01	DOWN
TraesCS1B02G376600	608335179	608335695	N.D.	N.D.	N.D.	N.D.	-360.01	DOWN
TraesCS2D02G497200	593218605	593220794	-4.30	FALSE	-18.08	FALSE	-360.34	DOWN
TraesCS1A02G031100	14467959	14469389	N.D.	N.D.	-2.14	FALSE	-373.15	DOWN
TraesCS6A02G163700	158943593	158947407	N.D.	N.D.	N.D.	N.D.	-405.29	DOWN
TraesCS1D02G260500	355072353	355079652	-1.08	FALSE	1.09	FALSE	-417.61	DOWN
TraesCS4A02G484800	738886426	738890686	N.D.	N.D.	N.D.	N.D.	-460.77	DOWN
TraesCS7D02G453500	572850416	572851057	N.D.	N.D.	N.D.	N.D.	-480.44	DOWN
TraesCS7B02G385200	650539649	650540050	N.D.	N.D.	N.D.	N.D.	-498.55	DOWN
TraesCS1B02G416600	641523153	641523608	-21.35	FALSE	-73.50	FALSE	-504.60	DOWN
TraesCS1A02G360100	541460217	541460850	N.D.	N.D.	N.D.	N.D.	-522.54	DOWN
TraesCS7A02G474200	669727531	669729352	N.D.	N.D.	-2.71	FALSE	-701.74	DOWN
TraesCS5D02G398500	464617022	464620872	N.D.	N.D.	1.06	FALSE	-739.04	DOWN
TraesCS2A02G108100	60077974	60079508	N.D.	N.D.	N.D.	N.D.	-766.85	DOWN
TraesCS2D02G107900	60021772	60023316	N.D.	N.D.	1.76	FALSE	-878.32	FALSE
TraesCS5A02G388400	585066965	585071106	N.D.	N.D.	N.D.	N.D.	-1074.91	DOWN
TraesCS4D02G252100	420939928	420944671	N.D.	N.D.	2.58	FALSE	-1755.91	DOWN
TraesCS2B02G599900	782684221	782688388	N.D.	N.D.	N.D.	N.D.	-2670.70	DOWN

N.D., not determined; FALSE, insignificant

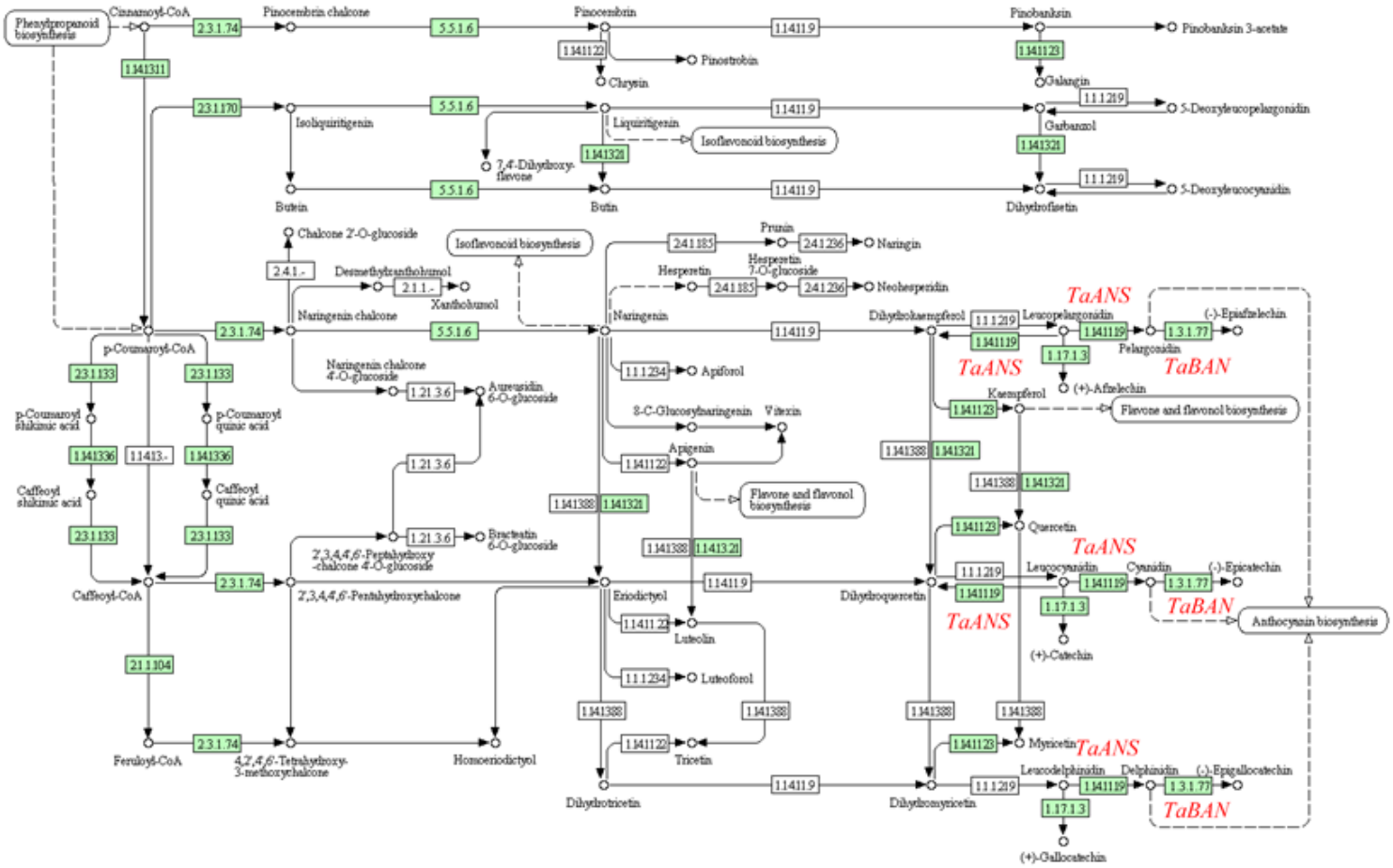
Note: the cut off value was set at ≥ ±300-fold, P ≤ 0.05 due to the detection of large number of DEGs

Appendix 9 Number of gene differentially expressed in any of the comparisons and enriched in different GO classes

Comparisons	Total no. of DEGs	Upregulated genes	Downregulated genes
D_0/ND_0	21	11	10
D_12/ND_12	1071	439	632
D_24/ND_24	16773	6913	9860
D_0/D_12	24261	11353	12908
D_12/D_24	3323	1675	1648
ND_0/ND_12	24232	11137	13095
ND_12/ND_24	10265	4072	6193

Sample descriptions are as shown in Table S2

Appendix 10 Flavonoid biosynthesis pathway enriched with genes differentially expressed between the seeds of dormant and non-dormant double haploid lines. Boxes highlighted in light green represent differentially expressed genes (P≤0.05). Genes with red and blue fonts represent those up- and downregulated in dormant seeds (≥3-fold), respectively.



Appendix 11 Differentially expressed genes (≥2-fold, P ≤ 0.05)) located in *QPhs.umb-4A* along with their orthologues and GO terms

	Oryza sativa spp. *japonica*		*Arabidopsis thaliana*		GO terms					
IWGSC gene ID	Orthologues	Gene ID	Orthologues	Gene ID	Molecular Function GO ID	Molecular Function description	Biological Process GO ID	Biological Process description	Cellular Component GO ID	Cellular Component description
TraesCS4A02G313300	*OsPM19L*	Os05g0381400	-	-	-	-	-	-	-	-
TraesCS4A02G313400	*OsPM19L*	Os07g0422100	-	-	-	-	-	-	-	-
TraesCS4A02G319700	*OsDR*	Os12g0464400	*AtHSD4*	AT5G50690	-	-	-	-	-	-
TraesCS4A02G315800	-	-	*AtMED13*	AT1G55325	GO:0003712	transcription coregulator activity	GO:0006357	regulation of transcription by RNA polymerase II	GO:0070847	core mediator complex
TraesCS4A02G311700	*OsACR10*	Os09g0570900	*AtACR10*	AT2G36840	-	-	-	-	-	-
TraesCS4A02G320200	-	Os03g0856500	-	-	GO:0043022	ribosome binding	GO:0017148	negative regulation of translation	GO:0022627	cytosolic small ribosomal subunit
TraesCS4A02G312200	-	Os04g0165300	*AtATI1*	AT2G45980	-	-	-	-	GO:0043231	intracellular membrane-bounded organelle
TraesCS4A02G313600	-	-	*AtUCB5*	AT1G63800	GO:0061650	ubiquitin-like protein conjugating enzyme activity	GO:0000209	protein polyubiquitination	GO:0005634	nucleus
TraesCS4A02G322600	-	Os01g0977600	*AtLTN1*	AT5G58410	GO:0061630	ubiquitin protein ligase activity	GO:0006414	translational elongation	GO:0032991	protein-containing complex
TraesCS4A02G322700	-	Os06g0204600	*AtFBL75*	AT4G14103	-	-	-	-	-	-
TraesCS4A02G323300	-	Os06g0204600	*AtFBL75*	AT4G14103	-	-	-	-	-	-
TraesCS4A02G311300	*OsPHT4;4*	Os09g0570400	*AtPHT4;4*	AT4G00370	-	-	-	-	-	-
TraesCS4A02G318900	-	Os03g0857500	-	-	-	-	-	-	-	-
TraesCS4A02G314400	-	Os12g0564800	*AtRPM3*	AT3G07040	-	-	-	-	-	-
TraesCS4A02G318800	-	Os03g0857600	-	-	-	-	-	-	-	-
TraesCS4A02G319100	*OsGA20OX1*	Os03g0856700	*AtGA20OX1*	AT4G25420	GO:0051213	dioxygenase activity	-	-	-	-
TraesCS4A02G312600	*OsCYP19-3*	Os09g0571400	*AtCYP19-3*	AT3G56070	-	-	-	-	-	-
TraesCS4A02G323700	-	Os03g0854100	*AtAGD10*	AT2G35210	-	-	GO:0048194; GO:0048199	Golgi vesicle budding; vesicle targeting, to, from or within Golgi	-	-
TraesCS4A02G323800	-	Os03g0854000	*AtMZA15.13*	-	-	-	-	-	-	-
TraesCS4A02G314300	-	Os03g0854000	*AtRPP13*	AT3G46530	-	-	-	-	-	-
TraesCS4A02G313200	*OsYUC6*	Os07g0437000	*AtYUC3*	AT2G33230	-	-	-	-	-	-
TraesCS4A02G323500	*OsGnk2RLK-8*	Os10g0136500	*AtRKS1*	AT1G11340	GO:0004674	protein serine/threonine kinase activity	GO:0006468	protein phosphorylation	GO:0005886	plasma membrane
TraesCS4A02G316700	*OsDUF668-5*	Os03g0858600	-	-	-	-	-	-	-	-
TraesCS4A02G323400	*OsGnk2RLK-7*	Os10g0136400	*AtCRK29*	AT4G21410	GO:0004674	protein serine/threonine kinase activity	GO:0046777	protein autophosphorylation	GO:0005886	plasma membrane
TraesCS4A02G317200	-	Os06g0121400	-	-	-	-	-	-	-	-

-, refers to no GO term

Appendix 12 Differentially expressed genes (≥10, P ≤ 0.05) located in *QPhs.umb-6B* along with their orthologues and GO terms

	Oryza sativa spp. *japonica*		*Arabidopsis thaliana*		GO terms					
IWGSC gene ID	Orthologues	Gene ID	Orthologues	Gene ID	Molecular Function GO ID	Molecular Function description	Biological Process GO ID	Biological Process description	Cellular Component GO ID	Cellular Component description
TraesCS6B02G325000	-	Os11g0533400	-	-	-	-	-	-	-	-
TraesCS6B02G296400	*OsAOX1C*	Os02g0700400	*AtAOX2*	AT5G64210	GO:0016679	oxidoreductase activity, acting on diphenols and related substances as donors	GO:0009060	aerobic respiration	GO:0005739	mitochondrion
TraesCS6B02G326000	*OsSYL2*	Os02g0733900	-	-	-	-	-	-	-	-
TraesCS6B02G314600	*OsETR4*	Os07g0259100	*AtERS2*	AT1G04310	GO:0038023; GO:0005488	signaling receptor activity; binding	-	-	GO:0005886; GO:0005783; GO:0005773	plasma membrane; endoplasmic reticulum; vacuole
TraesCS6B02G282100	*OsGLPT1*	Os02g0652800	*AtGLPT4*	AT4G17550	-	-	-	-	-	-
TraesCS6B02G317100	*OsDOF4*	Os02g0726300	*AtDOF2.1*	AT2G28510	-	-	-	-	-	-
TraesCS6B02G291800	*OsUBP7-2*	Os02g0795000	*AtUBP20*	AT4G17895	GO:0004197; GO:0004843	cysteine-type endopeptidase activity; thiol-dependent ubiquitin-specific protease activity	GO:0016579	protein deubiquitination	GO:0005829; GO:0005634	cytosol; nucleus
TraesCS6B02G331500	-	Os02g0751900	*AtIP5P5*	AT5G65090	GO:0016791	phosphatase activity	GO:0046488; GO:0046839	phosphatidylinositol metabolic process; phospholipid dephosphorylation	-	-
TraesCS6B02G288800	-	Os02g0686300	-	AT5G44820	-	-	-	-	-	-
TraesCS6B02G295400	-	Os07g0166500	-	-	-	-	-	-	-	-
TraesCS6B02G276800	*OsSAM2*	Os01g0323600	*AtMETK4*	AT3G17390	GO:0016765	transferase activity, transferring alkyl or aryl (other than methyl) groups	GO:0006790; GO:1901576; GO:0044249	sulfur compound metabolic process; organic substance biosynthetic process; cellular biosynthetic process	GO:0005829	cytosol
TraesCS6B02G303500	*OsRSL2*	Os02g0710300	*AtRHD6*	AT1G66470	GO:0000978; GO:0000981	RNA polymerase II cis-regulatory region sequence-specific DNA binding; DNA-binding transcription factor activity, RNA polymerase II-specific	GO:0006366; GO:0006357	transcription by RNA polymerase II; regulation of transcription by RNA polymerase II	GO:0005634	nucleus
TraesCS6B02G307400	*OsEIF3B*	Os10g0569200	*AtTIF3B1*	AT5G27640	GO:0003743	translation initiation factor activity	GO:0006413; GO:0006414	translational initiation; translational elongation	GO:0005852	eukaryotic translation initiation factor 3 complex
TraesCS6B02G247400	-	Os02g0607700	*AtSDR1*	AT3G61220	-	-	GO:0055114	oxidation-reduction process	GO:0016020	membrane
TraesCS6B02G313200	*OsENODL24*	Os12g0454600	*AtENODL18*	AT1G08500					GO:0046658	anchored component of plasma membrane
TraesCS6B02G289300	-	Os04g0586000	*AtFLZ3*	AT2G44670	-	-	-	-	-	-
TraesCS6B02G272200	-	Os02g0667300	-	-	-	-	-	-	-	-
TraesCS6B02G273400	*OsDHN1*	Os02g0669100	*AtERD14*	AT1G76180	-	-	GO:0009737; GO:0010035; GO:0009409	response to abscisic acid; response to inorganic substance; response to cold	GO:0016020;GO:0005829	membrane; cytosol
TraesCS6B02G275700	*OsEXPB11*	Os02g0658800	*AtEXPB6*	AT1G65681						
TraesCS6B02G311300	-	Os09g0246700	-	-	-	-	-	-	-	-
TraesCS6B02G259000	*OsPIP2;2*	Os02g0629200	*AtPIP2*	AT2G37170	GO:0015267; GO:0015318	channel activity; inorganic molecular entity transmembrane transporter activity	-	-	GO:0005886	plasma membrane
TraesCS6B02G288000	-	-	-	-	GO:0003677	DNA binding	GO:0006334	nucleosome assembly	-	-
TraesCS6B02G284500	-	-	-	-	-	-	-	-	-	-
TraesCS6B02G265300	*OsRH8*	Os02g0641800	*AtRH8*	AT4G00660	GO:0003729	mRNA binding	GO:0033962	cytoplasmic mRNA processing body assembly	GO:0010494	cytoplasmic stress granule

-, represents no GO terms and description

Appendix 13 Differentially expressed genes (≥ 10, P ≤ 0.05) located in 1D and 6D QTLs along with their orthologues and GO terms

	Oryza sativa spp. *japonica*		*Arabidopsis thaliana*		GO terms					
IWGSC gene ID	Orthologues	Gene ID	Orthologues	Gene ID	Molecular Function GO ID	Molecular Function description	Biological Process GO ID	Biological Process description	Cellular Component GO ID	Cellular Component description
						1D QTL				
TraesCS1D02G403200	*OsCYP71Y10*	Os06g0641500	*AtCYP71B11*	AT5G25120	-	-	-	-	-	-
TraesCS1D02G400300	*OsDOG1L-2*	Os05g0560200	*AtDOG1*	AT5G45830	-	-	-	-	-	-
TraesCS1D02G414700	*OsISC5*	Os05g0568100	*AtISU1*	AT4G04080	GO:0051539; GO:0008198	4 iron, 4 sulfur cluster binding; ferrous iron binding	GO:0006879	cellular iron ion homeostasis	GO:0005759	mitochondrial matrix
TraesCS1D02G395300	-	Os11g0503900	*AtSULT202B1*	AT3G45070	-	-	-	-	-	-
TraesCS1D02G413600	-	Os05g0567800	*AtERDL4*	AT1G19450	-	-	-	-	-	-
TraesCS1D02G412000	-	Os07g0274700	*AtAMY2*	AT1G76130	-	-	-	-	-	-
TraesCS1D02G396600	*OsH2B.7*	Os01g0152900	*AtHTB2*	AT5G22880	GO:0003677	DNA binding	GO:0006334	nucleosome assembly	-	-
TraesCS1D02G410700	-	Os02g0771200	-	-	-	-	-	-	-	-
						6D QTL				
TraesCS6D02G197600	*OsLEA4*	Os08g0327700	*AtLEA46*	AT5G06760	-	-	-	-	-	-
TraesCS6D02G241200	*OsOX3*	Os02g0684400	*AtOXS3*	AT5G56550	-	-	-	-	-	-
TraesCS6D02G224300	*OsGLPT1*	Os02g0652800	*AtGLPT4*	AT4G17550	-	-	-	-	-	-
TraesCS6D02G203800	-	Os02g0612900	*AtTIL*	AT5G58070	-	-	GO:1901700; GO:0006979; GO:0006629	response to oxygen-containing compound; response to oxidative stress; lipid metabolic process	GO:0005737	cytoplasm
TraesCS6D02G211600	*OsPT4*	Os04g0186400	*AtPHT1;4*	AT2G38940	-	-	-	-	-	-
TraesCS6D02G197500	*OsbHLH111*	Os04g0489600	*AtBHLH129*	AT2G43140	GO:0000978; GO:0000981	RNA polymerase II cis-regulatory region sequence-specific DNA binding; DNA-binding transcription factor activity, RNA polymerase II-specific	GO:0006366; GO:0006357	transcription by RNA polymerase II; regulation of transcription by RNA polymerase II	GO:0005634	nucleus
TraesCS6D02G222300	*OsALDH3E1*	Os02g0646500	*AtALDH3F1*	AT4G36250	GO:0016491	oxidoreductase activity	-	-	-	-
TraesCS6D02G199500	*OsCHI6*	Os02g0605900	*AtCHI4*	AT2G43570	-	-	-	-	-	-
TraesCS6D02G248800	*OsOPT4*	Os02g0695800	*AtOPT2*	AT1G09930	GO:0035673	oligopeptide transmembrane transporter activity	-	-	GO:0005887	integral component of plasma membrane
TraesCS6D02G200400	-	Os02g0607700	*AtSDR1*	AT3G61220	-	-	GO:0055114	oxidation-reduction process	GO:0016020	membrane
TraesCS6D02G222400	-	-	-	-	-	-	-	-	-	-
TraesCS6D02G226900	-	Os02g0666700	*AtNCS1*	AT5G03555	GO:0015205	nucleobase transmembrane transporter activity	GO:0071705; GO:0071702	nitrogen compound transport; organic substance transport	GO:0005886	plasma membrane
TraesCS6D02G235300	*OsRLCK82*	Os02g0670100	*AtKIN3*	AT5G25440	-	-	-	-	-	-
TraesCS6D02G207900	-	Os02g0620100	*AtMIS12*	AT5G35520	-	-	GO:0034501; GO:0000070; GO:0051382	protein localization to kinetochore; mitotic sister chromatid segregation; kinetochore assembly	GO:0032991 ; GO:0000778	protein-containing complex; condensed nuclear chromosome kinetochore
TraesCS6D02G212900	*OsPIP2-2*	Os02g0629200	*AtPIP2*	AT2G37170	GO:0015318; GO:0015267	inorganic molecular entity transmembrane transporter activity; channel activity	-	-	GO:0005886	plasma membrane
TraesCS6D02G241500	-	-	-	-	GO:0003677	DNA binding	GO:0006334	nucleosome assembly	-	-

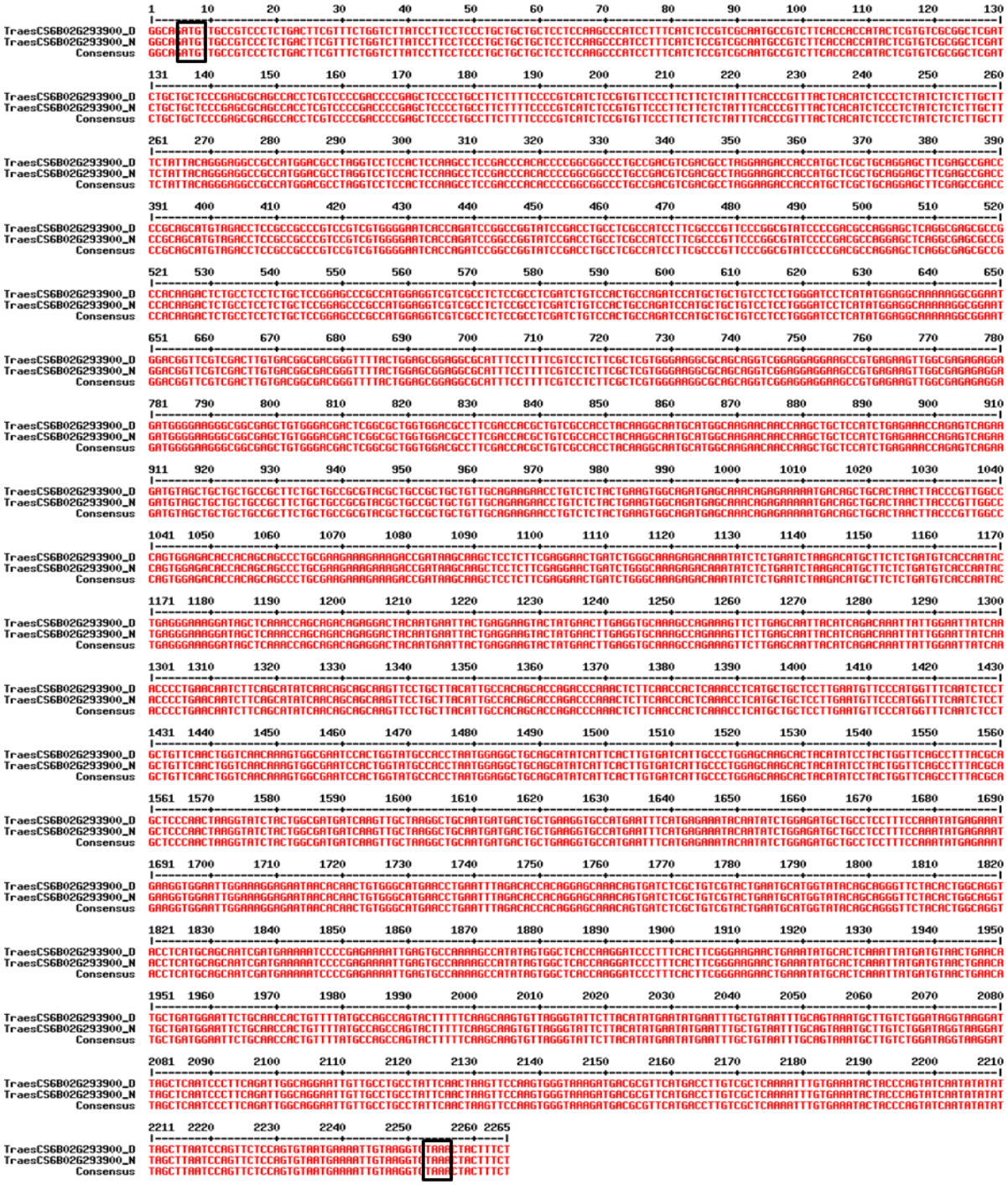
-, represents no GO terms and description

Appendix 14 Number of SNPs and InDels distributed across the 21 chromosomes of wheat

Chromosome	No. SNPs	No. of InDels
1A	8031	690
1B	9389	920
1D	2584	333
2A	7121	760
2B	11254	1095
2D	2232	254
3A	4944	524
3B	10784	961
3D	2037	315
4A	5890	573
4B	4168	424
4D	1289	192
5A	5885	646
5B	7004	779
5D	1954	313
6A	3857	408
6B	10362	918
6D	1747	240
7A	7969	692
7B	6971	588
7D	2515	233
Unassigned*	2688	125
Total	120675	11983

*Unassigned to any chromosome

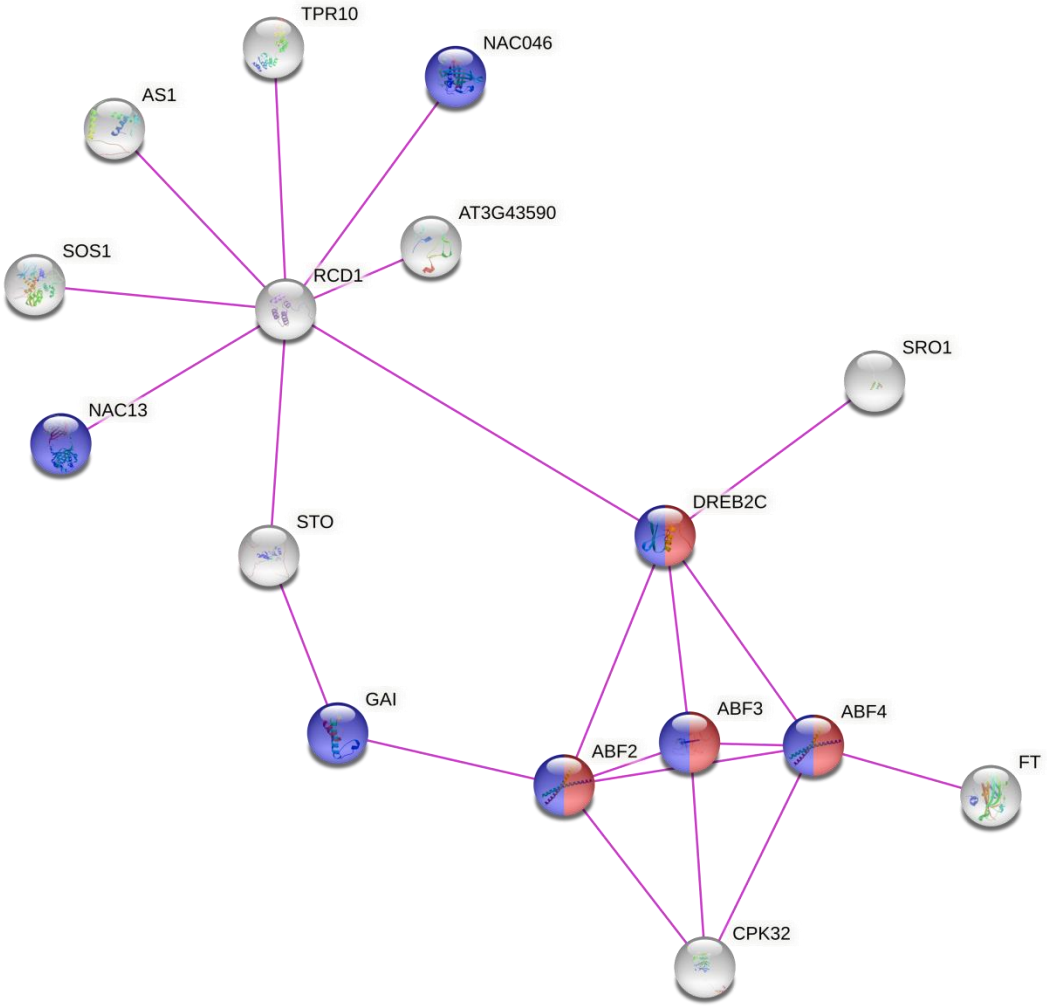
Appendix 15 Comparison of the coding sequences (CDS) of an uncharacterized gene (Gene ID: TraesCS6B02G293900) derived from dormant (D) and non-dormant (N) wheat seed. Start and stop codons of the gene are indicated by black rectangles.



Appendix 16 Number of differentially expressed genes encoding transcription factors (TFs) and transcriptional regulators (TRs)

Transcription factors			
TFs Family	No. of DEGs	TFs Family	No. of DEGs
AP2-EREBP	250	C2C2-CO-like	11
MYB	224	DBP	11
bHLH	204	LIM	10
mTERF	173	ARR-B	9
bZIP	165	C2C2-YABBY	9
HB	151	CSD	9
C2H2	139	VOZ	6
NAC	138	PBF-2-like	5
C3H	128	BBR/BPC	3
CCAAT	125	ULT	3
ABI3V/P1	109	HRT	2
WRKY	104	S1Fa-like	2
GRAS	90	FAR1	1
G2-like	85	**Total**	**3030**
ARF	59		
Trihelix	58	Transcriptional regulators	
HSF	56	TRs family	No. of DEGs
C2C2-Dof	53	Orphans	107
C2C2-GATA	53	PHD	102
LOB	45	SET	95
TCP	45	SNF2	85
MADS	44	AUX/IAA	64
FHA	41	GNAT	62
OFP	37	TRAF	55
TUB	31	Jumonji	31
SBP	29	SWI/SNF-BAF60b	31
BSD	28	HMG	23
RWP-RK	26	DDT	20
GeBP	25	IWS1	17
GRF	23	LUG	14
PLATZ	23	ARID	13
E2F-DP	22	SWI/SNF-SWI3	10
Tify	22	MBF1	9
zf-HD	22	Pseudo ARR-B	9
EIL	21	Coactivator p15	6
CPP	18	RB	6
Alfin-like	17	Rcd1-like	6
SRS	15	MED6	3
Sigma70-like	14	MED7	3
TAZ	14	SOH1	2
CAMTA	12	**Total**	**773**
BES1	11		

Appendix 17 Gene regulatory networks of *TaDREB2C-B* based on STRING database. Genes denoted by red color nods represent those involved in ABA signaling pathway while genes denoted with light violet colour represent those acting as transcriptional activator.



Appendix 18 List of abbreviations

°	Degree
PHS	Pre-harvest sprouting
ABA	Abscisic acid
GA	Gibberellin
QTL	Quantitative trait locus
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
SNP	Single nucleotide polymorphisms
BP	Before present
CWRS	Canada Western Red Spring
CWES	Canada Western Extra Strong
CPSR	Canada Prairie Spring Red
CWRW	Canada Western Red Winter
CPSW	Canada Prairie Spring White
CWSWS	Canada Western Soft White Spring
CWHWS	Canada Western Hard White Spring
CNHR	Canada Northern Hard Red
CWAD	Canada Western Amber Durum
CGC	Canadian Grain Commission
ET	Ethylene
CK	Cytokinin
BR	Brassinosteroids
JA	Jasmonic acid
SM	Secondary metabolites
ROS	Reactive oxygen species
BC	Backcross
DH	Doubled haploid
NIL	Near isogenic lines
RIL	Recombinant inbred lines
NAM	Nested association mapping
MAGIC	Multi-parent advanced generation inter-cross
RFLP	Restriction Fragment Length Polymorphism
STS	Sequence-Tagged Site
RAPD	Random Amplified Polymorphic DNA
EST	Expressed Sequence Tags
AFLP	Amplified Fragment Length Polymorphisms
KASP	Competitive Allele-Specific PCR
DArT	Diversity Array Technology
GBS	Genotyping-By-Sequencing
cM	Centimorgan
FN	Falling number
SI	Sprouting index

SR	Sprouting rate
GI	Germination index
IM	Interval mapping
MIM	Multiple interval mapping
CIM	Composite interval mapping
ICIM	Inclusive composite interval mapping
GC	Grain colour
AA	α-amylase activity
GR	Germination rate
DI	Dormancy index
SGI	Speed of germination index
GP	Germination percentage
NGS	Next generation sequencing
ANOVA	Analysis of variance
LOD	Logarithm of odds
DEG	Differentially expressed genes
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
HAI	Hours after imbibition
PCR	Polymerase chain reaction
CS	Chinese Spring
FC	Fold change
InDel	Insertion and Deletion
RT-qPCR	Reverse Transcriptase-quantitative PCR
DAI	Days after imbibition
ND	Non-dormant
D	Dormant
FPKM	Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced
TF	Transcription factor
TR	Transcription regulators