# Genomic Selection, Quantitative Trait Loci and Genome-Wide Association Mapping for Spring Bread Wheat (*Triticum aestivum* L.) Improvement

A Thesis Submitted to the College of

Graduate and Postdoctoral Studies

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

In the Department of Plant Sciences

University of Saskatchewan

Saskatoon

By

Teketel Astatike Haile

## Permission to Use

In presenting this thesis/dissertation in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis/dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis/dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

Requests for permission to copy or to make other uses of materials in this thesis/dissertation in whole or part should be addressed to:

Head of the Department of Plant Sciences
51 Campus Drive
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5A8
Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9
Canada

## Abstract

Molecular breeding involves the use of molecular markers to identify and characterize genes that control quantitative traits. Two of the most commonly used methods to dissect complex traits in plants are linkage analysis and association mapping. These methods are used to identify markers associated with quantitative trait loci (QTL) that underlie trait variation, which are used for marker assisted selection (MAS). Marker assisted selection has been successful to improve traits controlled by moderate to large effect QTL; however, it has limited application for traits controlled by many QTL with small effects. Genomic selection (GS) is suggested to overcome the limitation of MAS and improve genetic gain of quantitative traits. GS is a type of MAS that estimates the effects of genome-wide markers to calculate genomic estimated breeding values (GEBVs) for individuals without phenotypic records. In recent years, GS is gaining momentum in crop breeding programs but there is limited empirical evidence for practical application. The objectives of this study were to: i) evaluate the performance of various statistical approaches and models to predict agronomic and end-use quality traits using empirical data in spring bread wheat, ii) determine the effects of training population (TP) size, marker density, and population structure on genomic prediction accuracy, iii) examine GS prediction accuracy when modelling genotype-by-environment interaction (G × E) using different approaches, iv) detect marker-trait associations for agronomic and end-use quality traits in spring bread wheat, v) evaluate the effects of TP composition, cross-validation technique, and genetic relationship between the TP and SC on GS accuracy, and vi) compare genomic and phenotypic prediction accuracy. Six studies were conducted to meet these objectives using two populations of 231 and 304 spring bread wheat lines that were genotyped with the wheat 90K SNP array and phenotyped for nine agronomic and end-use quality traits. The main finding across these studies is that GS can accurately predict GEBVs for wheat traits and can be used to make predictions in different environments; thus, GS should be applied in wheat breeding programs. Each study provides specific insights into some of the advantages and limitations of different GS approaches, and gives recommendations for the application of GS in future breeding programs. Specific recommendations include using the GS model BayesB (especially for large effect QTL) for genomic prediction in a single environment, across-year genomic prediction using the reaction norm model, using a large TP size for making accurate genomic predictions, and not making across-population genomic predictions except for highly related populations.

**Acknowledgements**

I am very grateful to my supervisor, Dr. Curtis Pozniak, for his insightful guidance and strong support throughout my study. Dr. Pozniak gave me the opportunity to work independently, while providing excellent feedback whenever it is needed. Thank you for being a phenomenal academic advisor. I want to extend my sincere acknowledgement to members of my advisory committee: Dr. Pierre Hucl, Dr. Fiona Buchanan, Dr. Aaron Beattie and Dr. Yuguang Bai for their valuable suggestions and constructive comments on my research project and dissertation. I would also like to thank Dr. Clay H. Sneller, Ohio State University for agreeing to act as external examiner of this dissertation.

Financial support for this research has come from several sources. I want to thank Genome Canada, Saskatchewan Ministry of Agriculture, Saskatchewan Wheat Development Commission, and the Western Grains Research Foundation for funding this research. I am also very grateful to the Department of Plant Sciences, the College of Agriculture and Bioresources, and the College of Graduate and Postdoctoral Studies at the University of Saskatchewan, the Government of Saskatchewan and donors of the Daniel Geddes Graduate Scholarship for the financial support during my study.

My heartfelt thanks are due to my parents Astatike Haile and Assegedech Seteign, and siblings especially Dr. Beliyou Haile and Dr. Getinet Haile for your support, advice and constant encouragement throughout my journey. Special thanks are due to my wife Frehiwot Degefu, for your patience and support during my graduate study.

I want to express my gratitude for the technical assistance from the durum wheat breeding staff at the University of Saskatchewan: Russell Lawrie, Ryan Babonich, Heidi Lazorko, Charlene Tang, Chanese Beierle, Corey Howard, Krystalee Wiebe, Lexie Martin, Justin Coulson, and Jennifer Ens. Furthermore, technical support from Connie Briggs and Doug Medernach at the Grains Innovation lab is highly appreciated.

I also want to thank Dr. Amidou N'Diaye for his help during data processing, Dr. Sean Walkowiak and Dr. John Clarke for their valuable comments on my dissertation. I am also thankful to Dr. Amit Deokar at the University of Saskatchewan, Dr. Jennifer Spindel at the Joint Genome Institute, and Jicai Jiang at the University of Maryland for their help during the statistical analysis. Finally, I would also like to thank Dr. Hema Duddu for his support and friendship.

## Table of Contents

**Permission to Use** ..... **i**

**Abstract** ..... **ii**

**Acknowledgements** ..... **iii**

**Table of Contents** ..... **iv**

**List of Tables** ..... **viii**

**List of Figures** ..... **x**

**List of Abbreviations** ..... **xiii**

**1. Introduction** ..... **1**

**2. Literature Review** ..... **6**

2.1 Wheat Production ..... 6

2.2 Evolution of Wheat ..... 7

2.3 Molecular Markers in Wheat Breeding ..... 8

2.4 Traditional Applications of Molecular Markers in Wheat: Mapping and MAS ..... 10

2.5 Prospects for Genomic Selection ..... 13

2.6 Statistical Methods for Genomic Selection ..... 14

2.6.1 Ridge regression best linear unbiased prediction ..... 15

2.6.2 Bayesian regression models ..... 16

2.6.3 Kernel models ..... 18

2.6.4 Multiple-trait prediction models ..... 19

2.7 Factors Affecting GS Model Prediction Accuracy ..... 21

2.7.1 Training population size and composition ..... 21

2.7.2 Marker type and density ..... 22

2.7.3 Genetic relationships and marker-QTL LD ..... 24

2.7.4 Heritability and genetic architecture of traits ..... 25

2.7.5 Genotype-by-environment interaction and population structure ..... 26

2.8 Reported Genomic Prediction Accuracies in Wheat ..... 28

2.9 Thesis Objectives and Outline ..... 33

**3. Genomic Selection for Wheat Improvement: Comparison of Methods Based on Empirical Data** ..... **34**

3.1 Introduction ..... 34

3.2 Materials and Methods ..... 36
3.2.1 Plant material and phenotypic data ..... 36
3.2.2 Genotypic data ..... 37
3.2.3 Phenotypic data analysis ..... 37
3.2.4 Single-trait prediction model ..... 38
3.2.5 Multiple-trait prediction model ..... 39
3.2.6 Genomic prediction with significant markers from GWAS fitted as fixed effects ..... 39
3.2.7 Cross-validation techniques ..... 40
3.3 Results and Discussion ..... 41
3.3.1 Phenotypic evaluation ..... 41
3.3.2 Accuracy of single-trait genomic predictions ..... 45
3.3.3 Effect of trait heritability on prediction accuracy ..... 47
3.3.4 Assessment of prediction accuracy based on rank correlation ..... 48
3.3.5 Accuracy of multiple-trait genomic predictions ..... 51
3.3.6 Prediction accuracy of GS + de novo GWAS model ..... 55
3.4 Conclusion ..... 57
**4. Effects of Marker Density, Training Population Size and Population Structure on GS Accuracy in Wheat ..... 59**
4.1 Introduction ..... 59
4.2 Materials and Methods ..... 59
4.2.1 Effect of training population size on genomic prediction accuracy ..... 60
4.2.2 Effect of marker density on genomic prediction accuracy ..... 60
4.2.3 Effect of population structure on genomic prediction accuracy ..... 61
4.3 Results and Discussion ..... 62
4.3.1 Genomic prediction with increasing training population sizes ..... 62
4.3.2 Genomic prediction with increasing marker density ..... 65
4.3.3 Genomic prediction accounting for population structure ..... 67
4.4 Conclusion ..... 71
**5. Modelling Genotype-by-Environment Interaction for Genomic Prediction in Wheat ... 72**
5.1 Introduction ..... 72
5.2 Materials and Methods ..... 73
5.2.1 Phenotypic data analysis ..... 73
5.2.2 Statistical models used to incorporate genotype-by-environment interaction in GS ... 74
5.2.2.1 Marker-by-environment interaction GS model ..... 74
5.2.2.2 Reaction norm models that incorporate environmental covariates ..... 76
5.3 Results and Discussion ..... 79
5.3.1 Large phenotypic variations were observed among environments ..... 79
5.3.2 Genomic predictions using the marker-by-environment interaction model ..... 81

5.3.3 Genomic predictions using reaction norm models ..... 86
5.3.4 Comparing the accuracy of marker-by-environment interaction and reaction norm models ..... 89

5.4 Conclusion ..... 90

**6. Genome-wide Association Mapping of Agronomically Important Traits in Wheat ..... 91**

6.1 Introduction ..... 91

6.2 Materials and Methods ..... 92
6.2.1 Plant material, phenotypic and genotypic data ..... 92
6.2.2 Population structure and LD analyses ..... 93
6.2.3 Marker-trait association analysis ..... 94

6.3 Results and Discussion ..... 95
6.3.1 Population structure and LD decay ..... 95
6.3.2 Marker-trait associations ..... 100

6.4 Conclusion ..... 107

**7. Mapping of Quantitative Trait Loci Associated with Agronomically Important Traits in Wheat ..... 108**

7.1 Introduction ..... 108

7.2 Materials and Methods ..... 109
7.2.1 Plant material and phenotypic data ..... 109
7.2.2 Phenotypic data analysis ..... 110
7.2.3 Genotypic data ..... 110
7.2.4 QTL analyses ..... 111

7.3 Results and Discussion ..... 111
7.3.1 Wide variations were observed in measured phenotypes ..... 111
7.3.2 Twenty-three QTL were identified for six agronomic traits ..... 115
7.3.3 Coincident QTL and trait relationships ..... 123
7.3.4 Relationship between QTL from this study and known QTL ..... 124

7.4 Conclusion ..... 127

**8. Accuracy of Genomic and Phenotypic Predictions in Wheat Based on Different Cross-Validation Techniques ..... 128**

8.1 Introduction ..... 128

8.2 Materials and Methods ..... 130
8.2.1 Plant material and phenotypic data ..... 130
8.2.2 Genotypic data ..... 130
8.2.3 Statistical modelling and genomic prediction ..... 130

8.3 Results and Discussion ..... 133

8.3.1 Across-population genomic prediction ..... 133
8.3.2 Within-population genomic prediction ..... 138
8.3.3 Genomic prediction across years for combined locations ..... 140
8.3.4 Genomic prediction across years within locations ..... 142
8.4 Conclusion ..... 144

**9. General Discussion and Conclusions ..... 145**

9.1 Quantitative Trait Loci and Genome-Wide Association Mapping ..... 145

9.2 Comparison of GS Approaches and Models ..... 146

9.3 Effects of Training Population Size, Marker Density, Heritability and Population Structure on GS accuracy ..... 148

9.4 Modelling Genotype-by-Environment Interaction in GS ..... 150

9.5 Effects of Genetic Relatedness, Training Population Composition and Cross-Validation Technique on GS accuracy ..... 152

9.6 Future Research ..... 154

**References ..... 156**

**Appendices ..... 178**

## List of Tables

**Table 3-1.** *P*-values from mixed model ANOVA *F*-tests for nine agronomic and end-use quality traits as affected by genotype and covariance parameters. ..... 42

**Table 3-2.** Broad-sense heritability estimates for nine agronomic and end-use quality traits. .... 43

**Table 3-3.** Correlations among trait phenotypes averaged across environments. ..... 44

**Table 3-4.** Average and standard deviation of prediction accuracy (from fivefold cross-validation) based on Pearson's correlation (r) between trait phenotypes and GEBVs estimated using a single-trait and three multiple-trait prediction models. Prediction was made for different trait combinations..... 53

**Table 5-1.** Environmental covariate data used for modelling G × E using reaction norm models. ..... 77

**Table 5-2.** Phenotypic correlations across seven environments for grain yield (upper diagonal) and protein content (lower diagonal) of 81 lines. ..... 80

**Table 5-3.** Average and standard deviation of prediction accuracy based on Pearson's correlation between predicted and actual grain yield. Genomic predictions were made using 50 training-validation partitions and two cross-validation techniques (CV1 and CV2) across seven environments. ..... 83

**Table 5-4.** Average and standard deviation of prediction accuracy based on Pearson's correlation between predicted and actual grain protein. Genomic predictions were made using 50 training-validation partitions and two cross-validation techniques (CV1 and CV2) across seven environments. ..... 84

**Table 5-5.** Average and standard deviation of prediction accuracy (from fivefold cross-validation) based on Pearson's correlation between predicted and actual grain yield. Genomic predictions were made using six statistical models and two cross-validation techniques (CV1 and CV2) across seven environments. ..... 87

**Table 5-6.** Average and standard deviation of prediction accuracy (from fivefold cross-validation) based on Pearson's correlation between predicted and actual grain protein. Genomic predictions were made using six statistical models and two cross-validation techniques (CV1 and CV2) across seven environments. ..... 88

**Table 6-1.** Association of markers with eight agronomic and end-use quality traits detected with a mixed linear model. Associations that passed the FDR threshold are indicated in bold. ..... 106

**Table 7-1.** Correlations among adjusted traits of the selection candidates averaged across five environments. ..... 114

**Table 7-2.** Summary of QTL identified for six agronomic traits based on 304 RILs evaluated across five environments. QTL analyses were conducted using LS-means of each environment and averaged (combined) across all environments. ..... 118

**Table 8-1.** Prediction accuracy based on Pearson's correlations between GEBVs estimated using two statistical models and trait phenotypes. Predictions were made for the selection candidates using a training population with varying degrees of genetic relationships. ..... 134

**Table 8-2.** Prediction accuracy based on Pearson's correlations between GEBVs estimated using two statistical models and trait phenotypes. Predictions were made for six traits in two groups of selection candidates. ..... 137

**Table 8-3.** Average and standard deviation of prediction accuracy (from fivefold cross-validation) based on Pearson's correlation between GEBVs estimated using six statistical models and trait phenotypes of the selection candidates. ..... 139

**Table 8-4.** Across-year genomic and phenotypic prediction accuracies based on combined data from two sites. Predictions were made for six traits evaluated across two sites (Kernen and Rosthern) and three years for a total of five environments. ..... 141

**Table 8-5.** Across-year genomic and phenotypic prediction accuracies in each site. Predictions were made for six traits evaluated across two sites (Kernen and Rosthern) and three years for a total of five environments. ..... 143

## List of Figures

**Fig. 3-1.** Frequency distributions of nine agronomic and end-use quality traits used to estimate marker effects in the training population. Data were averaged across all environments. ..... 43

**Fig. 3-2.** Average prediction accuracy (from fivefold cross-validation) based on Pearson's correlation (r) between GEBVs estimated from nine statistical models and phenotypes for nine traits. HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight, PRO: grain protein, FN: falling number, SDS: sedimentation volume..... 45

**Fig. 3-3.** Average prediction accuracy (from fivefold cross-validation) based on Spearman's rank correlation (ρ) between GEBVs estimated from nine statistical models and phenotypes for nine traits. HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight, PRO: grain protein, FN: falling number, SDS: sedimentation volume..... 49

**Fig. 3-4.** Prediction accuracy based on Pearson's correlation (r) between GEBVs estimated from nine statistical models and phenotypes for nine traits. The varcomp population (77 lines) was used to estimate marker effects and the Co-op population (154 lines) was used as a validation set. HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight, PRO: grain protein, FN: falling number, SDS: sedimentation volume. 50

**Fig. 3-5.** Prediction accuracy based on Pearson's correlation (r) between GEBVs estimated from nine statistical models and trait phenotypes. The Co-op population (154 lines) was used to estimate marker effects and the varcomp population (77 lines) was used as a validation set. HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight, PRO: grain protein, FN: falling number, SDS: sedimentation volume. 51

**Fig. 3-6.** Average prediction accuracy (from fivefold cross-validation) based on Pearson's correlation (r) between GEBVs estimated using RR-BLUP and GS + de novo GWAS models and trait phenotypes. HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight, PRO: grain protein, FN: falling number, SDS: sedimentation volume. ..... 56

**Fig. 4-1.** Average prediction accuracy (from five training-validation partitions) based on Pearson's correlation (r) between GEBVs and trait phenotypes plotted against training population size. HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight, PRO: grain protein, FN: falling number, SDS: sedimentation volume. ..... 64

**Fig. 4-2.** Average prediction accuracy (from fivefold cross-validation) based on Pearson's correlation (r) between GEBVs estimated for different marker densities and trait phenotypes. HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight, PRO: grain protein, FN: falling number, SDS: sedimentation volume. 65

**Fig. 4-3.** A phylogenetic tree based on neighbor-joining and a Manhattan dissimilarity matrix showing clustering of the 231 wheat lines into three groups represented in different colors..... 68

**Fig. 4-4.** First two principal components showing marker-based K-means clustering of the 231 wheat lines into three groups (Group one = 30 lines, group two = 93 lines, and group three = 108 lines)..... 69

**Fig. 4-5.** Box plots showing prediction accuracy for grain yield estimated from 50 random training-validation partitions using Across-group, Within-group and Interaction BRR (top) and BayesB (bottom) models..... 70

**Fig. 5-1.** Boxplots for (A) grain yield and (B) grain protein content of 81 lines across seven environments. Data within each environment are averages of two replications. SWC: Swift Current. ..... 80

**Fig. 6-1.** The ΔK peak value at K = 2, that indicated the presence of two subpopulations based on Bayesian clustering. ..... 96

**Fig. 6-2.** Population structure analysis. (A) Hierarchical clustering using the Ward's method, (B) bar plots showing subpopulations represented by different colors based on the population membership coefficients obtained from Bayesian clustering. ..... 97

**Fig. 6-3.** Proportion of variance explained by marker based principal components. ..... 98

**Fig. 6-4.** First two principal components using 2,735 SNP markers run on 231 wheat lines. Each solid circle represents one line. The colors correspond to the two subpopulations identified from hierarchical clustering. ..... 98

**Fig. 6-5.** Genome-wide LD ($r^2$) values plotted against genetic distance (cM) for the 231 wheat lines. The green horizontal line indicates the $95^{th}$ percentile of the distribution of unlinked $r^2$ while the fitted curve (red line) indicates the LD decay. ..... 99

**Fig. 6-6.** Q-Q plots comparing the distribution of observed versus expected *P*-values for association analyses using five statistical models. (A) heading date, (B) plant height, (C) maturity, (D) grain yield, (E) test weight, (F) thousand-kernel weight, (G) grain protein, (H) falling number, (I) sedimentation volume. ..... 101

**Fig. 7-1.** Frequency distributions of six agronomic traits measured in the mapping population (selection candidates). Data were averaged across F6 to F8 generations and all environments. The values for the check cultivars (AC Barrie, CDC Utmost, CDC Plentiful, and Pasteur) are indicated with arrows. ..... 113

**Fig. 7-2.** The distribution of QTL identified for six agronomic traits based on the combined data across five environments. QTL are displayed on the right side of each chromosome with vertical bars showing the QTL confidence interval defined by 1-LOD drop. The QTL identified for each trait are indicated in different colors. ..... 116

**Fig. 8-1.** Heat map and dendrogram of a genomic relationship matrix estimated using the EMMA algorithm based on 16K SNPs among the 304 wheat lines and three parents. Color codes show groups of lines based on their genomic relationships. Both rows and columns represent the lines. (A) Lines that clustered with Pasteur or none of the parents, (B) lines that clustered with CDC Utmost and CDC Plentiful. ..... 136

## List of Abbreviations

| | |
|---|---|
| AFLP | Amplified fragment length polymorphism |
| ANOVA | Analysis of variance |
| BL | Bayesian Lasso |
| BLUP | Best linear unbiased prediction |
| BRR | Bayesian ridge regression |
| CDC | Crop Development Centre, University of Saskatchewan |
| CF | Call frequency |
| CNHR | Canada Northern Hard Red |
| cM | CentiMorgam |
| CPSR | Canada Prairie Spring Red |
| CPSW | Canada Prairie Spring White |
| CTAB | Cetyl trimethylammonium bromide |
| CWAD | Canada Western Amber Durum |
| CWES | Canada Western Extra Strong |
| CWHWS | Canada Western Hard White Spring |
| CWRS | Canada Western Red Spring |
| CWRW | Canada Western Red Winter |
| CWSWS | Canada Western Soft White Spring |
| DArT | Diversity array technology |
| DH | Double haploid |
| DNA | Deoxyribonucleic acid |
| ECs | Environmental covariates |
| FDR | False discovery rate |
| FN | Falling number |
| G-BLUP | Genomic best linear unbiased prediction |
| GBS | Genotyping-by-sequencing |
| GEBVs | Genomic estimated breeding values |
| GRM | Genomic relationship matrix |
| GS | Genomic selection |
| GWAS | Genome-wide association study |
| $h^2$ | Heritability (narrow sense) |

| | |
|---|---|
| $H^2$ | Heritability (broad sense) |
| HD | Heading date |
| HT | Plant height |
| Lasso | Least absolute shrinkage and selection operator |
| LD | Linkage disequilibrium |
| LOD | Logarithm of odds |
| MAF | Minor allele frequency |
| MAS | Marker assisted selection |
| MAT | Maturity |
| MCMC | Markov chain Monte Carlo |
| MLM | Mixed linear model |
| MSE | Mean-square error |
| OLS | Ordinary least square |
| PCA | Principal Component analysis |
| PCR | Polymerase chain reaction |
| PCs | Principal components |
| RAPD | Random amplified polymorphic DNA |
| RFLP | Restriction fragment length polymorphism |
| PRO | Grain protein content |
| QTL | Quantitative trait loci |
| RILs | Recombinant inbred lines |
| RKHS | Reproducing kernel Hilbert spaces regression |
| RKHS-KA | Reproducing kernel Hilbert spaces regression with kernel averaging |
| RR-BLUP | Ridge regression best linear unbiased prediction |
| SDS | Sedimentation volume |
| SC | Selection candidates |
| SNP | Single nucleotide polymorphism |
| SSR | Simple sequence repeats |
| TP | Training population |
| TKW | Thousand-kernel weight |
| TWT | Test weight |
| YLD | Grain yield |

## 1. Introduction

Wheat (*Triticum aestivum* L.) is an important cereal crop that accounts for more than 20% of the total calories consumed by humans globally and is a staple food for about 35% of the world's population (Breiman and Graur, 1995). Canada is the sixth largest wheat producing country in the world with a total production of 29.3 million tonnes in 2014 (FAOSTAT, 2015). In Canada, most of the wheat production is in the prairie provinces of Alberta, Saskatchewan and Manitoba, and a small proportion is grown in British Columbia and eastern Canada (McCallum and DePauw, 2008; Randhawa et al., 2013). Canada is the second largest exporter of wheat after the United States of America; 19.8 million of the 37.5 million tonnes of wheat grain produced in Canada in 2013 was exported (FAOSTAT, 2015). Canadian wheat is recognized globally for its high-quality end-use properties.

Wheat yields have increased to keep up with rising food demands from population growth (Gustafson et al., 2009). The increase in grain yield can be attributed to several factors but two of the commonly cited factors include the adoption of new cultivars and improved management practices (Rudd, 2009). Efforts to increase wheat grain yield must continue to meet projected food demands from a growing world population and increasing challenges from climate change, resource limitations and incidence of biotic and abiotic stresses. A large proportion of the anticipated yield increase will likely come from efforts in plant breeding to develop high yielding, stress tolerant and disease resistant cultivars with acceptable quality standards. This calls for the integration of conventional breeding approaches with innovative and new strategies to accelerate the breeding cycle and improve the precision and efficiency of selection strategies.

Wheat breeding involves the creation of new genetic variability through controlled hybridization of two or more parents followed by self-crossing and advancing generations by selecting offspring with desirable agronomic, disease resistance and end-use quality traits. These advanced wheat lines then undergo repeated field testing, and if they meet appropriate standards are released as new cultivars. This process normally takes 10 to 15 years and is resource intensive. Traditionally, selection of desirable plants within segregating populations is based on visual assessment of agronomic traits and laboratory tests of end-use quality traits, which are laborious and expensive. For quantitative traits, selection based on phenotype alone is subject to confounding effects from G × E so entries are evaluated over multiple locations and years. This makes phenotypic selection time consuming and expensive. Moreover, the short growing season of the

Canadian prairies (a frost-free period of 90 to 120 days) presents a challenge for large scale field evaluation and selection of a breeding material. The advent of molecular marker systems greatly improved the precision and speed of the breeding process through marker assisted breeding. When markers that are genetically linked to target genes are identified, they can be used for MAS. Marker assisted selection was developed to overcome the limitations of conventional breeding by changing the selection criteria from phenotypes to genes, either directly or indirectly (Francia et al., 2005).

The most commonly used methods for marker-trait association analysis have been QTL mapping with experimental populations and association mapping using natural populations. Quantitative trait loci mapping involves linking QTL underlying trait variation with known molecular markers in a segregating population developed through hybridization of inbred parental lines that are genetically variable for one or more target traits (Mackay, 2001). Association mapping is a method that relies on linkage disequilibrium (LD) to study the relationship between phenotypic variation and genetic polymorphisms across a set of germplasms with wide genetic diversity (Flint-Garcia et al., 2003). Association mapping exploits historic recombination and natural genetic diversity for high resolution mapping (Zhu et al., 2008). Linkage based QTL mapping has relatively lower resolution than association mapping because there are fewer recombination events (Mackay, 2001). Both QTL and association mapping studies are used to detect markers significantly associated with QTL underlying trait variation that can be used for MAS. Moreover, the genetic architecture revealed from these methods can be used to enhance genomic prediction of quantitative traits.

Several QTL for disease resistance, agronomic and end-use quality traits have been identified and molecular markers linked with these QTL were deployed for selection of these traits in several wheat cultivars released for commercial cultivation in Canada (Cuthbert et al., 2006; Knox et al., 2009; Randhawa et al., 2013; Wiebe et al., 2010). When few QTL that explain a large proportion of the variance of a quantitative trait are identified, the breeding strategy is to introgress/pyramid these QTL into elite germplasm through MAS (Bernardo, 2008). However, most traits of agronomic importance in wheat are quantitatively inherited and are regulated by many genes with small effects. For such traits, MAS has limited application because of the difficulty to identify and manipulate multiple genomic regions at the same time (Francia et al., 2005). Even when major genes or QTL underlie a trait, a large portion of the genetic variance may

be due to several minor effects QTL and introgression of the major genes or QTL will not capture the effects of minor QTL (Bernardo, 2014).

Advances in high-throughput genotyping technologies have resulted in the availability of abundant molecular markers covering the whole genome in many species. This new development has allowed the use of whole genome dense marker maps for the prediction of breeding values of individuals without phenotypic records (Heffner et al., 2009; Meuwissen, 2009; Meuwissen et al., 2001). This approach is commonly referred to as genomic selection or genome-wide selection (Meuwissen et al., 2001). Genomic selection is a type of MAS but is unique in that there is no need to identify marker-trait associations (Bernardo and Yu, 2007; Meuwissen, 2007). Genomic selection involves the estimation of marker effects from a training or reference population (TP) that has both genotypic and phenotypic data to predict GEBVs of selection candidates (SC) by combining their marker genotypes with the estimated marker effects (Meuwissen, 2009). These GEBVs are used to select breeding lines that should be removed or retained in future crosses (Jannink et al., 2010). The main assumption of GS is that by using dense genome-wide markers, all QTL will be in LD with at least one nearby marker and potentially all the genetic variance can be explained by markers (Calus, 2010; Goddard and Hayes, 2007).

Several approaches and models have been proposed for implementing GS (de los Campos et al., 2009a; de los Campos et al., 2010; Gianola et al., 2006; Gianola and van Kaam, 2008; Habier et al., 2011; Meuwissen et al., 2001; Park and Casella, 2008; VanRaden, 2008). Originally, GS models focused on prediction of a single trait evaluated in a single environment or averaged across environments. In recent years, multi-environment models that account for G × E (Burgueño et al., 2011; Burgueño et al., 2012; Crossa et al., 2015; Cuevas et al., 2017; Heslot et al., 2014; Jarquín et al., 2014a; Jarquín et al., 2017; Lopez-Cruz et al., 2015; Pérez-Rodríguez et al., 2015; Sukumaran et al., 2017; Technow et al., 2015) and multiple-trait GS models (Aguilar et al., 2011; Calus and Veerkamp, 2011; Guo et al., 2014a; Hayashi and Iwata, 2013; Hayes et al., 2017; Jia and Jannink, 2012; Jiang et al., 2015) have been proposed. Moreover, models that combine the results of genome-wide association study (GWAS) with GS have been proposed (Bentley et al., 2014; Bernardo, 2014; Spindel et al., 2016; Zhang et al., 2014; Zhao et al., 2014). These studies reported improved prediction accuracy using multi-environment and multiple-trait analysis. The performance of single and multiple-trait prediction models, models that account for G × E, and

methods that combine GWAS with GS have not been evaluated using the same cross-validation folds in wheat.

Simulation and empirical studies from plant and animal breeding programs indicated that GS accuracy depends on several factors including the marker type and density (Combs and Bernardo, 2013; Heffner et al., 2011a; Moser et al., 2010), TP size (Calus and Veerkamp, 2007; Combs and Bernardo, 2013; Heffner et al., 2011a; VanRaden et al., 2009), trait heritability (Combs and Bernardo, 2013; Moser et al., 2010), genetic relationship between the TP and SC (Solberg et al., 2008; VanRaden et al., 2009; Wientjes et al., 2013), population structure (de los Campos et al., 2015; de Roos et al., 2009; Guo et al., 2014b; Isidro et al., 2015), G × E (Burgueño et al., 2012; Crossa et al., 2015; Jarquín et al., 2014a; Jarquín et al., 2017; Lopez-Cruz et al., 2015; Pérez-Rodríguez et al., 2015; Sukumaran et al., 2017) and the statistical method used for prediction (Calus, 2010). These factors are interrelated in a complex manner (Desta and Ortiz, 2014). Genetic relationship between the TP and SC was cited as one of the factors that affect prediction accuracy. However, it is not clear what measure of genetic relationship is appropriate and the extent of genetic relationship that is sufficient to obtain an acceptable level of accuracy. Moreover, most of these factors are population and environment specific and it is important to assess different statistical models and model parameters for their predictive ability in different breeding populations and environments.

In recent years, GS is gaining momentum in crop breeding programs. Despite a growing interest, there remain uncertainties in the practical application of GS for crop improvement. Most studies in GS evaluated model prediction accuracy based on simulated data but increasing numbers of studies are now reporting empirical evidence, with mixed results. Model prediction accuracy is commonly evaluated through a cross-validation approach by systematically partitioning the same population into training and validation folds or based on k-fold cross-validation methods (Lorenz et al., 2011). However, successful implementation of GS in crop breeding programs largely depends on its potential to accurately estimate GEBVs of individuals in a population different from the one used to estimate marker effects. Moreover, it is often difficult to extrapolate the results reported in previous studies because the statistical methods, model parameters, population characteristics, and environments are variable.

This research was conducted to evaluate different GS approaches, statistical models and model parameters to design selection strategies for complex traits in wheat under the short growing

seasons of western Canada. The primary hypothesis was that GS has the potential to accurately predict GEBVs for wheat lines and facilitate rapid gains from selection. The objectives of this study were to i) evaluate the performance of various statistical approaches and models to predict agronomic and end-use quality traits using empirical data in spring bread wheat, ii) determine the effects of TP size, marker density, and population structure on genomic prediction accuracy, iii) examine GS prediction accuracy when modelling G × E using different approaches, iv) detect marker-trait associations for agronomic and end-use quality traits in spring bread wheat, v) evaluate the effects of TP composition, cross-validation technique, and genetic relationship between the TP and SC on GS accuracy, and vi) compare genomic and phenotypic prediction accuracy.

## 2. Literature Review

### 2.1 Wheat Production

Wheat is the most widely grown cereal crop in the world, with a total harvested area of 221.6 million ha in 2014 (FAOSTAT, 2015). In 2014, wheat ranked third in production (729 million tonnes) among the world grain crops after maize (*Zea mays*) and rice (*Oryza sativa*) (FAOSTAT, 2015). Wheat is a staple food for about 35% of the world's population and accounts for more than 20% of the total calories consumed globally (Breiman and Graur 1995). It is grown throughout the temperate, tropical and sub-tropical parts of the northern and southern hemispheres (Zohary and Hopf, 2000). Canada is the sixth largest wheat producing country in the world with the total production of 29.3 million tonnes in 2014 (FAOSTAT, 2015). In Canada, nearly all the wheat is grown in the prairie provinces of Alberta, Saskatchewan and Manitoba, and a small proportion is grown in British Columbia and eastern Canada (McCallum and DePauw, 2008; Randhawa et al., 2013). Both spring wheat, planted in the spring and harvested in late summer or early fall, and winter wheat, planted in the fall and harvested in summer, are grown in Canada. Of the total wheat produced in Canada in 2016, spring hexaploid wheat accounted for 62%, winter hexaploid wheat accounted for 13%, and durum wheat accounted for 25% (Statistics Canada, 2017). Canada is the second largest exporter of wheat after the United States of America; 19.8 million of the 37.5 million tonnes of wheat grain produced in Canada in 2013 was exported (FAOSTAT, 2015).

Canadian wheat is classified into different market classes based on end-use quality parameters of grain protein content, gluten strength, and kernel colour (DePauw et al., 2011). Currently, the wheat grown in western Canada is classified into nine milling classes, including Canada Northern Hard Red (CNHR), Canada Prairie Spring Red (CPSR), Canada Prairie Spring White (CPSW), Canada Western Amber Durum (CWAD), Canada Western Extra Strong (CWES), Canada Western Hard White Spring (CWHWS), Canada Western Red Spring (CWRS), Canada Western Red Winter (CWRW), and Canada Western Soft White Spring (CWSWS) (Canadian Grain Commission, 2015). Each of these classes have unique characteristics and end-uses. CWRS and CWAD are the predominant classes of wheat grown in western Canada (McCallum and DePauw, 2008). The CWRS cultivars have a premium price in world trade due to their superior milling and baking quality under different manufacturing conditions (McCallum and DePauw, 2008). The CWAD cultivars are mainly used for semolina production for making pasta and

couscous. Breeding efforts in western Canada developed CWAD cultivars with high semolina yield, high yellow pigment content, strong gluten and low cadmium content which led to durum wheat selling at a price premium over CWRS wheat (McCallum and DePauw, 2008).

### 2.2 Evolution of Wheat

Wheat is one of the first domesticated food crops. While some species of wheat contain a single genome, others have multiple homoeologous genomes that resulted from natural hybridization. Diploid genomes from ancestral grasses that have been involved in hybridization events are labelled using letter codes, with 'A', 'D', and 'S' genomes. There are two species of wheat at the diploid level, *Triticum monococcum* L. ($A^mA^m$), which is also called einkorn wheat and *Triticum urartu* (AA) (Nevo et al., 2002). *Triticum monococcum* includes cultivated sub-species (ssp.) *monococcum* and wild ssp. *aegilopoides*, while *T. urartu* exists only in its wild form (Nevo et al., 2002). The diploids are the most primitive wheats and have limited range of morphology (Riley, 1975).

There are two species of wheat at the tetraploid level, *T. turgidum* L. (BBAA) and *T. timopheevi* (GGAA), each having cultivated and wild ssp. (Riley, 1975). The wild forms in *T. turgidum* L. and *T. timopheevi* are designated ssp. *dicoccoides* and *araraticum*, respectively (Riley, 1975). The cultivated tetraploids include, ssp. *dicoccum*, *timopheevi*, *durum*, *turgidum*, *polonicum*, *carthlicum* and *orientale* (Riley, 1975). It is believed that the cultivated forms in the diploid and tetraploid groups are derived from their respective wild forms (McFadden and Sears, 1946). Studies have shown that the 'A' genome of the polyploid wheats are equivalent to that of *T. urartu* (Dvorak et al., 1988; Petersen et al., 2006). The 'B' genome in *T. turgidum* and the 'G' genome in *T. timopheevi* are believed to have originated from an annual diploid species *Aegilops speltoides* (genome type: S) or its close relatives (Dvorak and Zhang, 1990). The 'B' and 'G' genomes are widely considered to be modified 'S' genomes that have undergone massive changes following polyploid formation (Gustafson et al., 2009). *T. turgidum* ssp. *durum* (durum wheat) is the second most commonly grown domesticated wheat and it is used for making pasta and couscous.

Hexaploid wheat also has two species, *T. aestivum* L. (BBAADD) and *T. zhukovskyi* ($GGAAA^mA^m$), which are both cultivated forms (Riley, 1975). *T. zhukovskyi* is morphologically similar to the tetraploid wheat *T. timopheevi* and it is believed to have originated following hybridization involving *T. timopheevi* and *T. monococcum* (Riley, 1975). *T. aestivum* L. is believed

to have originated in the Caspian Sea region about 9,000 years ago from a hybridization between domesticated emmer wheat (*T. turgidum* ssp. *dicoccum*) and the diploid *Aegilops tauschii* (DD), which contributed the D genome (Peng et al., 2011). *T. aestivum* ssp. *vulgare* is commonly known as bread wheat and the other ssp. in this group are *compactum*, *sphaerococcum*, *spelta*, *macha*, and *vavilovii* (Riley, 1975). *T. aestivum* ssp. *vulgare* is the most commonly grown domesticated wheat and it is valued for the baking of high rising bread.

### 2.3 Molecular Markers in Wheat Breeding

Markers can be classified into three broad classes: 1) morphological markers based on visually assessable traits, 2) biochemical markers based on allelic variants of enzymes (isozymes), and 3) molecular markers based on Deoxyribonucleic acid (DNA) assay (Collard et al., 2005; Koebner and Summers, 2003). Morphological markers are simple but seldom used because most phenotypes are determined by allelic variation at more than one locus and are also affected by the environment (Koebner and Summers, 2003; Winter and Kahl, 1995). Biochemical markers are single-locus-based but are rare because each marker requires its own biochemical assay (Koebner and Summers, 2003). Moreover, the expression of biochemical markers is often restricted to specific developmental stages or tissues (Winter and Kahl, 1995). Molecular markers are the most widely used markers since their discovery in the 1980s. Molecular markers score different types of sequence variation, such as single nucleotide polymorphisms (SNPs), rearrangements, insertions and deletions, or length differences (Paterson, 1996). Molecular markers have advantages since they are not influenced by the environment and can be detected at any stage of the plant development.

Over the years, several DNA-based marker analysis methods have been developed. Hybridization-based restriction fragment length polymorphisms (RFLP) were the first to be discovered and applied in human genome mapping (Botstein et al., 1980), and subsequently applied for linkage mapping in plants (Helentjaris et al., 1985; Paterson et al., 1988). Restriction fragment length polymorphisms detect DNA polymorphisms as the difference in the length of DNA fragments after digestion of DNA with sequence specific endonucleases (Botstein et al., 1980). Restriction fragment length polymorphisms are unlimited in number and are not confounded by polyploidy; they are co-dominant markers and are able to detect individual loci in each of the three genomes of hexaploid wheat simultaneously (Chao et al., 1989). However, RFLPs

have been relatively less useful in wheat because they are time-consuming and labor-intensive (Gupta et al., 1999).

With advances in biotechnology, other markers systems based on the polymerase chain reaction (PCR) were developed (Saiki et al., 1988). Some of the commonly used PCR-based markers include random amplified polymorphic DNA (RAPD) based on amplification of random DNA segments with single primers of arbitrary nucleotide sequence (Williams et al., 1990), amplified fragment length polymorphism (AFLP) based on selective PCR amplification of restriction fragments from a total digest of genomic DNA (Vos et al., 1995), and amplification of microsatellites or simple sequence repeats (SSR) (Weber and May, 1989). Compared to RFLPs, PCR-based markers offer the potential to reduce the time, effort and expense required for molecular mapping (Gupta et al., 1999). However, in wheat, RAPD technology has not been widely used due to the complexity of the large wheat genome, low level of polymorphisms, and lack of reproducibility of results (Gupta et al., 1999). Similarly, AFLP markers were not widely used in molecular breeding due to the lengthy and laborious detection method which was not amenable to automation (Mammadov et al., 2012). Microsatellite markers were the most popular and useful molecular markers for wheat before the discovery of SNP markers because they are abundant, easy to detect, dispersed throughout the genome, and show higher levels of polymorphism relative to RFLP and RAPD markers (Gupta et al., 1999; Langridge et al., 2001).

As more DNA sequence information became available, new marker systems that detect SNPs were developed (Chee et al., 1996; Wang et al., 1998). Single nucleotide polymorphisms are variations detected at the level of a single nucleotide base in the genome, which are the most abundant source of variation in plant and animal genomes (Xu and Crouch, 2008). Single nucleotide polymorphisms are the marker system of choice in most plant species due to their abundance and amenability to high-throughput automation (Mammadov et al., 2012). In wheat, high-density SNP arrays, such as the 9K (Cavanagh et al., 2013), 90K (Wang et al., 2014a), and 820K (Winfield et al., 2016) gene associated SNPs, have provided an enormous opportunity to dissect complex traits and advance marker assisted breeding. Moreover, advances in next-generation sequencing (NGS) technologies (Shendure and Ji, 2008), have made whole genome sequencing feasible in many species. This has also led to genotyping-by-sequencing (GBS), which combines SNP discovery and genotyping (Elshire et al., 2011; He et al., 2014). Genotyping-by-sequencing uses restriction enzymes to capture a reduced representation of the target genome

which is sequenced by NGS platforms (Poland et al., 2012). Currently, several marker systems are used in wheat, with several emerging high-throughput systems producing dense marker data sets that span the whole genome.

### 2.4 Traditional Applications of Molecular Markers in Wheat: Mapping and MAS

One of the main challenges in modern genetic analysis is determining the genetic basis of quantitative trait variation. The regions within genomes that contain genes associated with a quantitative trait are called quantitative trait loci (Collard et al., 2005). Two of the most commonly used methods to dissect complex traits in plants are linkage analysis (QTL mapping) and association mapping, also called LD mapping (Zhu et al., 2008). Linkage analysis has been used to map qualitative and quantitative traits. In plants, linkage analysis is normally conducted by establishing experimental (biparental) mapping populations such as F2 populations, backcross populations, RILs, and doubled haploids (Collard et al., 2005; Flint-Garcia et al., 2003). Linkage analysis in these populations detects only those QTL that are polymorphic in the population (Bernardo, 2008). Association mapping is a method that relies on LD to study the relationship between phenotypic variation and genetic polymorphisms across a set of germplasms with wide genetic diversity (Flint-Garcia et al., 2003). The main principle of association mapping is that LD tends to be maintained over many generations between loci that are genetically linked to one another (Neumann et al., 2011). Linkage disequilibrium, also known as gametic phase disequilibrium, is the non-random association of alleles at different loci, which in random mating populations is generated by mutation and genetic drift, and decays by recombination (Breseghello and Sorrells, 2006a). Two common approaches for association analysis are candidate gene association mapping and genome-wide association mapping. Candidate gene association mapping tests the relation between DNA polymorphism of a candidate gene with the trait of interest; whereas, genome-wide association mapping scans the whole genome for casual genetic variation using dense genome-wide markers (Rafalski, 2002; Zhu et al., 2008). Unlike linkage analysis, association mapping evaluates genetic diversity across natural populations to identify polymorphisms that correlate with phenotypic variation (Flint-Garcia et al., 2003). The advantage of association mapping over linkage analysis is that it uses a more diverse population, and therefore examines a broader set of genetic variation for marker-trait correlations; this results in enhanced mapping resolution and broader allele coverage (Abdurakhmonov and Abdukarimov, 2008;

Neumann et al., 2011; Zhu et al., 2008). Association mapping panels can also be used to study several traits within a breeding program and can save time and money because there is no need to develop a biparental mapping population for each trait (Abdurakhmonov and Abdukarimov, 2008; Neumann et al., 2011; Zhu et al., 2008). However, association analysis may lead to a high frequency of false-positive associations due to population structure and cryptic relatedness that may arise from the origins and history of the populations used for mapping. Therefore, statistical methods that account for multiple levels of relatedness need to be used to detect true associations. Both linkage and association mapping studies are useful to identify genomic regions associated with traits and are often used together as complementary approaches.

Many agriculturally important genes have been mapped using different types of molecular markers. Molecular markers that are genetically linked to target genes can be used for MAS. Marker assisted selection can accelerate the breeding cycle because it can be carried out at the seedling stage or on single plants in early generations (Collard and Mackill, 2008). Marker assisted selection also improves the efficiency and precision of conventional plant breeding by indirectly selecting molecular markers linked to target genes (Collard and Mackill, 2008; Gupta et al., 1999). Molecular markers are also important for marker-assisted evaluation of breeding material, marker-assisted backcrossing, and pyramiding of several disease resistance genes into a single cultivar (Collard and Mackill, 2008). Marker assisted evaluation of breeding material includes confirming the identity of cultivars, assessing genetic diversity, parent selection, and confirmation of hybrids (Collard and Mackill, 2008). In wheat, several QTL have been identified that affect disease resistance, agronomic performance, and end-use quality traits (Cuthbert et al., 2006; Cuthbert et al., 2008; Knox et al., 2009; McCartney et al., 2005). However, markers identified in preliminary genetic mapping studies require further testing and development before they can be used in MAS (Collard et al., 2005). Further development includes high resolution mapping, marker validation in different genetic backgrounds, and marker conversion when there are problems with reproducibility or when the marker technique is complicated, time consuming and expensive (Collard et al., 2005). High resolution mapping identifies markers that are closely linked with the desired trait (Mohan et al., 1997). Suitable markers for MAS should be tightly linked to the target loci (preferably less than 5 cM genetic distance), highly polymorphic in the breeding material, and amenable to high-throughput detection methods that are cost effective (Collard and Mackill, 2008).

Genetic analysis in wheat is challenged by the size and complexity of the wheat genome; however, recent advances in molecular techniques, NGS, and bioinformatics tools have accelerated marker discovery and analysis of the wheat genome. To date, more than 30 genes have been cloned in common wheat and its relatives, and 97 functional (gene-specific) markers for wheat agronomic, disease resistance and end-use quality traits have been developed (Liu et al., 2012). Most of these markers are available in the public domain (http://maswheat.ucdavis.edu) and can be used for MAS. Functional markers are derived from polymorphic sites within gene coding sequences that cause phenotypic trait variation (Bagge and Lübberstedt, 2008; Varshney et al., 2005). These markers can be used for MAS in different genetic backgrounds without revalidating the marker-QTL associations (Varshney et al., 2005). Unlike random markers linked to a locus, functional markers are more reliable for MAS because there is no recombination between the marker and the target locus (Bagge and Lübberstedt, 2008). Currently, international efforts are underway to complete a high-quality genome sequence of all 21 chromosomes in wheat (http://www.wheatgenome.org/). An ordered draft sequence of the 17-gigabase hexaploid wheat genome has been produced and 124,201 gene loci have been annotated across the homologous chromosomes (International Wheat Genome Sequencing Consortium, 2014). Moreover, more than 3.6 million marker loci have been identified; these include all publicly available molecular markers, insertion site-based polymorphism and SNP markers identified from recent whole-genome shotgun and transcriptome sequencing, and GBS tags that have been mapped on individual chromosomes of the bread wheat genome (International Wheat Genome Sequencing Consortium, 2014). Recently, Avni et al. (2017) reported a 10.1-gigabase assembly of the genome of wild emmer wheat (*T. turgidum* ssp. *dicoccoides*), a tetraploid wheat known to be the direct ancestor of economically important wheats. This allowed detection of casual mutations in genes controlling shattering, an important domestication trait in wheat (Avni et al., 2017). This shows that the advent of fully assembled genome will facilitate the discovery of additional genes and markers associated with important traits which will enhance the precision and efficiency of MAS in wheat.

In plant breeding, MAS has been successful to improve traits controlled by few QTL with large effects. When few QTL with large effects explain much of the variation of a quantitative trait, the breeding strategy is to find and introgress/pyramid these QTL into elite cultivars through MAS (Bernardo, 2008). However, when a trait is controlled by many QTL with small effects, MAS has limited application because estimates of effects for minor QTL are often unreliable

(Bernardo, 2008). Most agriculturally important traits in cereals are quantitatively inherited, which makes identification of the genes underlying variation for these traits difficult (Neumann et al., 2011). In this case, it is necessary to utilize all QTL affecting the trait in MAS (Meuwissen et al., 2001). The methodology for this approach, called genomic selection, was first proposed by Meuwissen et al. (2001).

### 2.5 Prospects for Genomic Selection

Traditional MAS has been successful at identifying and selecting major effect QTL; however, complex traits involving many small effect QTL are more difficult to monitor using MAS or phenotypic selection alone, and GS offers an attractive solution. Genomic selection differs from the traditional MAS strategies in that instead of only using markers that have a significant association with a trait based on a predefined significance threshold, all markers are used to predict GEBVs for individuals without phenotypic records (Heffner et al., 2009; Meuwissen et al., 2001). The GEBVs are calculated as the sum of the effects of markers or marker haplotypes across the entire genome, thereby potentially capturing all the QTL that underlie trait variation (Hayes et al., 2009c). Fitting all markers simultaneously avoids multiple testing and bias when estimating the effects of minor and major effect QTL (Jia and Jannink, 2012). The advantage of GS is its potential to predict GEBVs with an accuracy that is sufficient for selection over several generations without repeated phenotyping, which reduces the cost and generation intervals (Habier et al., 2007). Heffner et al. (2010) indicated that GS can dramatically accelerate genetic gain through short breeding cycles if moderate selection accuracies can be achieved. Genomic selection uses a training population that has both genotypic and phenotypic data to develop a statistical model that takes genotypic data from a candidate population and calculates GEBVs of untested individuals (Heffner et al., 2011a; Jannink et al., 2010). These GEBVs say nothing of the function of the underlying genes but selection of new breeding lines is based on these values (Jannink et al., 2010). Usually a unique set of individuals, commonly called a validation population, that have phenotypic and genotypic data are used to assess the predictive performance of the models. Genomic estimated breeding values are predicted for individuals in the validation population and the correlation of these values to the actual phenotype is considered as prediction accuracy of the model. In theory, the prediction accuracy is the correlation between the true breeding value (TBV) and GEBVs. However, the TBV is known only in simulated data and in real data sets the actual phenotype is

used to measure prediction accuracy (Charmet and Storlie, 2012). Together, GS is an emerging tool that provides several advantages over traditional MAS.

In recent years, there is a growing interest to implement GS in crop breeding programs; however, there is still limited information regarding the practical application of GS for crop improvement. Since the first simulation study that reported prediction accuracies in the range of 0.73 to 0.85 (Meuwissen et al., 2001), the application of GS has been evaluated in animal breeding, plant breeding and human genetic studies using both simulated and empirical data. The potential of GS was evaluated in a number of annual and perennial plant species such as maize (Albrecht et al., 2011; Bernardo and Yu, 2007; Bernardo, 2009; Beyene et al., 2015; Mendes and de Souza, 2016; Windhausen et al., 2012; Zhao et al., 2012); rice (Grenier et al., 2015; Onogi et al., 2016; Spindel et al., 2015; Xu et al., 2014); wheat (Charmet and Storlie, 2012; Charmet et al., 2014; Crossa et al., 2014; Crossa et al., 2016; Daetwyler et al., 2014; Dawson et al., 2013; de los Campos et al., 2009b; Gianola et al., 2011; He et al., 2016; Heffner et al., 2011a; Huang et al., 2016; Lado et al., 2016; Longin et al., 2015; Michel et al., 2016; Pérez-Rodríguez et al., 2012; Poland et al., 2012; Rutkoski et al., 2015; Rutkoski et al., 2012; Rutkoski et al., 2011; Rutkoski et al., 2014; Storlie and Charmet, 2013); barley (Iwata and Jannink, 2011; Zhong et al., 2009); oat (Asoro et al., 2011); rye (Wang et al., 2014b); rapeseed (Würschum et al., 2014); pea (Burstin et al., 2015); soybean (Jarquín et al., 2014b); alfalfa (Annicchiarico et al., 2015); intermediate wheatgrass (Zhang et al., 2016); sugarcane (Gouy et al., 2013); sugar beet (Würschum et al., 2013); tomato (Yamamoto et al., 2016); apple (Kumar et al., 2012a; Kumar et al., 2012b); cassava (de Oliveira et al., 2012; Ly et al., 2013); oil palm (Wong and Bernardo, 2008); loblolly pine (Resende et al., 2012b); white spruce (Beaulieu et al., 2014) and eucalyptus (Denis and Bouvet, 2013; Resende et al., 2012a). These studies reported that GS has a potential to accelerate the breeding cycle and increase genetic gain.

### 2.6 Statistical Methods for Genomic Selection

Genomic selection has been increasing in popularity; as a result, numerous statistical methods have been developed to estimate marker effects and compute GEBVs. These methods differ primarily in the way the marker data are weighted in the models and in their underlying assumptions about the variance of marker effects. In GS models, phenotypes are regressed on genome-wide markers, which is often challenging because the number of markers ($p$) greatly

exceeds the number of phenotyped individuals (*n*). When *p* is large relative to *n* (*p* >> *n*), it may result in overfitting of the model and poor prediction of GEBVs (de los Campos et al., 2013; Pérez and de los Campos, 2014). If marker effects are fitted as fixed effects, *p* >> *n* increases the variance of estimates and the pooled experimental error (i.e. mean square error; MSE) (de los Campos et al., 2013). To overcome this problem, whole genome regression models use estimation methods that perform variable selection, shrinkage of estimates, or a combination of both to reduce the dimensions of the marker data (de los Campos et al., 2013). These procedures introduce bias but reduce the variance of estimates and when *p* >> *n*, the use of shrinkage yields smaller MSE than that of standard estimation procedures such as ordinary least squares (OLS), or maximum likelihood (de los Campos et al., 2013). In GS, marker genotypes are treated as random variables and prior assumptions are made about the variance explained due to their effects (Clark and van der Werf, 2013). The various GS models can be grouped into shrinkage models, variable selection models, and kernel methods (Lorenz et al., 2011). The three groups of models are described below.

### 2.6.1 Ridge regression best linear unbiased prediction

Ridge regression best linear unbiased prediction (RR-BLUP), also called ridge regression or best linear unbiased prediction (BLUP), was one of the first models proposed for the prediction of total genetic value using genome-wide markers (Meuwissen et al., 2001). This method was proposed to overcome the limitations of OLS analysis where the effects of loci are set either to zero or their full effect, based on whether they are below or above an arbitrarily chosen significance threshold (Meuwissen et al., 2001). Ridge regression was first proposed by Whittaker et al. (2000) for genomic prediction as an alternative to variable selection procedures and provided a method that includes all markers in a regression model. Whittaker et al. (2000) introduced a penalty parameter 'λ' in an OLS estimator that shrinks marker effects uniformly towards zero assuming every marker has equal contribution to the genetic variance. The assumption that all loci explain equal amount of variance requires estimation of only the total genetic variance and the variance per locus is obtained by dividing the total variance by the number of loci (Meuwissen et al., 2001). Although this assumption is unrealistic, RR-BLUP performs well when predictions are made for traits controlled by many loci with small effects (Lorenz et al., 2011). However, Gianola (2013) reported that in RR-BLUP, shrinkage is allele frequency and sample size dependent but effect-size independent. There is some differential shrinkage in small sample sizes, but with little or no

differential shrinkage otherwise unless alleles are rare (Gianola, 2013). For fixed sample sizes, BLUP performs less shrinkage of markers that have intermediate allelic frequencies (Gianola, 2013).

Genetic relationships between individuals can be estimated from SNPs and can readily be incorporated into BLUP models (Habier et al., 2013; VanRaden, 2008). In this case, the genomic relationship matrix (GRM), which estimates the realized proportion of the genome that is shared by two individuals, is used to predict the genetic merit of individuals (Goddard et al., 2011). This method was first proposed by VanRaden (2008) and is referred to as genomic BLUP (G-BLUP). G-BLUP is equivalent to RR-BLUP, but it is computationally more efficient because the dimension of the marker data is reduced in the mixed model equations (Clark and van der Werf, 2013).

### 2.6.2 Bayesian regression models

The Bayesian approach was proposed for GS to overcome the limitation of homogenous shrinkage of marker effects, a characteristic of BLUP. The main difference between BLUP and the Bayesian models is the prior distribution for the variance of marker effects (Asoro et al., 2011). In Bayesian approach, the variance explained by each locus can vary and is assumed to come from a prior distribution (Meuwissen et al., 2001). This allows it to perform marker specific shrinkage of estimates by specifying an appropriate prior density. This assumption agrees with the fact that some chromosome segments contain QTL with large effects, some contain QTL with small effects, and some have no QTL. The prior density of marker effects determines the extent and type of shrinkage induced and whether the model will induce variable selection and shrinkage or shrinkage only (de los Campos et al., 2013). In Bayesian methods, parameter estimates cannot be obtained analytically and Markov chain Monte Carlo (MCMC) sampling is commonly used to approximate parameter estimates through repeated sampling from their posterior distributions and compute appropriate summary statistics, such as the mean or median of the distributions (Kärkkäinen and Sillanpää, 2012; Lorenz et al., 2011). This process increases the computational time compared to ridge regression. There are several Bayesian regression models proposed to date for GS. Gianola et al. (2009) used the term ‘Bayesian alphabet’ to denote the various letters of the alphabet used to name these methods. The first Bayesian models proposed for GS were BayesA and BayesB (Meuwissen et al., 2001). Since then, several Bayesian models were developed that include

Bayesian Lasso (BL) (Park and Casella, 2008); Bayesian ridge regression (BRR) (Pérez et al., 2010); BayesCπ and Dπ (Habier et al., 2011); BayesR (Erbe et al., 2012); weighted Bayesian shrinkage regression (wBSR), which is a fast version of BayesB (Hayashi and Iwata, 2010); MCBayes and varBayes (Hayashi and Iwata, 2013), etc. Moreover, there are different models denoted by variants of the same letter such as Bayes-B1 and B2 (Zhong et al., 2009); BayesD0, D1, D2 and D3 (Wellmann and Bennewitz, 2012); EBL (Mutshinda and Sillanpää, 2010); BayesTA, TB and TCπ (Wang et al., 2013), etc. The main difference among these models is the assumption of the prior distribution of variances of marker effects. The assumptions of some of the Bayesian models assessed in this thesis are described below.

BayesA, BayesB and BayesCπ assign non-uniform variances to markers with different levels of effect. In BayesA, proportion of markers with no effect (π) are treated as zero so that all markers have non-zero effect and are included in the model but estimates of their effects are shrunk by assuming a normal distribution with mean of zero and locus-specific variances (Habier et al., 2011; Hayashi and Iwata, 2013). Whereas in BayesB, π is greater than zero to accommodate the assumption that many SNPs have no effect and are excluded from the model (Habier et al., 2011). Habier et al. (2011) indicated that BayesA and BayesB have drawbacks because these models treat π, which affects shrinkage of SNP effects, as known and proposed another model called BayesCπ in which π is treated as an unknown that is inferred from the data. In BayesB and BayesCπ it is assumed that only a fraction of loci of 1 – π contribute to the genetic variance (Zhao et al., 2013). The prior used in BayesA has a scaled-t distribution (Meuwissen et al., 2001; Pérez and de los Campos, 2014; Zhao et al., 2013), while BayesB and BayesCπ use a mixture of two priors with a point of mass at zero and a slab that can either be a Gaussian in BayesCπ (Habier et al., 2011) or a scaled-t in BayesB (Meuwissen et al., 2001). Previous studies indicated that Bayesian models that differ in their prior assumptions produce different inferences about individual marker effects and GEBVs; however, they often have similar predictive performance in cross-validation studies (Gianola, 2013; Heslot et al., 2012).

The least absolute shrinkage and selection operator (Lasso) was proposed by Tibshirani (1996) for estimation of linear models. This method combines the good features of ridge regression and variable selection by shrinking some coefficients and setting others to zero (Tibshirani, 1996). Tibshirani (1996) reported that the prediction accuracy of Lasso is superior to variable selection and ridge regression when there are small to moderate number of medium-sized effects, but ridge

regression outperforms Lasso when there are large number of small effects. The BL was first proposed by Park and Casella, (2008) using conditional Laplace (double-exponential) priors. This method was later extended by de los Campos et al. (2009b) to accommodate pedigree information and covariates other than markers.

Bayesian ridge regression is the Bayesian counterpart of RR-BLUP which was proposed by Pérez et al. (2010). Bayesian ridge regression and RR-BLUP have similar assumptions where marker effects are shrunk to a similar extent, but the level of shrinkage in BRR is estimated in a Bayesian hierarchical model (Heslot et al., 2012; Pérez and de los Campos, 2014). The main difference between BRR and BL is in the shrinkage priors applied in these methods. Bayesian ridge regression uses a Gaussian prior that shrinks the effects of all markers more heavily than the double-exponential prior density used in BL (de los Campos et al., 2013; Park and Casella, 2008; Pérez et al., 2010).

### 2.6.3 Kernel models

Kernel based statistical approaches have been proposed for GS to capture non-additive genetic effects. Reproducing kernel Hilbert spaces (RKHS) regression is a common kernel method used for GS. This method was first proposed by Gianola et al. (2006) and later elaborated by Gianola and van Kaam (2008). The RKHS regression model is a semi-parametric approach which captures both the additive and non-additive effects among loci by creating a kernel matrix that includes interactions among markers (Gianola and van Kaam, 2008). It uses a kernel function to convert the marker data set into a set of distances between pairs of observations that results in a square matrix to be used in a linear model (Heslot et al., 2012). de los Campos et al. (2009a) later presented a Bayesian view of RKHS regression using Gaussian processes. In Bayesian RKHS regression, the genetic values are regarded as random variables coming from a Gaussian process with a covariance function proportional to the evaluations of a reproducing kernel (Crossa et al., 2010; de los Campos et al., 2009a; de los Campos et al., 2010). The kernel function implemented in Bayesian RKHS is the Gaussian kernel evaluated as the average squared-Euclidean distance between genotypes:

$$K_{(\mathrm{x}_i,\,\mathrm{x}_{i'})} = \exp\left\{-h \times \frac{\sum_{k=1}^{p}\left(x_{ik}-x_{i'k}\right)^2}{p}\right\} \qquad (2.1)$$

where $x_i$ and $x_{i'}$ are the pairs of vectors of genotypes, $p$ refers to the total number of markers and $h$ is a bandwidth parameter that controls how fast the co-variance function drops as the distance between pairs of vector genotypes increases (de los Campos et al., 2009a; Pérez and de los Campos, 2014). No specific interpretation can be attached to the bandwidth parameter because the co-variance function is not derived from mechanistic consideration (de los Campos et al., 2010). The value of *h* can be chosen either through cross-validation or with Bayesian methods (Pérez and de los Campos, 2014). The cross-validation approach involves fitting models by assigning different values of *h* and identifying the value that maximizes the likelihood function, while for the Bayesian approach, *h* is estimated from the data (Pérez and de los Campos, 2014). The Bayesian approach is computationally intensive because the reproducing kernel needs to be re-estimated every time *h* is updated (de los Campos et al., 2010; Pérez and de los Campos, 2014). Kernel methods also allow the use of multiple kernels by evaluating the Gaussian kernel over a range of $h$ values, which was termed kernel averaging (KA) (de los Campos et al., 2010). Kernel averaging offers a computationally convenient method for kernel selection to overcome the computational demand when selecting *h* either through cross-validation or the Bayesian approach. Kernel averaging involves defining a set of kernels based on sensible values of *h* and fitting a multi-kernel model with the number of random effects equal to the number of kernels used (Pérez and de los Campos, 2014). When the variance parameters associated with each kernel are known, KA is equivalent to a model with a single-kernel $K$, where $K$ is the weighted average of all kernels used in multi-kernel model (de los Campos et al., 2010; Pérez and de los Campos, 2014).

### 2.6.4 Multiple-trait prediction models

Originally, GS models were developed for single-trait analysis; however, models now exist that allow for GEBVs to be estimated across multiple traits. When developing elite breeding material, plant breeders need to select for several important traits simultaneously to ensure that standards are met for registration. Several multiple-trait prediction models have been reported in recent years, mainly from livestock breeding programs (Aguilar et al., 2011; Calus and Veerkamp, 2011; Guo et al., 2014a; Hayashi and Iwata, 2013; Hayes et al., 2017; Jia and Jannink, 2012; Jiang et al., 2015; Tsuruta et al., 2011). Improved prediction accuracies were reported when multiple-trait GS models were used instead of single-trait prediction models (Calus and Veerkamp, 2011; Guo et al., 2014a; Hayes et al., 2017; Jiang et al., 2015; Tsuruta et al., 2011). The advantage of

multiple-trait models is that they use genetic correlation between traits to improve prediction accuracy (Guo et al., 2014a; Hayashi and Iwata, 2013; Jia and Jannink, 2012). However, the increase in accuracy comes at a cost because multiple-trait models greatly increase the computational burden depending on the TP size and the number of traits incorporated into the model (Calus and Veerkamp, 2011; Hayashi and Iwata, 2013; Tsuruta et al., 2011).

Several studies have shown that the prediction accuracy for traits with low heritability ($h^2 < 0.2$) can be increased when a correlated trait with higher heritability ($h^2 \geq 0.5$) is included in multiple-trait prediction models (Guo et al., 2014a; Hayashi and Iwata, 2013; Jia and Jannink, 2012; Jiang et al., 2015). For uncorrelated traits, multiple-trait prediction was either less accurate or comparable with single-trait prediction (Hayashi and Iwata, 2013; Jia and Jannink, 2012). This could be due to a non-zero estimate of genetic correlation between traits in the TP and using that erroneous information to predict traits in the validation population (Jia and Jannink, 2012). Studies also showed that multiple-trait models had higher prediction accuracy than single-trait models when phenotypic records are missing for some of the individuals and traits (Calus and Veerkamp, 2011; Guo et al., 2014a; Jia and Jannink, 2012). Calus and Veerkamp (2011) indicated that for individuals with missing phenotypes, an increase in accuracy was observed when using multiple-trait models compared to single-trait analysis. Missing phenotypes of individuals in the TP can be inferred from a correlated trait in a multiple-trait models (Calus and Veerkamp, 2011; Jia and Jannink, 2012). However, Hayashi and Iwata (2013) reported that the technique of imputing missing phenotypes in a multiple-trait prediction model was inconclusive. In contrast, multiple-trait models were not better than single-trait models for traits with a higher heritability and complete phenotypic data (Guo et al., 2014a; Hayashi and Iwata, 2013; Jia and Jannink, 2012). The genetic architecture of a trait is another factor that affects the relative advantage of multiple-trait models over single-trait models (Jia and Jannink, 2012). Jia and Jannink (2012) reported that under a major QTL genetic architecture, multiple-trait analysis performed well but for polygenic traits, multiple-trait analysis provided a slight improvement; this may indicate that multiple-trait prediction models capture the genetic correlation between traits more efficiently when major QTLs are present.

## 2.7 Factors Affecting GS Model Prediction Accuracy

Simulation and empirical studies, both from plant and animal breeding programs, indicated that the accuracy of GS model prediction depends on a number of factors related to the marker type and density (Combs and Bernardo, 2013; Heffner et al., 2011a; Moser et al., 2010), TP size (Calus and Veerkamp, 2007; Combs and Bernardo, 2013; Heffner et al., 2011a; VanRaden et al., 2009), trait heritability (Combs and Bernardo, 2013; Moser et al., 2010), genetic relationships between individuals in the training and validation population (Solberg et al., 2008; VanRaden et al., 2009; Wientjes et al., 2013), population structure (de los Campos et al., 2015; de Roos et al., 2009; Guo et al., 2014b), G × E (Burgueño et al., 2012; Crossa et al., 2015; Jarquín et al., 2014a; Jarquín et al., 2017; Lopez-Cruz et al., 2015; Pérez-Rodríguez et al., 2015; Sukumaran et al., 2017), and the statistical method used for prediction (Calus, 2010). Each of these factors are discussed below.

### 2.7.1 Training population size and composition

The TP size and composition are key elements in determining the prediction accuracy of GS (Bassi et al., 2016). Several studies reported that increasing the TP size increases the accuracy of GS by providing more data to estimate marker effects (Asoro et al., 2011; Bentley et al., 2014; Meuwissen et al., 2001; Muir, 2007; Saatchi et al., 2010; VanRaden et al., 2009). Meuwissen et al. (2001) showed that TP sizes of 500, 1000 and 2200 gave prediction accuracy of 0.58, 0.66, and 0.73 for the BLUP estimation method and 0.71, 0.79, and 0.85 for BayesB, respectively. Similarly, Saatchi et al. (2010) showed that a TP of 500, 1000, and 2000 gave prediction accuracy of 0.57, 0.63, and 0.69, respectively, based on the BLUP estimation method. Bentley et al. (2014) tested the effect of TP size on the prediction of agronomic and end-use quality traits in European wheat based on bootstrap resampling of 50, 100 and 200 individuals as the TP. Improved prediction accuracy was reported for flowering time, thousand-kernel weight and protein content as the TP increased, but no marked improvement in accuracy was observed for grain yield (Bentley et al., 2014). This suggests that the benefit of increasing TP size to improve accuracy may depend on the characteristics of the trait. In winter wheat, Heffner et al. (2011a) reported that decreasing the TP size from 288 to 192 and 96 reduced the average GS prediction accuracy across all traits by 11 and 30%, respectively. Taken together, increasing the TP size has a positive effect on GS prediction accuracies.

The composition of the TP is another factor that affects the accuracy of GS. The TP needs to be diverse to reflect the whole range of phenotypes and genotypes for a reliable prediction across a wide variety of lines (Calus, 2010; Daetwyler et al., 2014). Based on simulated data, Saatchi et al. (2010) reported that reduced number of phenotypic records from recent generations in the TP resulted in higher accuracy than larger phenotypic records from distant generations. However, increasing phenotypic records by combining older and recent generations in the TP resulted in small increase in accuracy (Saatchi et al., 2010). Using empirical data, Asoro et al. (2011) grouped oat lines based on their first year of entry into uniform trials to evaluate the effect of including older (historical) lines in the TP on prediction accuracy. The results showed that the older lines resulted in lower accuracy for two out of five traits, but showed no difference for the remaining traits when compared to more recent populations in the TP. This showed that the effect of including older lines in the TP was variable for different traits and may have confounding effect from LD and genetic relationship with the SC. However, Asoro et al. (2011) also showed that increasing the TP size by adding older lines increased or maintained prediction accuracy indicating that older generations retained useful information. In recent years, algorithms based on the prediction error variance and the coefficient of determination were reported for sampling of an optimized TP from a larger population to maximize the reliability of prediction in the SC (Akdemir et al., 2015; Isidro et al., 2015; Rincet et al., 2012; Rincet et al., 2017). These algorithms sample an optimized TP based on the marker information and estimate a priori the reliability of predictions in the SC using the selected individuals (Rincet et al., 2012). Optimization of the TP is especially useful when genotypic data is available for many lines, but phenotyping resources are limited (Akdemir et al., 2015; Rincet et al., 2012), or when there is strong population structure in the dataset (Isidro et al., 2015; Rincet et al., 2017). Several studies tested these methods in maize, rice, wheat, and Arabidopsis datasets and reported more accurate predictions when an optimized TP is used compared to randomly sampled individuals in the TP (Akdemir et al., 2015; Isidro et al., 2015; Rincet et al., 2012; Rincet et al., 2017). These algorithms can be helpful to reduce the bias due to TP composition and its relatedness to the SC when making genomic predictions in wheat.

### 2.7.2 Marker type and density

Different types of genetic markers can be used in GS but the most common ones are SNPs, Diversity Array Technology (DArT) markers, microsatellites, and GBS markers. Studies have

shown that the type of marker used influences GS prediction accuracy (Poland et al., 2012; Solberg et al., 2008). In a simulation study, Solberg et al. (2008) showed that SNP markers required two to three times higher density compared with microsatellites to achieve a similar accuracy using the same genetic architecture. Using comparable numbers of GBS and DArT markers, Poland et al. (2012) also showed that GBS markers led to higher genomic prediction accuracies compared to DArT markers in advanced breeding lines of wheat. Gains in accuracy up to 0.15 were obtained for yield and heading date with the GBS platform, but for thousand-kernel weight no difference was observed in accuracy between the GBS and DArT markers (Poland et al., 2012). The higher accuracy with the GBS platform could be because the GBS markers are free of the genotypic ascertainment bias that is found with fixed array genotyping or due to increased genome coverage from a more uniform distribution of GBS markers than DArT markers (Poland et al., 2012). Overall, marker type is an important factor that affects genomic prediction accuracies, but its effect may vary with traits and the density of markers.

Several studies reported that lower marker densities reduce the accuracy of predicting GEBVs (Heffner et al., 2011a; Moser et al., 2010; Solberg et al., 2008). However, it was suggested that a low-density assay of evenly spaced SNPs can deliver similar gains as a high-density SNP assay (Moser et al., 2010). Heffner et al. (2011a) reported that reducing marker density from 1158 to 768 and 384 resulted in a small decrease in GS accuracy in winter wheat; however, a further reduction to 192 reduced the average GS accuracy by 10%. The benefit of increasing marker density is to maximize the number of QTL in LD with at least one marker, which also maximizes the number of QTL whose effects will be captured by markers (Heffner et al., 2009). Based on a simulation study, Solberg et al. (2008) reported that the accuracy of prediction for both microsatellite and SNP markers increased about 1.04 to 1.07-fold when the marker density was doubled. A recent study on biparental and mixed populations of maize, barley and wheat also indicated that the accuracy of genomic predictions increased as the number of markers increased; however, gains in the accuracy began to plateau at 40 to 80% of the total marker density (Combs and Bernardo, 2013). Gains in accuracy began to plateau despite the small number of markers used in the study: 1213 markers (maize biparental population), 223 markers (barley biparental population), 1178 markers (barley mixed population), and 731 markers (wheat mixed population) (Combs and Bernardo, 2013). Similarly, Moser et al. (2010) showed that prediction accuracy for several traits in cattle reached a plateau when SNP density exceeded 1,000. Other studies in cattle

also reported that prediction accuracy for several traits was not significantly different when as many as 75% of the original markers were masked (Luan et al., 2009; VanRaden et al., 2009). Bassi et al. (2016) indicated that uniform distribution of markers across the genome and their ability to tag the QTL underlying traits are important considerations for GS than marker number. Increasing the number of markers without increasing the TP size may also reduce accuracy because it increases collinearity among markers (Muir, 2007). Increasing the TP size has a more important effect on accuracy than marker number (Lorenz et al., 2011; VanRaden et al., 2009).

Genomic predictions can be made using single markers, haplotypes of markers, or using an identical by descent (IBD) approach (Calus et al., 2008; Goddard and Hayes, 2007; Meuwissen et al., 2001). The computational burden increases as one moves from using single marker genotypes to haplotypes or further to IBD probabilities (Goddard and Hayes, 2007). Based on a simulation study, Solberg et al. (2008) compared the effects of using singe marker or haplotypes of two neighbouring markers on prediction accuracy using microsatellites and SNP markers. The results showed that for both marker types, using haplotypes resulted in similar or reduced accuracies compared with using direct marker effects. Similarly, Calus et al. (2008) compared different ways of including marker information in GS using single marker, haplotypes of two adjacent markers, and haplotypes of two or ten surrounding markers with IBD probabilities between different haplotypes at the same locus. The results indicated that the model with individual marker genotypes yielded the lowest accuracy at low marker density and the highest accuracy at high marker density. At low marker density, the advantage of using haplotypes instead of single markers is that a QTL that is not in complete LD with any individual marker may be in complete LD with a multi-marker haplotype (Goddard and Hayes, 2007). However, there does not seem to be a benefit to use marker haplotypes if marker density is sufficiently high and some SNPs are closely linked to important QTL (Calus et al., 2008).

### 2.7.3 Genetic relationships and marker-QTL LD

Genomic selection models can encompass the information from marker-QTL LD as well as genetic relationships among individuals to estimate GEBVs (Habier et al., 2007). Therefore, the accuracy of GS depends on the extent of marker-QTL LD (Calus, 2010; Toosi et al., 2010). In GS, GEBVs are estimated from marker genotypes instead of the actual QTL (Goddard and Hayes, 2007). This requires that markers are in LD with the actual QTL controlling traits. The assumption

of GS is that by using dense marker coverage, each QTL is in LD with at least one nearby marker and potentially all the genetic variance can be explained by markers (Calus, 2010; Goddard and Hayes, 2007; Meuwissen et al., 2001). The breakup of marker-QTL LD in subsequent generations requires re-estimation of marker effects to maintain the accuracy of GS (Calus, 2010).

The degree of genetic relationship between the TP and SC is another important factor that affects the accuracy of GS prediction (Clark et al., 2012; Habier et al., 2010; Habier et al., 2007; Hayes et al., 2009; Riedelsheimer et al., 2013). This genetic relationship is influenced by and results from generations of descent or population stratification (Asoro et al., 2011). When the TP is closely related to the SC, prediction of GEBVs is more reliable (Calus, 2010; Clark et al., 2012). The reliability of genomic predictions across populations is determined by the extent of marker-QTL LD phase between the TP and the SC and is related to the divergence between the two populations (de Roos et al., 2009). The SC are assumed to come from the same population as the TP, so the marker-QTL LD persists from the TP to the SC (de Roos et al., 2009). Several studies reported very low accuracy when GEBVs were predicted using unrelated populations (Charmet et al., 2014; Crossa et al., 2014; Riedelsheimer et al., 2013; Windhausen et al., 2012). In the absence of close relationships, prediction accuracy is driven by distant relationships which will be useful when there is strong LD in the population (Clark et al., 2012). Prediction of breeding values of unrelated individuals requires a substantially higher marker density and number of training records than when making prediction for offspring of training individuals (Meuwissen, 2009). Based on a simulation study, Meuwissen (2009) reported that breeding values of unrelated individuals can be predicted with accuracies of 0.88 to 0.93 using $2\times N_e\times L$ number of records and $10\times N_e\times L$ markers, where $N_e$ refers to effective population size and L the genome size in Morgan.

### 2.7.4 Heritability and genetic architecture of traits

When the same TP is used to predict multiple traits, the prediction accuracy will be lower for some traits than for the other traits, in the same way that $h^2$ is lower for some traits than for others (Combs and Bernardo, 2013). There is strong relationship between the accuracy of genomic prediction and the $h^2$ of a trait, with the prediction being more accurate for traits with higher $h^2$ (Combs and Bernardo, 2013; Heffner et al., 2011a; Moser et al., 2010; Saatchi et al., 2011). However, this is not always the case, because higher prediction accuracies were reported for traits with low $h^2$ (Combs and Bernardo, 2013). Combs and Bernardo (2013) indicated that the product

of $h^2$ and TP size is the key factor that determines prediction accuracy. Hayes et al. (2009b) estimated that nearly 9,000 individuals are required to get a prediction accuracy of 0.7 for a trait with $h^2 = 0.2$ but about 1,000 individuals are required to get similar prediction accuracy when $h^2 = 0.8$. This suggests that a decrease in the accuracy of prediction due to low $h^2$ can be compensated by using a larger number of observations to estimate marker effects (Combs and Bernardo, 2013; de Roos et al., 2009; Heffner et al., 2011a; Saatchi et al., 2010).

The genetic architecture of traits is another factor that needs to be considered when choosing statistical method for GS. Genetic architecture refers to the number and position of loci affecting a trait, the magnitude of their effects, and the relative contributions of additive and non-additive gene effects (Holland, 2007). Lorenz et al. (2011) reported that when traits are controlled by many loci with small effects, models such as RR-BLUP or G-BLUP are expected to work well but for traits that are controlled by few large effects QTL, Bayesian variable selection methods such as BayesB should be preferred. This agrees with the assumptions made by these models and is supported by several studies which assessed various models using simulated data (Clark et al., 2011; Daetwyler et al., 2010). On the other hand, studies based on real data reported that RR-BLUP has similar performance to Bayesian variable selection models for traits with different genetic architectures (Heslot et al., 2012; Riedelsheimer et al., 2012). The difference in results between simulation and empirical studies might be because the effect sizes of the QTL used in simulation may be substantially larger than the effect sizes of real QTL (Goddard and Hayes, 2007). Lorenz et al. (2011) also reported that the number of QTL has a strong influence on marker density and TP size. This indicates that the different factors affecting GS accuracy are interrelated, which is why multiple GS models are usually tested to select a model that performs optimally given the architecture of a trait.

### 2.7.5 Genotype-by-environment interaction and population structure

Genotype-by-environment interaction is a key issue in plant breeding. Multi-environment trials play an important role for studying G × E and genotype stability, and for predicting the performance of untested individuals (Burgueño et al., 2012). Originally, GS models were developed for the prediction of a single trait evaluated in a single environment (Crossa et al., 2015; Lopez-Cruz et al., 2015). However, plant breeding trials are designed to capture the variability across environments, and selection is made for multiple traits to develop lines that perform well

across certain agro-ecological regions. The most important G × E interactions are those in which genotype rank changes across environments, also known as crossover interaction (Crossa et al., 2004). The presence of crossover interaction might affect GS in a similar way that it affects phenotypic selection. Traits that have weak G × E were reported to have higher GS prediction accuracy than traits with strong G × E (Heffner et al., 2011a). In recent years, several studies reported improved prediction accuracy by modelling G × E in GS (Crossa et al., 2015; Cuevas et al., 2017; Heslot et al., 2014; Jarquín et al., 2014a; Jarquín et al., 2017; Lopez-Cruz et al., 2015; Pérez-Rodríguez et al., 2015; Sukumaran et al., 2017; Technow et al., 2015).

In GS, G × E can be modelled using different techniques. Integration of crop growth models and environmental covariates (ECs) into GS predictions was suggested to model G × E and predict performance in different environments (Heslot et al., 2014; Technow et al., 2015). Crop growth models are equations developed using growth data for a few genotypes under a range of growing conditions, while also integrating environment data such as weather and soil characteristics (Heslot et al., 2014). Crop growth models can explain the effect of G × E on phenotypes by explaining the impact of functional relationships between plant physiology and the environment (Technow et al., 2015). Genotype-by-environment interaction can also be incorporated in GS by modelling marker-by-environment interactions (M × E) when genomic and environmental covariate data are available (Lopez-Cruz et al., 2015). Lopez-Cruz et al. (2015) indicated that the prediction accuracy of the M × E GS model was substantially higher than that of an across-environment model that ignores G × E. Crossa et al. (2015) extended the M × E GS model into a Bayesian approach and reported that the M × E GS model performed better than the across-environment model. In another study, Jarquín et al. (2014a) developed reaction norm models that integrate G × E in GS by modelling interactions between markers and environments or ECs using covariance functions. Jarquín et al. (2014a) reported a 17-34% increase in prediction accuracy when G × E terms were included in the model. Pérez-Rodríguez et al. (2015) assessed the reaction norm models for prediction of cotton yield using pedigrees instead of molecular markers and reported a 2.7-fold increase in prediction accuracy when incorporating G × E terms. Recently, Sukumaran et al. (2017) also used the reaction norm models with pedigree-based relationship matrices to predict grain yield of spring bread wheat lines and obtained the highest accuracy when modelling G × E. Similarly, Jarquín et al. (2017) applied the reaction norm models to predict grain yield in winter wheat and reported 16 to 82% higher accuracy when G × E terms were included in the model. This indicates

that modelling G × E can enhance the accuracy of GS for similar training and prediction target environments.

Population structure is another important factor that affects the accuracy of genomic predictions (Asoro et al., 2011; Guo et al., 2014b; Isidro et al., 2015). Population structure is known to cause spurious associations between a phenotype and unlinked candidate loci in association-mapping studies (Pritchard and Rosenberg, 1999). In GS, however, rare spurious associations will not affect the accuracy of prediction but consistency of LD across subpopulations is important (Lorenz et al., 2011). In GWAS, the focus is to avoid false positive associations due to population structure but in GS the focus shifts to maintaining prediction accuracy despite a structured TP (Lorenz et al., 2011). Population structure and differences in allele frequencies and marker-QTL LD between subpopulations are likely to induce heterogeneity of marker effects even under the assumption that QTL effects are homogeneous (de los Campos et al., 2015). Guo et al. (2014b) reported that the effect of population structure on genomic prediction accuracy varies based on prediction strategies, genetic architectures of traits and populations. In the presence of strong population structure, accuracy is generally low when predicting the performance of one subpopulation based on marker effects estimated in the other subpopulations (Akdemir et al., 2015; Windhausen et al., 2012). de los Campos et al. (2015) reported improved prediction accuracy by modelling marker effects in subpopulations as the sum of an effect that is group specific and other that is common to all groups. On the other hand, Guo et al. (2014b) reported that when population structure existed in both training and validation sets, correcting for population structure led to a significant decrease in genomic prediction accuracy, but when prediction was limited to a specific subpopulation, population structure showed little effect on accuracy. Similarly, Crossa et al. (2016) reported that accounting for population structure decreased prediction accuracy by 15-20% as compared to prediction accuracy obtained when not accounting for population structure. This suggests that the benefit of accounting for population structure in GS may vary based on the prediction approach and other population characteristics, and thus incorporation of structure into GS models should be assessed on a case by case basis.

### 2.8 Reported Genomic Prediction Accuracies in Wheat

Several studies highlighted the potential application of GS in wheat. The application of GS for plant breeding was first demonstrated using simulation studies in maize by Bernardo and Yu

(2007). de los Campos et al. (2009b) later showed the application of GS for prediction of wheat yield using empirical data. Since then, several studies evaluated GS models for the prediction of agronomic, end-use quality and disease resistance traits in wheat and varying levels of accuracies have been reported depending on the population, trait characteristics, marker type and density, G × E, the statistical method and cross-validation techniques used for prediction. It is difficult to make direct comparisons of the reported accuracies from different studies, even for similar traits and statistical models, because model parameters and cross-validation designs are often different (Daetwyler et al., 2013). In wheat, GS has been reported for the prediction of several traits, including grain yield, spike grain number, heading date, plant height, lodging, preharvest sprouting, test weight, thousand-kernel weight, flour yield and softness, protein content, gluten strength, water absorption, damaged starch, arabinoxylan and partially hydrated gliadin content, time to young microspore, and resistance to fungal diseases (wheat rust, fusarium head blight, and septoria tritici blotch) (Charmet et al., 2014; Charmet and Storlie, 2012; Crossa et al., 2010; Daetwyler et al., 2014; Dawson et al., 2013; Heffner et al., 2011a; Heffner et al., 2011b; Miedaner et al., 2013; Ornella et al., 2012; Poland et al., 2012; Rutkoski et al., 2011; Rutkoski et al., 2012; Rutkoski et al., 2014; Rutkoski et al., 2015; Thavamanikumar et al., 2015; Zhao et al., 2013; Zhao et al., 2014). Heffner et al. (2010) reported that the genetic gain for GS would be twofold higher than that of MAS if prediction accuracies of 0.5 could be achieved in wheat. Various GS schemes in wheat were shown to have similar cost to phenotypic selection, but the potential of increasing the genetic gain is the main deriving force for GS in wheat breeding (Bassi et al., 2016).

Grain yield is the most important trait in crop breeding and is controlled by many minor effect QTL. The potential of GS for predicting wheat grain yield has been evaluated in several studies. Crossa et al. (2010) evaluated six different GS models in 599 historical wheat lines for prediction of grain yield and reported accuracies ranging from 0.36 to 0.61 based on a tenfold cross-validation. Gianola et al. (2011) applied machine learning algorithms for genomic prediction in the same data set used by Crossa et al. (2010) and reported accuracies in the range of 0.48 to 0.59. Similarly, Long et al. (2011) used the same wheat data set and reported prediction accuracies for grain yield in the range of 0.50 to 0.58 averaged over 50 training-testing replicates based on BL and support vector regression, a machine learning algorithm developed for classification and regression. In a different study, Charmet and Storlie (2012) evaluated ridge regression, G-BLUP, BRR and Lasso for the prediction of wheat grain yield and reported accuracies around 0.5 based

on the average of 100 cross-validations. Charmet et al. (2014) also evaluated G-BLUP, BRR, BL, RKHS and Random Forest regression, a machine learning method that could capture non-additive effects, in three bi-parental wheat populations and reported grain yield prediction accuracies in the range of 0.2 to 0.5 depending on the cross-validation technique and populations used. On the other hand, Heffner et al. (2011a) reported very low accuracy (0.17 to 0.22) for grain yield in a population of 374 soft winter wheat varieties and F5 derived advanced breeding lines using RR-BLUP, BayesA, BayesB and BayesCπ. Historical data from regular breeding trials have also been used in several GS studies. Dawson et al. (2013) used a highly unbalanced historical data set from a period of 17 years for prediction of wheat grain yield in G-BLUP and reported accuracies in the range of 0.43 to 0.56 depending on the cross-validation methods. Similarly, He et al. (2016) compared RR-BLUP, G-BLUP, extended G-BLUP and RKHS and reported grain yield prediction accuracy up to 0.65 averaged across 100 different combinations of training and test sets in 2325 historic elite winter wheat lines. In hybrid wheat, Zhao et al. (2013) compared five different models (RR-BLUP, BayesA, BayesB, BayesC and BayesCπ) and reported accuracies in the range of 0.58 to 0.63 for grain yield based on fivefold cross-validation, but when unrelated individuals were used for cross-validation the accuracy decreased on average by 44%. Taken together, the literature suggests that GS shows promise for predicting wheat grain yield with moderate accuracy.

Moderate to high GS prediction accuracies were reported for wheat agronomic traits other than yield. Heffner et al. (2011a) reported accuracies for heading date (0.72 to 0.76), height (0.72 to 0.75), and lodging (0.23 to 0.28) based on ridge regression and three Bayesian models in winter wheat. Pérez-Rodríguez et al. (2012) compared various linear and non-linear models for the prediction of heading date and reported accuracies ranging from 0.48 to 0.60 averaged over 12 environments but in each environment the accuracy ranged from 0.02 to 0.69. Charmet et al. (2014) also reported a prediction accuracy of approximately 0.7 for heading date in three bi-parental wheat populations. Based on two bi-parental doubled haploid (DH) wheat populations, Thavamanikumar et al. (2015) reported prediction accuracies of 0.51 to 0.84 for time to young microspore, a flowering time related trait, and 0.10 to 0.45 for spike grain number based on tenfold cross-validation. In hybrid wheat, Zhao et al. (2014) reported accuracies in the range of 0.4 to 0.6 for heading date and plant height using RR-BLUP, BayesCπ and W-BLUP and a cross-validation technique where training and validation sets were not related via shared parental lines. Though

methods and accuracies were highly variable across studies, populations, and environments, GS may be suitable for making more accurate predictions for some agronomic traits other than yield.

In wheat, end-use quality traits are also prime targets for GS. Despite the development of several indirect tests to measure quality traits in wheat, these methods require enormous time and resources for accurate phenotyping. Moreover, many of the end-use quality traits in wheat are evaluated in advanced generations because their tests often require a large amount of grain which is not available in early generations (Battenfield et al., 2016). The application of GS enables wheat breeders to assess end-use quality traits early in the breeding cycle and advance only lines with promising quality standards which saves both time and resources. The superiority of GS over conventional MAS was shown by Heffner et al. (2011a) who reported 28% higher accuracy in GS than conventional MAS averaged across 13 agronomic and end-use quality traits in winter wheat. Heffner et al. (2011a) compared RR-BLUP, BayesA, BayesB and BayesCπ, and reported a wide range of prediction accuracies across different traits and models ranging from 0.17 (for grain yield in BayesCπ) to 0.76 (for flour yield in BayesA). However, only slight differences were reported among the evaluated models for each trait, with BayesA having the highest mean accuracy across all traits (Heffner et al., 2011a). Similarly, Heffner et al. (2011b) evaluated ridge regression and BayesCπ for the prediction of nine different grain quality traits in two DH winter wheat populations and reported accuracies ranging from 0.27 (flour softness in RR) to 0.74 (damaged starch in BayesCπ). Test weight is another important quality trait in wheat, which is a measure of grain density, and accuracies for test weight were in the range of 0.5 to 0.6 by Heffner et al. (2011a, b) while Charmet et al. (2014) reported accuracies ranging from 0.3 to 0.7. Battenfield et al. (2016) also evaluated five GS models for prediction of ten processing and end-use quality traits in advanced spring bread wheat lines. The reported accuracies ranged from 0.32 (grain hardness) to 0.62 (mixing time) based on forward prediction where data in one year were used to make predictions in the following year (Battenfield et al., 2016). Because of the increased selection intensity with GS, Battenfield et al. (2016) reported genetic gain that was 1.4 to 2.7 times higher across all traits than phenotypic selection. Recently, Hayes et al. (2017) evaluated single and multiple-trait models for prediction of 19 end-use quality traits in diverse bread wheat accessions and reported accuracies ranging from 0 to 0.69 using different cross-validation designs. Multiple-trait analysis improved the accuracy compared to single-trait analysis and the reported accuracies were greater than 0.5 for many of the evaluated traits (Hayes et al., 2017). This indicates that end-

use quality traits can also be amendable to GS and provide increased accuracy over conventional MAS.

Disease resistance is another important trait of interest for wheat breeders. The potential of GS for predicting quantitative disease resistance in wheat has been evaluated in several studies. Based on 1,055 elite wheat hybrids, Miedaner et al. (2013) reported prediction accuracy of 0.28 and 0.32 for septoria tritici blotch resistance using BayesCπ and RR-BLUP, respectively. This low accuracy could be due to the cross-validation technique used, where training and validation sets were not related to each other via common parental lines. The reported accuracies were corrected by dividing the correlation between GEBVs and observed phenotypes by the square root of $h^2$. Lorenz et al. (2011) indicated that adjusting the correlation by $h^2$ would result in upwardly biased estimates of accuracy when training and validation data were collected in the same environment. Daetwyler et al. (2014) assessed the accuracy of GS for the prediction of leaf rust, stem rust and stripe rust in diverse landraces of wheat collected from 32 countries. Based on fivefold cross-validation, the accuracy of genomic prediction averaged across years was 0.35, 0.27 and 0.44 for leaf rust, stem rust and stripe rust using G-BLUP and 0.33, 0.38 and 0.30 for leaf rust, stem rust and stripe rust using BayesR, respectively (Daetwyler et al., 2014). Rutkoski et al. (2012) reported prediction accuracies for mycotoxin (deoxynivalenol) levels produced by a fungus causing fusarium head blight, ranging from 0.2 to 0.65 across different models and marker types using a fivefold cross-validation. The highest mean accuracy was obtained from a random forest regression model that combined markers and correlated traits such as disease incidence, severity, and, kernel quality index as predictor variables. Similarly, Arruda et al. (2015) evaluated three GS models (RR-BLUP, Lasso and elastic net) to predict six traits related to fusarium head blight resistance in winter wheat. The reported accuracies ranged from 0.40 (Lasso for disease severity) to 0.82 (RR-BLUP for fusarium-damaged kernels) based on the maximum number of SNPs and a fivefold cross-validation. For all traits except disease incidence, the highest prediction accuracies were obtained in RR-BLUP (Arruda et al., 2015). Overall, these studies showed that there are differences in the accuracy of genomic predictions for different traits, populations and statistical methods. Therefore, it is important to evaluate different statistical models and model parameters for each breeding population and environment.

## 2.9 Thesis Objectives and Outline

The goal of this thesis was to investigate various GS approaches, statistical models, and model parameters to design selection strategies for complex traits in wheat under the short growing seasons of western Canada. The specific objectives were to i) evaluate the performance of various statistical approaches and models to predict agronomic and end-use quality traits using empirical data in spring bread wheat, ii) determine the effects of TP size, marker density, and population structure on genomic prediction accuracy, iii) examine GS prediction accuracy when modelling G × E using different approaches, iv) detect marker-trait associations for agronomic and end-use quality traits in spring bread wheat, v) evaluate the effects of TP composition, cross-validation technique and genetic relationship between the TP and SC on GS accuracy, and vi) compare genomic and phenotypic prediction accuracy.

Six studies were conducted to meet these objectives based on a TP of 231 spring bread wheat lines and SC composed of 304 RILs. The results from these studies are organized into six experimental chapters in this thesis. The first experimental chapter (Chapter three) addresses the first objective. We compared various single and multiple-trait GS models and a GS model that incorporates the results from de novo GWAS using a fivefold cross-validation design in the TP. The best statistical model identified from this study was then used in Chapter four to evaluate the effects of TP size, marker density, and population structure on GS prediction accuracy in the TP (second objective). In Chapter five, a subsample of 81 lines from the TP were used to evaluate the benefit of modelling G × E in GS using two different approaches (third objective). In Chapter six, genome-wide association mapping was conducted for nine agronomic and end-use quality traits in the TP (fourth objective). Similarly, in Chapter seven, QTL analyses were conducted for six agronomic traits in the SC. Markers significantly associated with the QTL underlying the evaluated traits were fitted as fixed effects to enhance the accuracy of genomic predictions in the TP (Chapter three) and SC (Chapter eight). In Chapter eight, we used the statistical models and model parameters evaluated in the previous chapters for genomic predictions in the SC using different cross-validation techniques and a TP with varying levels of relatedness (fifth objective). Moreover, across-year genomic and phenotypic prediction accuracies were compared in the SC (sixth objective). Finally, the results from all experimental chapters were discussed and conclusions and future directions are presented in Chapter nine.

# 3. Genomic Selection for Wheat Improvement: Comparison of Methods Based on Empirical Data

## 3.1 Introduction

Genomic selection is a novel MAS approach that is shown to improve the genetic gain of quantitative traits in plant and animal breeding programs. Genomic selection estimates all marker effects across the genome to calculate GEBVs for individuals having only marker data (Meuwissen et al., 2001). Traditional MAS involves estimating the effects of QTL that are significantly associated with a trait of interest. Genomic selection differs from traditional MAS strategies in that instead of using markers that have a predefined significant correlation with a trait, all markers are used to estimate breeding values of individuals (Heffner et al., 2009; Jannink et al., 2010). The main assumption of GS is that by using dense marker coverage, each QTL is in LD with at least one nearby marker and potentially all the genetic variance can be explained by markers (Calus, 2010; Goddard and Hayes, 2007; Meuwissen et al., 2001). Fitting all markers simultaneously avoids multiple testing and ensures that marker-effect estimates are unbiased (Jia and Jannink, 2012).

Genomic selection uses a training (reference) population that has been genotyped and phenotyped to develop a statistical model that takes genotypic data from a candidate population of untested individuals and calculates GEBVs (Heffner et al., 2011a; Jannink et al., 2010). These GEBVs say nothing about the function of the underlying genes but selection is subsequently based on these values (Jannink et al., 2010). The advantage of GS is its potential to predict GEBVs with an accuracy that is sufficient to allow selection over several generations without repeated phenotyping; this reduces the cost and generation intervals of breeding programs (Habier et al., 2007).

Originally, GS models were developed for prediction of a single trait evaluated in a single environment or averaged across environments. In recent years, several multiple-trait prediction models have been reported (Aguilar et al., 2011; Calus and Veerkamp, 2011; Guo et al., 2014a; Jia and Jannink, 2012; Jiang et al., 2015; Tsuruta et al., 2011). Multiple-trait GS models take advantage of genetic correlation between traits to improve the accuracy of prediction (Guo et al., 2014a; Hayashi and Iwata, 2013; Jia and Jannink, 2012). In the absence of genetic correlation between traits, multiple-trait models were inferior to single-trait models (Jia and Jannink, 2012). Studies mainly based on simulated data reported increased prediction accuracy when multiple-trait

prediction models were used instead of single-trait prediction models (Calus and Veerkamp, 2011; Guo et al., 2014a; Jiang et al., 2015; Tsuruta et al., 2011). Recently, Hayes et al. (2017) reported improved prediction accuracy for wheat end-use quality traits by including values of the same trait measured by either near infrared or nuclear magnetic resonance as a correlated trait in a multiple-trait model. Their aim was to develop a large TP using quality traits predicted by near infrared or nuclear magnetic resonance and to combine these with the available end-use quality data based on standard assays in a multiple-trait analysis. However, there is no empirical evidence on the performance of multiple-trait models for genomic prediction of two or more correlated traits in wheat.

Statistical methods used for genomic prediction treat markers as random effects and assign prior assumptions about the variance explained due to their effects (Clark and van der Werf, 2013). Fitting markers as random effects and shrinking their effects uniformly as performed in BLUP do not explicitly model the effects of major QTL versus unknown background QTL with minor effects (Bernardo, 2014). Bernardo, (2014) suggested that when a few major genes each accounting for more than 10% of the genetic variance are present for a quantitative trait, these major genes should be fitted as fixed effects instead of random effects in GS models. In recent years, methods that combine GWAS and GS have been proposed (Bernardo, 2014; Rutkoski et al., 2014; Spindel et al., 2016; Spindel et al., 2015; Zhang et al., 2014). These methods have advantages because GS and GWAS can be performed on the same population and highly significant SNPs identified from GWAS can be fitted as fixed effects in a GS model without shrinking their effects while the remaining genome-wide markers are treated as random effects (Begum et al., 2015; Rutkoski et al., 2014; Spindel et al., 2016; Zhang et al., 2014). Alternatively, markers tagging candidate genes or previously identified QTL can also be included in GS as fixed effects (Bentley et al., 2014; Spindel et al., 2016). Several studies have reported improved genomic prediction accuracy by integrating the results from GWAS (Begum et al., 2015; Spindel et al., 2016; Spindel et al., 2015; Zhang et al., 2014). Moreover, the genetic architecture of a trait identified from the GWAS can be used to inform GS models (Begum et al., 2015). However, the performance of this new approach has not been evaluated thoroughly for genomic prediction in wheat.

With increasing application of GS, numerous statistical methods have been proposed to estimate marker effects and compute GEBVs. Many of these models have been validated mainly in animal breeding programs using computer simulation studies and empirical data. However,

there is limited information on the use of GS for practical application in wheat breeding. The objective of this study was to evaluate the performance of various GS approaches and statistical models to predict agronomic and end-use quality traits using empirical data in spring bread wheat.

## 3.2 Materials and Methods

### 3.2.1 Plant material and phenotypic data

A training and validation population of 231 spring hexaploid wheat varieties and advanced breeding lines was used to assess the accuracy of various GS models (Appendix A). Data for these lines were obtained from two different experiments. The first experiment was a variety comparison (hereafter called varcomp) experiment composed of 100 commercial wheat varieties that were evaluated at Kernen Crop Research Farm, Saskatoon, SK, (lat 52°08', long 106°32') from 2011 to 2014 and at Swift Current, SK (lat 50°16', long 107°44') from 2012 to 2014. This population was composed of both contemporary and historic Canadian bread wheat varieties. The field experiments were laid out in 200 plots, each plot having an area of 4.25 $m^2$, with five seeded rows at Kernen and an area of 3.65 $m^2$, with four seeded rows at Swift Current. A seeding rate of 300 and 275 seeds per $m^2$ were used at Kernen and Swift Current, respectively. The second experiment was composed of 200 spring hexaploid wheat varieties and advanced breeding lines (hereafter called Co-op) that were selected from the Central (Manitoba and eastern Saskatchewan), Western (southern Saskatchewan and Alberta) and Parkland (northern Alberta, Saskatchewan and Manitoba) Bread Wheat Co-operative Tests, grown to provide data for registration and commercialization of new spring hexaploid wheat cultivars. Each of these lines were evaluated in 2014 at the Seed Farm of the Crop Development Centre (CDC) in Saskatoon, SK (lat 52°08' long 106°36') on a 0.74 $m^2$ plot area, with two seeded rows and again in 2015 at Kernen and Rosthern, SK (lat 52°41' long 106°19') on a 4.25 $m^2$ plot area, with five seeded rows. The seeding rate of plots was 300 seeds per $m^2$. The two experiments were connected through 27 common lines. Both experiments were arranged in an alpha-lattice design with two replications in each environment. The field experiments were seeded in early to mid-May and harvested in mid to late September in each year.

Traits including heading date, plant height, maturity, grain yield, test weight, thousand-kernel weight, grain protein content, falling number, and SDS sedimentation volume were measured. Heading date was recorded for each plot as the number of days from seeding to when

50% of the heads emerged out of the flag leaf sheath. Plant height was measured for each plot when the plants approached physiological maturity by taking the length of the main stem from the soil surface to the tip of the spike, excluding the awns. Maturity was recorded as the number of days from seeding to when 50% of the spikes in a plot turn to a straw color. Plots were harvested using a small plot combine at harvest maturity. Grain yield was measured by taking the mass of grain harvested from each plot after the grains were air dried to constant moisture. Test weight was measured as the weight of dockage-free grain in grams required to fill a level 0.5 L container. Grain yield and test weight were reported in kg $ha^{-1}$ and kg $hL^{-1}$, respectively. Thousand-kernel weight in grams was determined from a sub-sample of 200 kernels that were free from foreign material and broken kernels. For quality analysis, grain samples were ground using a UDY Cyclone mill (UDY Corp., Fort Collins, CO, USA). Grain protein content (%) was determined by combustion N analysis (AACCI approved method 46-30.01) using a LECO model FP-528 (Leco Instruments Corp., St. Joseph, MI, USA). Falling number was determined according to AACCI approved method 56-81.03. SDS sedimentation volume was measured following the method of Axford et al. (1978).

### 3.2.2 Genotypic data

Genomic DNA was extracted from fresh leaves of one-week-old seedlings using a modified CTAB approach (CIMMYT, 2005). DNA quality was assessed on agarose gels in a standard gel electrophoresis. All lines were genotyped using the wheat 90K SNP array (Wang et al., 2014a). Genotype calling was performed using the GenomeStudio Polyploid Clustering Module v1.0 (Illumina, San Diego, CA), and erroneous lines and markers were filtered from the analysis (Appendix B). A total of 28,081 polymorphic SNPs were generated. After filtering SNPs with call frequency (CF) less than 90% (0.52%) and minor allele frequency (MAF) smaller than 10% (35.78%), 17,887 polymorphic SNPs remained for this analysis. Missing marker genotypes were replaced with the population mean for that marker using the function 'A.mat' in R package rrBLUP, v4.4 (Endelman, 2011).

### 3.2.3 Phenotypic data analysis

The varcomp and Co-op data sets were analyzed separately as well as combined and analyzed simultaneously. The phenotypic data were analyzed using ANOVA with SAS Mixed

models, v9.4 (SAS Institute Inc., 2015). Genotypes (i.e. lines) were considered as fixed effects, while replication nested in environment, block nested in replications and environment, environment (site-years), and G × E were considered as random effects. The Kenward-Roger degrees of freedom approximation method (DDFM=Kr) was used to compute the degrees of freedom for means. The phenotypic data analyses included all lines in the varcomp and Co-op datasets for a better estimate of (co)variances, but lines with genotyping errors and lines from the CWAD and CWSWS wheat classes were excluded and LS-means of the remaining 231 lines were used for this study. The phenotypic data were inspected and met the assumptions of ANOVA. Broad-sense heritability ($H^2$) of traits were calculated across environments using the equation $\sigma_g^2/(\sigma_g^2 + \sigma_\varepsilon^2)$, where $\sigma_g^2$ and $\sigma_\varepsilon^2$ are the estimated genetic and residual variance components, respectively. The variance components were calculated using the 'lmer' function in the R package lme4, v1.1-7 (Bates et al., 2016). Pearson correlation coefficients were estimated in SAS among the evaluated traits.

### 3.2.4 Single-trait GS model prediction

Genomic predictions were made using trait phenotypes and 17,887 polymorphic SNP markers. Nine single-trait prediction models, RR-BLUP, G-BLUP, BayesA, BayesB, BayesCπ, BL, BRR, RKHS, and RKHS-KA were evaluated. These models were fitted in R (R Development Core Team, 2016), using the Bayesian generalized linear regression (BGLR) package, v1.0.4, (Pérez and de los Campos, 2014), and the ridge regression and other kernels for genomic selection (rrBLUP) package, v4.4 (Endelman, 2011). The default settings of BGLR (5 degrees of freedom and the scale parameter based on sample variance of the phenotypes) were used (Pérez and de los Campos, 2014). The rrBLUP package was developed mainly for genomic prediction with mixed models (Endelman, 2011). The main function of the package is 'mixed.solve', which calculates maximum-likelihood solutions for mixed models with a single variance component other than the error (Endelman, 2011). Assumptions of the various single-trait prediction models evaluated in this study are described in Section 2.6. For the G-BLUP model, the GRM was computed as $XX'$ (VanRaden, 2008), where X is a matrix containing scaled and centered marker genotypes. Marker genotypes were centered by subtracting the sample mean of each marker from the original genotypes and standardized by dividing the resulting centered genotypes by the sample standard deviation of the marker (de los Campos et al., 2015). For RKHS, we followed recommendations

by Pérez and de los Campos (2014) and used the Gaussian kernel (Equation 2.1) with an arbitrarily chosen bandwidth parameter of 0.5 for the single-kernel RKHS and $h = 1/M \times \{1/5, 1, 5\}$ for the RKHS-KA, where M is the median squared Euclidean distance between all lines calculated using off-diagonals only.

### 3.2.5 Multiple-trait GS model prediction

Joint predictions of multiple traits were made using antedependence-based multiple-trait BayesA (Jiang et al., 2015), and multiple-trait BayesA (Jia and Jannink, 2012) models. Multiple-trait BayesA model takes advantage of the genetic correlation between traits to improve prediction accuracy (Jia and Jannink, 2012). The antedependence-based multiple-trait BayesA considers correlations between multiple traits as well as between SNP effects simultaneously to achieve higher prediction accuracy (Jiang et al., 2015). The antedependence model considers the potential nonstationary correlations between SNP effects near to the QTL and was developed to overcome the limitation of the standard GS approaches where marker effects are assumed to be independently distributed (Yang and Tempelman, 2012). The antedependence-based multiple-trait BayesA has two different approaches by setting the antedependence parameter as either a matrix or a scalar (Jiang et al., 2015). Detailed description of the matrix and scalar models has been provided in Jiang et al. (2015). In this study, the prediction performance of multiple-trait BayesA, and multiple-trait BayesA scalar and matrix models were compared with the standard single-trait BayesA model (Meuwissen et al., 2001) using different trait combinations. All multiple trait predictions were made using C language programs, available at, https://sites.google.com/site/jicaijiang/mtgenpred (Jiang et al., 2015).

### 3.2.6 Genomic prediction with significant markers from GWAS fitted as fixed effects

This study was based on a GS + de novo GWAS model that combines RR-BLUP with significant markers identified from GWAS fitted as fixed effects on the RR-BLUP training data (Spindel et al., 2016). The GS + de novo GWAS model is equivalent to the standard RR-BLUP when no marker is included as fixed effect. Spindel et al. (2016) provided a detailed description of this method. In this study, the prediction performance of the GS + de novo GWAS model was evaluated using a fivefold cross-validation technique. In each fold, marker-trait associations were determined in TASSEL, v3.0 (Bradbury et al., 2007) based on the phenotypic and genotypic data

of the training set using a mixed linear model (MLM) that combines both population structure information and kinship as covariates. Genome-wide association study was conducted based on a subset of 1908 evenly spaced SNPs that were selected from the full marker density based on genetic distance (mean genetic distance of 1.8 cM between SNPs) using MapThin, v1.11 (Howey and Cordell, 2011). Population structure was accounted using five marker-derived principal components (PCs). Principal components and kinship were computed from the marker data using TASSEL. The MLM analysis was conducted using the default settings of TASSEL (optimum compression level and PD3 variance component estimation). The *P*-value from the GWAS output were sorted from low to high and multiple testing correction was performed for all SNPs based on a False Discovery Rate (FDR) using the function "p.adjust" and "BH" method (Benjamini and Hochberg, 1995), in R (R Development Core Team, 2016). Up to three most significant markers (FDR = 0.2) were selected separately for each fold in the fivefold cross-validation method. In cases where no marker met this threshold, the most significant marker was selected. The selected markers were then included in the GS + de novo GWAS model as fixed effects while all the remaining markers from the full marker density (17,887 SNPs) were included as random effects. The GS + de novo GWAS model was fitted in R (R Development Core Team, 2016) using the function 'kinship.BLUP' in the rrBLUP package (Endelman, 2011).

### 3.2.7 Cross-validation techniques

The cross-validation technique is used in GS to assess the performance of a model and its ability to predict the breeding value of individuals that are different from the ones used to train the model. In this study, two cross-validation techniques were used to assess the prediction accuracy of the evaluated models. The first technique involved a fivefold cross-validation that randomly partitioned the data into five mutually exclusive groups of approximately equal sizes. In each fold, four groups were combined to form the training set and the one remaining group was used as the validation set. The process was repeated until each of the five groups was used as a validation set. In each fold, the prediction accuracy of the models was assessed by computing Pearson's correlation ($r$) and Spearman's rank correlation ($\rho$) between the predicted GEBVs and observed phenotypes of the individuals in the validation set. This technique was used to assess single-trait, multiple-trait, and GS + de novo GWAS models. For each model-trait combination, the reported accuracies were the averages across the five folds. Analysis of variance and student's *t*-test were

conducted on the cross-validation results for each trait separately to determine differences in prediction accuracy among the statistical models. For a reliable comparison of these models, the same cross-validation folds were used in each model-trait combination. This technique increases the computation load but the difference in estimates across groups will help to estimate uncertainty due to random sampling of the training and validation sets (Morota and Gianola, 2014; Pérez and de los Campos, 2014).

The second technique involved cross-validation using an independent population. In this case, marker effects were estimated using one population and genomic predictions were made for a population different from the one used to estimate marker effects. For this approach, the varcomp population was used as training set while the Co-op population was used as validation set and vice versa. GEBVs were calculated for lines in the validation population by multiplying the vector of the marker scores for each line by the vector of marker effects estimated from the TP. Prediction accuracy of the models was assessed by computing Pearson's correlations ($r$) between the predicted GEBVs and observed phenotypes of the individuals in the validation set. In this case, there is only one prediction accuracy measurement. Across-population genomic prediction accuracy was used to assess performance of single-trait prediction models. Inferences for all Bayesian models were based on 50,000 iterations obtained after discarding 5,000 samples as a burn-in.

### 3.3 Results and Discussion

#### 3.3.1 Phenotypic evaluation

Analysis of variance indicated that there were significant differences among lines ($P < .001$) for all traits (Table 3-1). There was also significant G × E for all traits, indicating that the environment also had strong influence on measured phenotypes (Table 3-1). Frequency distributions of the various traits were broad and normally distributed, except for sedimentation volume which is skewed to the left because of the low sedimentation volume (40 ml) of a historical wheat cultivar (Red Fife) (Fig. 3-1).

For each trait, broad sense heritability was estimated using the combined varcomp and Co-op data sets. The highest heritability was observed for thousand-kernel weight (0.84) and sedimentation volume (0.84) followed by heading date (0.82) and grain protein content (0.64) (Table 3-2). Moderate to high heritability estimates were obtained for grain yield (0.51), test

weight (0.57), falling number (0.57), maturity (0.58), and plant height (0.62) (Table 3-2). These results suggest that most of the observed phenotypic variation in these traits is heritable and has a genetic cause. Cross et al. (2014) reported that when trait heritability is low, genomic information does not improve prediction accuracy. Because the trait data from this study has high heritability, accurate genomic predictions can be made using these data sets.

The correlation among the agronomic and end-use quality traits agrees with the commonly known trait relationships in wheat. Heading date was positively correlated with days to maturity ($r$ = 0.612) (Table 3-3). Plant height was negatively correlated with grain yield ($r$ = -0.295) and positively correlated with grain protein content ($r$ = 0.364), indicating that shorter varieties were high yielding but with lower protein content (Table 3-3). Also, grain yield was negatively correlated with grain protein content ($r$ = -0.329). There was also significant positive correlation of grain protein content with falling number ($r$ = 0.319) and sedimentation volume ($r$ = 0.426).

**Table 3-1.** *P*-values from mixed model ANOVA *F*-tests for nine agronomic and end-use quality traits as affected by genotype and covariance parameters.

| **Source of variation** | **HD†** | **HT** | **MAT** | **YLD** | **TWT** | **TKW** | **PRO** | **FN** | **SDS** |
|---|---|---|---|---|---|---|---|---|---|
| | days | cm | days | kg ha$^{-1}$ | kg hL$^{-1}$ | g | % | sec | ml |
| Genotype (G) | *** | *** | *** | *** | *** | *** | *** | *** | *** |
| **Covariance parameter** | | | | | | | | | |
| Environment (E) | * | * | * | * | * | * | * | * | * |
| Replication (Env) | NS | * | NS | * | * | * | * | * | * |
| Block (Rep × Env) | *** | *** | *** | *** | *** | *** | *** | ** | *** |
| G × E | *** | *** | *** | *** | *** | *** | *** | *** | *** |

*, **, ***, Significant at the 0.05, 0.01 and 0.001 probability level, respectively, and *NS* not significant.

†HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight, PRO: grain protein, FN: falling number, SDS: sedimentation volume.

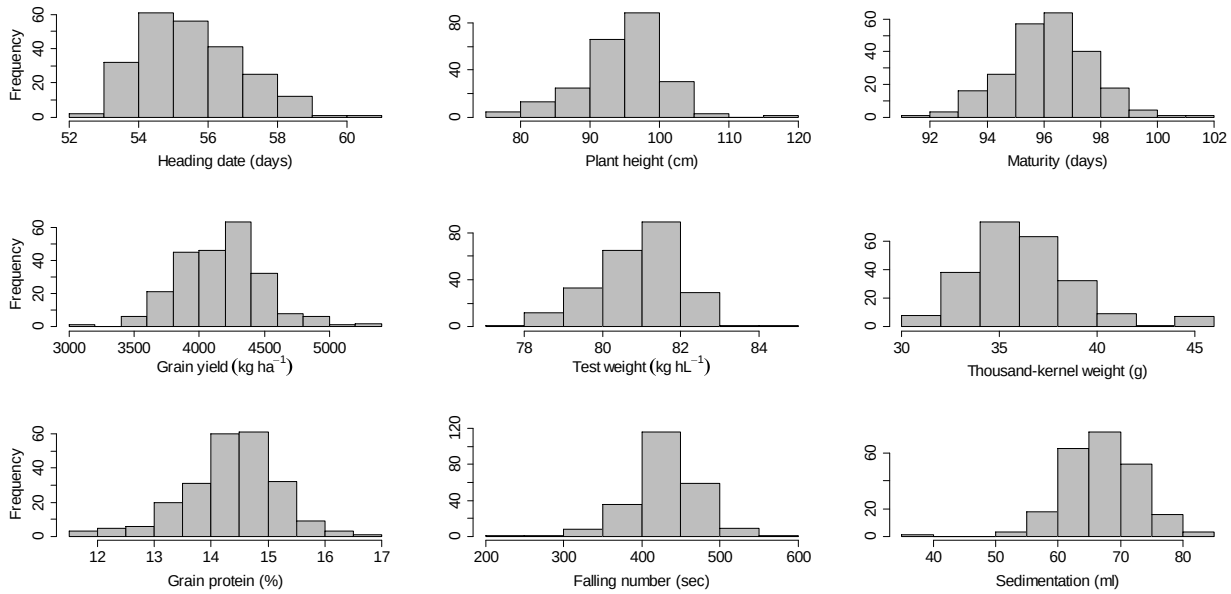



**Fig. 3-1.** Frequency distributions of nine agronomic and end-use quality traits used to estimate marker effects in the training population. Data were averaged across all environments.

**Table 3-2.** Broad-sense heritability estimates for nine agronomic and end-use quality traits.

| Trait† | Number of environments | Heritability |
|---|---|---|
| HD | 7 | 0.82 |
| HT | 9 | 0.62 |
| MAT | 8 | 0.58 |
| YLD | 10 | 0.51 |
| TWT | 9 | 0.57 |
| TKW | 10 | 0.84 |
| PRO | 9 | 0.64 |
| FN | 9 | 0.57 |
| SDS | 9 | 0.84 |

†HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, and TKW: thousand-kernel weight, PRO: grain protein, FN: falling number, SDS: sedimentation volume.

**Table 3-3.** Correlations among trait phenotypes averaged across environments.

| | | | Pearson's correlations, N = 271 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Trait†** | **Mean** | **SD‡** | **HD** | **HT** | **MAT** | **YLD** | **TWT** | **TKW** | **PRO** | **FN** |
| HD | 55.7 | 1.5 | 1 | | | | | | | |
| HT | 94.3 | 6.0 | 0.096NS | 1 | | | | | | |
| MAT | 96.3 | 1.6 | 0.612*** | -0.126* | 1 | | | | | |
| YLD | 4190.0 | 371.8 | 0.073NS | -0.295*** | 0.286*** | 1 | | | | |
| TWT | 80.9 | 1.0 | -0.018NS | -0.004NS | 0.379*** | 0.204*** | 1 | | | |
| TKW | 36.5 | 3.1 | 0.301*** | -0.189** | 0.407*** | 0.242*** | -0.017NS | 1 | | |
| PRO | 14.3 | 0.9 | -0.204*** | 0.364*** | -0.233*** | -0.329*** | 0.165** | -0.256*** | 1 | |
| FN | 425.7 | 41.8 | -0.024NS | 0.215*** | -0.337*** | -0.212*** | -0.113NS | -0.370*** | 0.319*** | 1 |
| SDS | 65.1 | 9.7 | -0.356*** | 0.116NS | -0.430*** | -0.051NS | 0.122* | -0.314*** | 0.426*** | 0.185** |

*, **, ***, Significant at the 0.05, 0.01 and 0.001 probability level, respectively, and *NS* not significant.

†HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, and TKW: thousand-kernel weight, PRO: grain protein, FN: falling number, SDS: sedimentation volume.

‡SD: Standard deviation of the trait.

### 3.3.2 Accuracy of single-trait genomic predictions

Prediction accuracies were different among the evaluated traits. The average prediction accuracies, based on Pearson's correlation, ranged from 0.55 to 0.77 across different model-traits combinations (Fig. 3-2). This was expected, because when the same TP is used to predict multiple traits, the prediction accuracy will be lower for some traits than for the other traits, in the same way that heritability is lower for some traits than for others (Combs and Bernardo, 2013). The accuracy obtained for thousand-kernel weight was better than accuracies obtained for all other traits. This indicates that some traits may be more amenable to GS in future breeding programs.

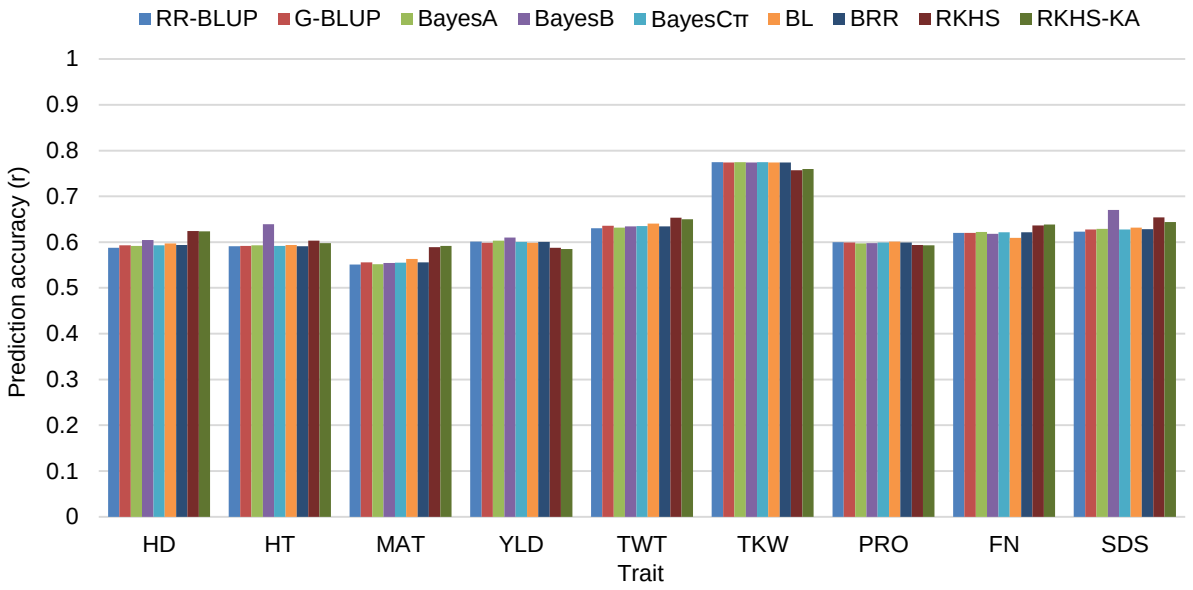



**Fig. 3-2.** Average prediction accuracy (from fivefold cross-validation) based on Pearson's correlation (*r*) between GEBVs estimated from nine statistical models and phenotypes for nine traits. HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight, PRO: grain protein, FN: falling number, SDS: sedimentation volume.

Although prediction accuracies were different among traits, they were not significantly different among different models (Fig. 3-2). These results agree with Heslot et al. (2012) who also reported a similar level of accuracy among eleven GS models and machine learning methods evaluated using wheat (*Triticum aestivum* L.), barley (*Hordeum vulgare* L.), *Arabidopsis thaliana*, and maize (*Zea mays* L.) data sets. Similarly, Daetwyler et al. (2013) compared several GS

methods using both simulated and real data sets and reported that based on simulated data, most methods performed similarly in traits influenced by large number of QTL, but in traits influenced by fewer QTL, variable selection methods did have some advantages. However, based on real data sets all methods had very similar accuracies. Charmet et al. 2014 also reported very similar accuracies with different prediction methods for heading date and test weight in a RIL population and two DH populations of wheat. Heslot et al. (2012) reported that the level of overfitting, computation time and the distribution of marker effect estimates varied widely among the models, although the level of accuracy was similar.

In this study, all models generated similar prediction accuracies perhaps because the evaluated traits are controlled by many minor effect loci. The main difference among the models is how they treat the prior distribution of the variance of marker effects. Ridge regression BLUP assigns uniform variance to all markers while Bayesian methods allow non-uniform variances for markers (Asoro et al., 2011; Meuwissen et al., 2001). Lorenz et al. (2011) indicated that models such as RR-BLUP are expected to work well for traits that are controlled by many loci with small effects, but variable selection methods such as BayesB should be preferred for traits that are controlled by few large effects QTL. In this study, BayesB showed 7 to 8% and 4 to 8% higher accuracies compared to the other models for plant height and sedimentation volume, respectively (Fig. 3-2). However, the average prediction accuracy was not significantly different among the evaluated methods across all traits. Similarly, Hayes et al. (2009c) reviewed the progress of GS for dairy cattle breeding in three different countries and concluded that for most dairy traits, the BLUP method with the assumption of many genes of small effects and few or none with moderate to large effects is close to reality.

We applied the RKHS to determine if including non-additive effects improves prediction accuracy. The RKHS captures both the additive and non-additive effects among loci; therefore, the predicted values are not GEBVs (Heslot et al., 2012). In this study, the RKHS and RKHS-KA models were not superior to the other models (RR-BLUP, G-BLUP, BayesA, BayesB, BayesCπ, BRR and BL), which are based on additive effects. The absence of significant difference between RKHS and other models in this study could be because the contribution of non-additive genetic effects to the total genetic variance was too small to be captured by the models. There are inconsistent reports on the benefit of including non-additive effects to improve GS prediction accuracy. Sallam et al. (2015) reported similar accuracies among simple additive models (RR-

BLUP and BayesCπ) and models that account for both additive and non-additive effects for the prediction of fusarium head blight resistance, yield and plant height using empirical data in barley. Lorenzana and Bernardo (2009) also reported no advantage of including epistatic effects for the prediction of genetic values using biparental maize, *Arabidopsis* and barley data sets. In contrast, Zhao et al. (2013) reported either equal or improved prediction accuracy when ignoring dominance effects in simulated and commercial hybrid wheat data sets. On the other hand, Pérez-Rodríguez et al. (2012) reported better predictive accuracy in RKHS compared to BL, BRR, BayesA and BayesB using empirical wheat data sets. Improved prediction accuracy was also reported in wheat when predictions were made using models that include epistasis (Crossa et al., 2010; He et al., 2016; Jiang and Reif, 2015; Wang et al., 2012). The discrepancies in the usefulness of including non-additive effects to improve accuracy of predictions could be due to the nature of the traits investigated or the population used in these studies. However, models that capture non-additive genetic effects may not be attractive for practical application of GS because gains from selection might be lower than expected in the long term.

The most appropriate model for GS may be different depending on the nature of trait and population of interest. Heslot et al. (2012) suggested that GS in plant breeding could use a reduced set of models, such as the BL, wBSR, and random forest regression. However, Daetwyler et al. (2013) indicated that no single method can serve as a benchmark for genomic prediction and recommended comparing accuracy and bias of new methods to results from G-BLUP and a variable selection approach such as BayesB, because these methods are appropriate for a range of genetic architectures. Lorenz et al. (2011) also indicated that there is no single best model across traits and populations because these models assume different genetic architectures. When models have comparable accuracies, it may also be advantageous to select a model that is less computationally demanding. In this study, the computation time for the Bayesian models was much longer because parameter estimates cannot be obtained analytically and must be estimated through repeated sampling from their posterior distributions, which requires several thousands of iterations.

### 3.3.3 Effect of trait heritability on prediction accuracy

In this study, there was no direct relationship between trait heritability and prediction accuracy (Table 3-2, Fig. 3-2). Based on heritability the traits can be ranked as: thousand-kernel

weight = sedimentation volume > heading date > grain protein > plant height > maturity > test weight = falling number > grain yield (Table 3-2). However, the prediction accuracy based on Pearson's correlation was comparable for all traits except for thousand-kernel weight (Fig. 3-2). Previous studies reported a strong relationship between the accuracy of genomic predictions and trait heritability, with the prediction being more accurate for traits with higher heritability than for traits with low heritability (Combs and Bernardo, 2013; Heffner et al., 2011a; Moser et al., 2010; Saatchi et al., 2010). However, this is not always the case because Combs and Bernardo (2013) reported situations where high prediction accuracy was attained for traits with low heritability. The lack of a direct relationship between trait heritability and prediction accuracy in this study could be because of the high heritability estimates obtained for all traits (Table 3-2).

### 3.3.4 Assessment of prediction accuracy based on rank correlation

Two different methods were used to assess model prediction accuracy in this study, Spearman's rank correlation and Pearson's correlation. Spearman's rank correlation was used to test for rank agreements for lines based on GEBVs and observed phenotypes. There was no significant difference in the average prediction accuracy based on Spearman's rank correlation among the evaluated methods for all traits (Fig. 3-3). However, BayesB showed 3 to 7% and 10 to 17% higher accuracies than the other models for plant height and sedimentation volume, respectively (Fig. 3-3). Rank correlations were generally lower for all traits except for heading date, test weight and falling number where the results were comparable to Pearson's correlations (Fig. 3-2 and 3-3). Rank correlation was low for grain protein content (in the range of 0.33 to 0.36) compared to Pearson's correlation (in the range of 0.59 to 0.60). Similarly, rank correlation was low for sedimentation volume (in the range of 0.43 to 0.46) compared to Pearson's correlation (in the range of 0.62 to 0.66). This indicated that the GEBVs did not capture the rankings of lines based on the actual phenotypes and this might affect selection decisions for these traits, especially when making truncation selection. Li et al. (2015) also reported that accuracies based on rank correlations were slightly lower than accuracies based on Pearson's correlation, but the rankings obtained for all methods exhibited the same trend for both correlations. This suggests that ranking of different models is similar based on Pearson's or rank correlations.

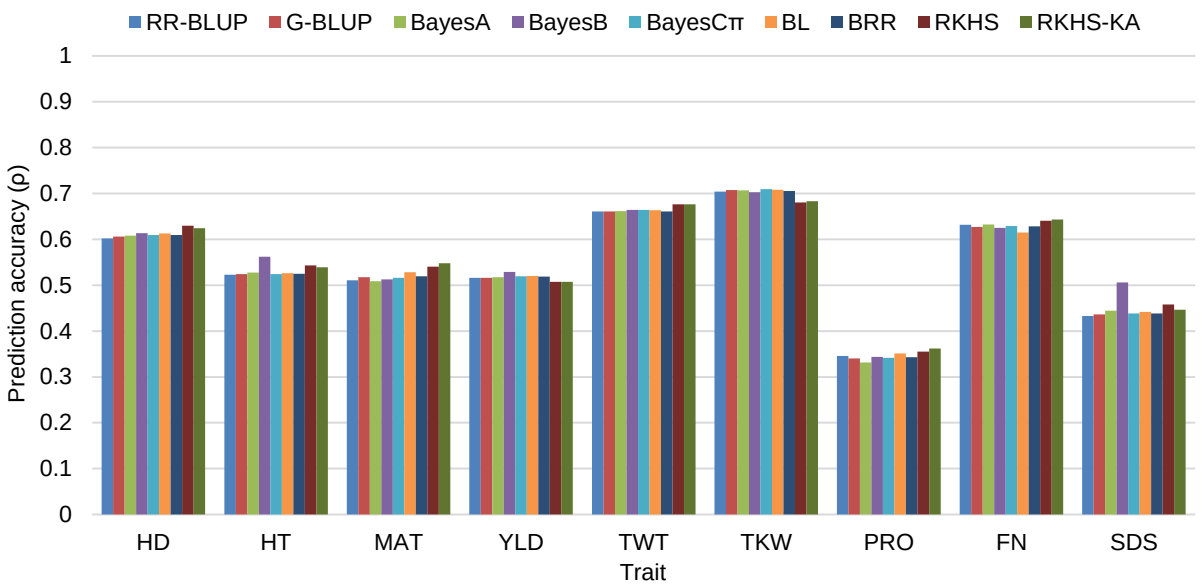



**Fig. 3-3.** Average prediction accuracy (from fivefold cross-validation) based on Spearman's rank correlation (ρ) between GEBVs estimated from nine statistical models and phenotypes for nine traits. HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight, PRO: grain protein, FN: falling number, SDS: sedimentation volume.

In addition to using different metrics for assessing prediction accuracies, we also performed cross-validation using two different methods; fivefold cross-validation and cross-validation using an independent validation set. The cross-validation results using an independent population were inconclusive. When the varcomp population was used as TP and the Co-op population for validation, prediction accuracy was substantially lower or comparable to fivefold cross-validation. Accuracies using the independent cross-validation method were lower for height (0.39-0.41), grain yield (0.28-0.31), grain protein (0.24-0.28), and sedimentation volume (0.31-0.38), were slightly lower for heading date (0.50-0.54), maturity (0.51-0.52), test weight (0.57-0.60), and thousand-kernel weight (0.65-0.67), and was similar for falling number (0.62) when compared to fivefold cross-validation (Fig. 3-4). All models showed comparable performance for each trait except for the RKHS that had 16 to 21% higher accuracy compared to the other models for sedimentation volume (Fig. 3-4). Similarly, when the Co-op population was used as TP and the varcomp population for validation, prediction accuracy was substantially lower for grain yield (0.31-0.37), and sedimentation volume (0.19-0.22), slightly lower for heading date (0.54-0.55) and test weight (0.49-0.55), similar for thousand-kernel weight (0.76-0.78), but slightly higher for maturity (0.56-

0.61), grain protein (0.54-0.64), and falling number (0.64-0.67) compared to the accuracy based on fivefold cross-validation (Fig. 3-5). Generally, a larger TP (Co-op) gave improved accuracy, except for test weight and sedimentation volume where slightly higher accuracy was obtained when the varcomp was used as TP. Previous reports also indicated that validation using an independent population of unrelated individuals is inferior to randomly drawn subsets of the same population (Goddard and Hayes, 2007; Thavamanikumar et al., 2015). This may be because some markers are correlated with a QTL in one population but not in the other (Goddard and Hayes, 2007). However, Habier et al. (2007) indicated that accuracy is non-zero even without LD because the accuracy of GEBVs depends on both tight LD between QTL and markers as well as the information from genetic relationships among individuals. The results of cross-validation using an independent population were satisfactory for most traits because we used elite breeding materials characterized by complex pedigrees and different levels of relatedness. Although the varcomp and Co-op populations were evaluated in the field separately, individuals in both populations have various levels of relatedness and cannot be considered as totally unrelated.

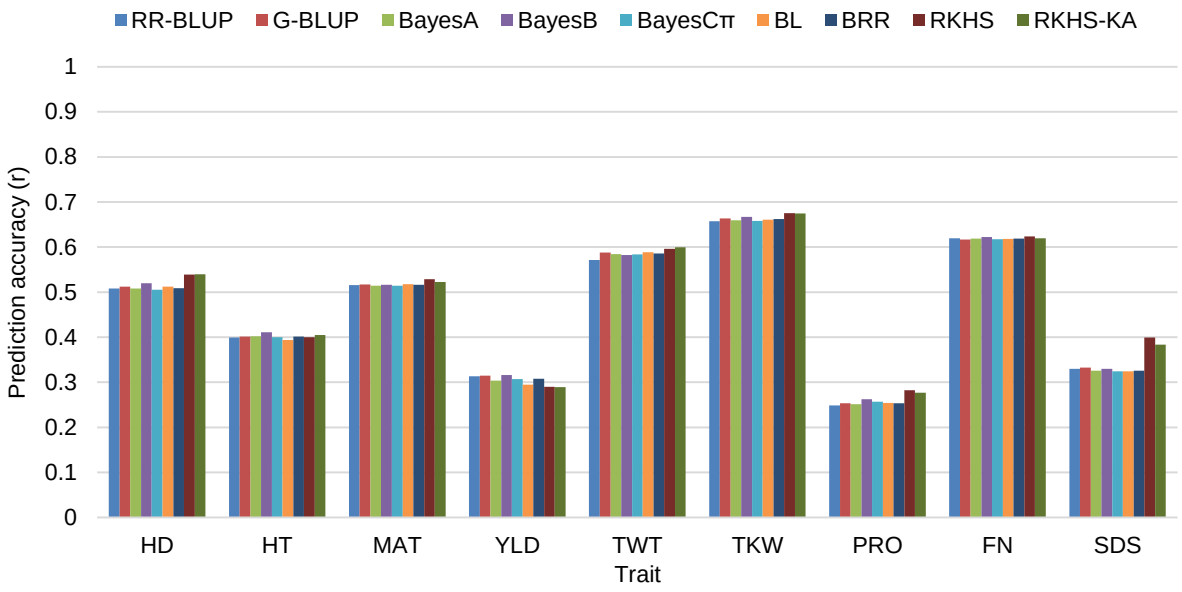



**Fig. 3-4.** Prediction accuracy based on Pearson's correlation (*r*) between GEBVs estimated from nine statistical models and phenotypes for nine traits. The varcomp population (77 lines) was used to estimate marker effects and the Co-op population (154 lines) was used as a validation set. HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight, PRO: grain protein, FN: falling number, SDS: sedimentation volume.

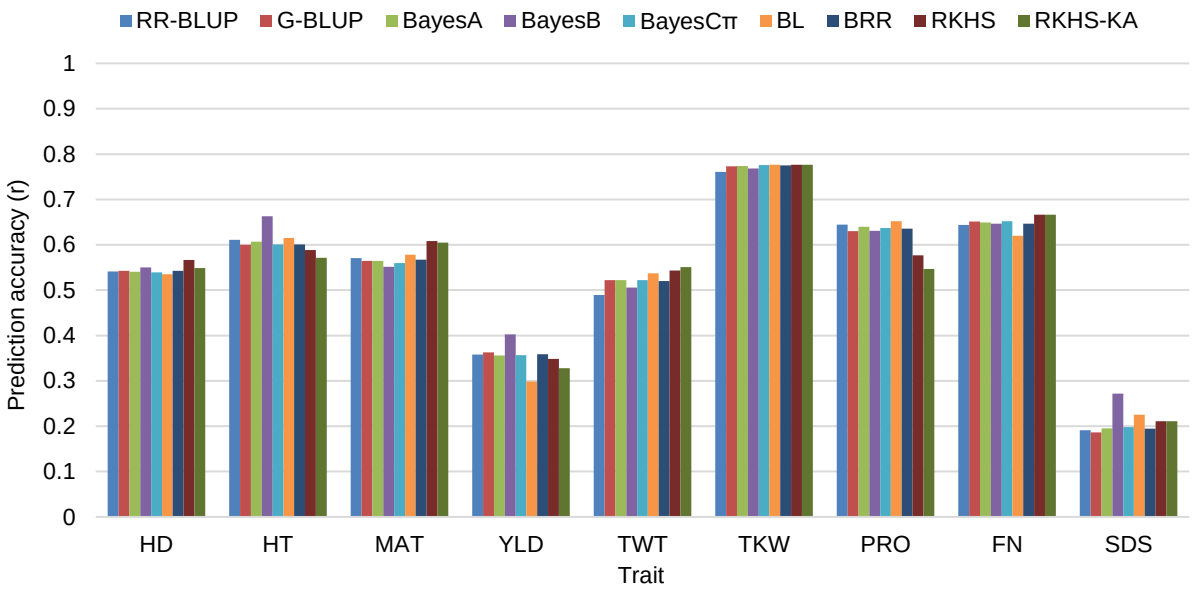



**Fig. 3-5.** Prediction accuracy based on Pearson's correlation (r) between GEBVs estimated from nine statistical models and trait phenotypes. The Co-op population (154 lines) was used to estimate marker effects and the varcomp population (77 lines) was used as a validation set. HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight, PRO: grain protein, FN: falling number, SDS: sedimentation volume.

### 3.3.5 Accuracy of multiple-trait genomic predictions

Breeders are often tasked with selecting for multiple traits simultaneously. Originally most GS models were designed for assessing a single trait; however, some multiple-trait prediction models have recently been developed which we assessed for their performance in making joint predictions for important traits in wheat. Predictive performance of the evaluated multiple-trait prediction models varied depending on the traits selected for joint prediction. Overall, none of the evaluated multiple-trait models gave consistently higher prediction accuracy than the single-trait BayesA model across different combinations of traits.

The prediction accuracy of the multiple-trait models was comparable to the single-trait model for joint prediction of grain yield and protein content (trait combination one) (Table 3-4). Lack of improvement in accuracy when yield and protein were jointly predicted could be due to negative correlation between these traits (Table 3-3). The negative correlation between traits in the TP might negatively affect the prediction of traits in the validation population. Jia and Jannink, (2012) indicated that multiple-trait models take advantage of genetic correlation between traits, whether it is favorable or unfavorable, but they are not designed to break the undesirable genetic

correlation. However, the observed negative correlations between grain yield and protein content may not have genetic causes. Earlier studies suggested that the negative phenotypic correlation between grain yield and protein content in wheat were not caused by genetic factors, but because of environmental factors, source-sink interactions and dilution of protein by carbohydrates (Kibite and Evans, 1984). When joint prediction was made for grain yield, test weight and thousand-kernel weight (trait combination two), multiple-trait models were not better than the single-trait model for prediction of grain yield, but slightly lower or comparable accuracy were obtained for test weight and thousand-kernel weight (Table 3-4). Similarly, when predictions were made for trait combination three (height, grain yield, and grain protein) and four (heading date, maturity and grain yield), no obvious benefit was obtained from all three multiple-trait models compared to the single-trait model for all traits (Table 3-4). Finally, when joint prediction was made for all nine traits (trait combination five), prediction accuracies of the multiple-trait models were either comparable or lower than the single-trait model for all traits except for sedimentation volume, where multiple-trait BayesA and multiple-trait BayesA Matrix models gave a slightly higher accuracy (Table 3-4).

The poor performance of the multiple-trait models may be because the traits in this study are controlled by several minor effect loci. Grain yield is a complex trait that is often difficult to maintain or improve while enhancing other correlated traits, as such it is an ideal candidate for multiple-trait prediction. However, the evaluated multiple-trait models performed poorly compared to the single-trait model for the prediction of grain yield in all trait combinations (Table 3-4). Jia and Jannink (2012) reported that the genetic architecture of a trait affects the relative advantage of multiple-trait models over single-trait models and multiple-trait genomic prediction captures the genetic correlation between traits more efficiently when major QTL are present. Jiang et al. (2015) also showed that the antedependence-based models were superior to the multiple-trait BayesA model when the number of simulated QTL was 30 but there was no difference in accuracy between these models when 300 QTL each having small effects were simulated. Overall, there was no advantage of the evaluated multiple-trait prediction models over the single-trait BayesA for all trait combinations, possibly because the traits in this study, including grain yield, are controlled by many QTL with small effects.

**Table 3-4.** Average and standard deviation of prediction accuracy (from fivefold cross-validation) based on Pearson's correlation (*r*) between trait phenotypes and GEBVs estimated using a single-trait and three multiple-trait prediction models. Prediction was made for different trait combinations.

| Trait combination | Trait† | Single-trait BayesA | Multiple-trait BayesA | Multiple-trait BayesA Matrix | Multiple-trait BayesA Scalar |
|---|---|---|---|---|---|
| 1 | YLD | 0.603(0.089) | 0.589(0.093) | 0.583(0.095) | 0.571(0.130) |
| | PRO | 0.597(0.149) | 0.577(0.174) | 0.593(0.156) | 0.581(0.172) |
| 2 | YLD | 0.603(0.089) | 0.542(0.148) | 0.559(0.141) | 0.578(0.116) |
| | TWT | 0.631(0.070) | 0.630(0.079) | 0.623(0.069) | 0.637(0.078) |
| | TKW | 0.775(0.042) | 0.762(0.057) | 0.758(0.058) | 0.768(0.047) |
| 3 | HT | 0.593(0.129) | 0.610(0.127) | 0.568(0.134) | 0.627(0.102) |
| | YLD | 0.603(0.089) | 0.551(0.162) | 0.528(0.115) | 0.570(0.099) |
| | PRO | 0.597(0.149) | 0.606(0.113) | 0.581(0.127) | 0.582(0.150) |
| 4 | HD | 0.592(0.110) | 0.588(0.102) | 0.586(0.086) | 0.593(0.110) |
| | MAT | 0.552(0.076) | 0.563(0.083) | 0.552(0.087) | 0.559(0.084) |
| | YLD | 0.603(0.089) | 0.554(0.118) | 0.580(0.084) | 0.548(0.113) |
| 5 | HD | 0.592(0.110) | 0.581(0.110) | 0.582(0.112) | 0.583(0.116) |
| | HT | 0.593(0.129) | 0.598(0.109) | 0.563(0.109) | 0.554(0.097) |
| | MAT | 0.552(0.076) | 0.537(0.087) | 0.533(0.085) | 0.534(0.093) |
| | YLD | 0.603(0.089) | 0.575(0.089) | 0.548(0.086) | 0.556(0.087) |
| | TWT | 0.631(0.070) | 0.619(0.079) | 0.608(0.074) | 0.603(0.065) |
| | TKW | 0.775(0.042) | 0.763(0.035) | 0.764(0.047) | 0.763(0.040) |
| | PRO | 0.597(0.149) | 0.574(0.135) | 0.573(0.149) | 0.543(0.155) |
| | FN | 0.622(0.049) | 0.531(0.046) | 0.553(0.076) | 0.507(0.111) |
| | SDS | 0.629(0.129) | 0.647(0.125) | 0.650(0.120) | 0.624(0.138) |

†HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, and TKW: thousand-kernel weight, PRO: grain protein, FN: falling number, SDS: sedimentation volume.

Genetic correlation between traits and physical correlation between markers may also affect multiple-trait GS model performance. Several studies reported improved prediction accuracy when multiple-trait GS models were used instead of single-trait models in the presence of genetic correlation between traits (Calus and Veerkamp, 2011; Guo et al., 2014a; Jiang et al., 2015; Tsuruta et al., 2011). However, in the absence of genetic correlation, multiple-trait models were inferior to

single-trait models (Jia and Jannink, 2012). This could be due to a nonzero estimate of genetic correlation between traits in the TP and using this wrong information to predict traits in the validation population (Jia and Jannink, 2012). The two antedependence-based multiple-trait prediction models evaluated in this study (Multiple-trait BayesA Matrix and Scalar) were designed to benefit from correlations between traits as well as between SNP effects to enhance whole genome prediction (Jiang et al., 2015). Based on simulated and real mice data sets, Yang and Tempelman (2012) reported that the prediction accuracy of antedependence-based models (ante-BayesA and ante-BayesB), which consider SNP effects as being spatially correlated based on their relative physical location along the chromosome, was up to 3.6% higher compared to their classical counterparts (BayesA and BayesB). Similarly, based on simulated data, Jiang et al. (2015) reported that the prediction accuracy of the antedependence-based model was 2.8 to 8.6% higher compared to the accuracy of the best single-trait prediction model. However, based on real mice data set, the antedependence-based models were slightly better than the single and multiple-trait BayesA models (Jiang et al., 2015). In this study, there was no consistent benefit of the antedependence-based models compared to the single-trait model or multiple-trait BayesA model that only takes advantage of the genetic correlation between traits.

Heritability also has an important role in making accurate multiple-trait predictions. Based on simulated data, several studies reported improved prediction accuracy for traits with low heritability ($h^2 < 0.2$) when a correlated trait with higher heritability ($h^2 \geq 0.5$) was included in multiple-trait prediction models (Guo et al., 2014a; Hayashi and Iwata, 2013; Jia and Jannink, 2012; Jiang et al., 2015). Jia and Jannink, (2012) reported that the prediction accuracy for a low heritability trait ($h^2 = 0.1$) increased from 0.49 obtained from a single-trait model to 0.67 and 0.70 when it was jointly predicted with a medium ($h^2 = 0.5$) and a high heritability trait ($h^2 = 0.8$) using multiple-trait prediction model. However, the prediction accuracy for the medium or high heritability traits did not change as the heritability of the correlated trait changed and when the genetic correlation between traits increased from 0.1 to 0.9 (Jia and Jannink, 2012). Similarly, Jiang et al. (2015) reported that a low heritability trait ($h^2 = 0.1$) benefited from joint prediction of multiple-traits but no obvious advantage was observed for a high heritability trait ($h^2 = 0.5$). Using a Bayesian multiple-trait prediction model, Hayashi and Iwata (2013) also reported increased prediction accuracy for a low heritability trait ($h^2 = 0.1$) that had genetic correlation of 0.7 with a high heritability trait ($h^2 = 0.8$); however, there was no difference in prediction accuracy between

multiple-trait and single-trait models for the high heritability trait. Guo et al. (2014a) also showed that the multiple-trait model gave more accurate prediction than the single-trait model for a low heritability trait ($h^2 = 0.05$), but there was no difference between these models for a high heritability trait ($h^2 = 0.3$). In our study, the lowest heritability estimate was obtained for grain yield ($H^2 = 0.51$) (Table 3-2), which was equivalent to the highest heritability considered in these simulation studies. The high heritability estimates of the traits in our study could be the reason why we did not find improved prediction accuracy when multiple-trait models were used instead of the single-trait prediction model.

### 3.3.6 Prediction accuracy of GS + de novo GWAS model

Another GS approach assessed in this study used significant markers from GWAS as fixed effects when making genomic predictions. Prediction accuracies were evaluated for the standard RR-BLUP and GS + de novo GWAS model that included one to three of the most significant SNPs identified from fold-specific GWAS as fixed effects. Strong marker trait associations were not detected after FDR-based multiple testing correction for all traits except plant height and sedimentation volume (Appendix C). After including significant markers as fixed effects, no significant difference was observed between accuracies obtained from the standard RR-BLUP and GS + de novo GWAS models for all traits (Fig. 3-6). However, including up to three significant markers detected from GWAS resulted in 13 and 7% higher accuracy for plant height and sedimentation volume, respectively. In contrast, Spindel et al. (2016) reported that the GS + de novo GWAS model outperformed the standard RR-BLUP in all cases. Gains in prediction accuracy up to 30% were reported for flowering time, for which a large GWAS peak was identified (Spindel et al., 2016). Spindel et al. (2016) reported that GS + de novo GWAS works best for traits with one or more medium to large effect QTL segregating in the population but it has no advantage if a trait has no significant GWAS peaks or peaks that are very near the significance threshold after applying multiple testing correction. Spindel et al. (2016) recommended the GS + de novo GWAS model over alternatives such as RR-BLUP or random forest regression when the -log (FDR corrected *P*-values of the most significant SNPs) $\geq 2.0$. Plant height and sedimentation volume were the only traits from this study to have an observed increase in accuracy for GS + de novo GWAS when compared to RR-BLUP, likely because these were also the only traits that had significant markers in all five folds after accounting for the FDR.

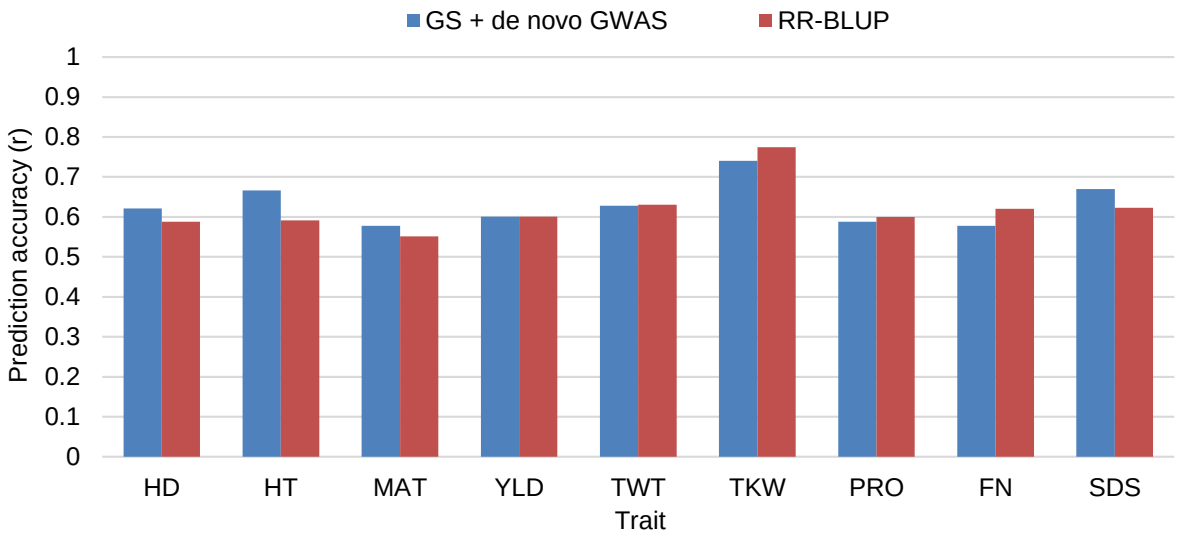



**Fig. 3-6.** Average prediction accuracy (from fivefold cross-validation) based on Pearson's correlation (*r*) between GEBVs estimated using RR-BLUP and GS + de novo GWAS models and trait phenotypes. HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight, PRO: grain protein, FN: falling number, SDS: sedimentation volume.

Trait heritability and the proportion of genetic variance explained by the QTL are important considerations when deciding to fit markers associated with known genes or QTL as fixed effects. Based on a simulation study, Bernardo (2014) indicated that fitting markers associated with a known major gene as fixed effect is more advantageous in GS when the percentage of the genetic variance explained by the gene ($R^2$) and trait heritability ($h^2$) are high. Bernardo (2014) reported that the ratios of selection response when one or more major genes had a fixed effect versus when none of the major genes had a fixed effect was always significantly greater than one with $R^2 \geq 25\%$ and $h^2 \geq 0.50$. In this study, estimates of broad-sense heritability ranged from 0.51 to 0.84 for all traits (Table 3-2), while the $R^2$ values of the markers fitted as fixed effects ranged from 7 to 11% (heading date), 7 to 15% (plant height), 6 to 8% (maturity, grain yield and test weight), 5 to 7% (thousand-kernel weight), 7 to 12% (grain protein and sedimentation volume), 6 to 12% (falling number) (Appendix C). Bernardo (2014) reported that modelling a single major gene as fixed effect was never disadvantageous, except when $R^2 < 10\%$. Spindel et al. (2016) also indicated that adding fixed markers identified from GWAS never decreased accuracy compared to RR-BLUP or other statistical methods tested. Similarly, modelling markers having $R^2 < 10\%$ as fixed

effects never reduced accuracy in this study. Because heritability for most traits was high in this study, lack of improvement in accuracy of GS + de novo GWAS could be due to the small $R^2$ value of the markers fitted as fixed effects.

In addition to fitting significant markers from GWAS in the TP as fixed effects, as was performed in this study, significant markers from a different population could also be incorporated into GS as fixed effects. Markers of known effect, markers tagging candidate genes, or markers tagging previously identified QTL can also be included in GS models as fixed effect (Bentley et al., 2014; Spindel et al., 2016; Zhang et al., 2014). However, Bentley et al. (2014) indicated that significant marker selection can lead to bias because some of these markers may not have any effect within the TP. Zhang et al. (2014) used existing GWAS results from publicly available databases to build trait-specific GRM by assigning large weights to significant markers to account for the variance explained by the corresponding loci. Zhang et al. (2014) tested this new approach using real cattle and rice data sets and reported improved accuracy of genomic predictions compared to the standard G-BLUP and BayesB models. Spindel et al. (2016) also compared GS + de novo GWAS model with GS + historical GWAS model, a method equivalent to the first model but the markers fitted as fixed effects were selected from previously published GWAS results. The results indicated that the accuracy of GS + historical GWAS model was either similar or worse than the GS + de novo GWAS model depending on the trait and environment. This indicates that significant SNPs identified from previously published GWAS results may not always be relevant to a given population. On the other hand, Zhao et al. (2014) reported improved prediction accuracy using a different approach that assigns larger weights to known functional markers (less shrinkage) independently of the genome-wide markers. This suggests that instead of fitting random markers associated with a trait, modelling functional markers or markers located within gene coding sequence as fixed effects might enhance GS prediction accuracy.

### 3.4 Conclusion

We evaluated single-trait, multiple-trait and GS + de novo GWAS models for the prediction of several agronomic and end-use quality traits using empirical data in spring bread wheat. Comparison of different statistical methods in this study suggested that GS can be implemented in wheat using either G-BLUP or BayesB models. G-BLUP is computationally efficient because the dimension of the marker data is reduced when using the GRM. Though more computationally

demanding, BayesB gave more accurate predictions for plant height and sedimentation volume. Prediction accuracy of multiple-trait and single-trait models was similar. For some traits, multiple-trait prediction accuracy was lower than single-trait prediction accuracy, but this was dependent on the inter-trait correlation and the genetic architecture of the traits. The GS + de novo GWAS model that included up to three significant SNPs as fixed effects performed similar to the standard RR-BLUP for most traits, likely because few large effect QTL were identified in this population. The GS + de novo GWAS model might be promising when major genes exist and should be investigated further.

Overall, the prediction accuracies based on fivefold cross-validation are encouraging given the small size of the TP used in this study. Different cross-validation techniques gave variable results indicating that it is important to choose a cross-validation technique that mimics the actual prediction problem. In GS, prediction accuracy is commonly assessed by systematically partitioning the same population into training and validation sets. However, the ultimate use of GS is to predict breeding values in a population different from the one used to estimate marker effects. Therefore, the cross-validation technique should resemble the actual application of GS and across-population cross-validation would be more realistic when the intention is to make genomic predictions in a population different from the one used to estimate marker effects.

# 4. Effects of Marker Density, Training Population Size and Population Structure on GS Accuracy in Wheat

## 4.1 Introduction

Genomic selection is growing in popularity as a tool for crop breeding, but empirical studies outlining the factors influencing GS prediction accuracy in wheat are limited. With the increasing application of GS, numerous statistical methods have been proposed to predict GEBVs from genotypic data. In Chapter three, we investigated some of these methods and determined that, for the most part, models have similar prediction accuracies and single-trait models such as G-BLUP and BayesB may offer the best trade-offs for computational demand and accuracy for the wheat populations used in this study. Studies based on simulated and empirical data indicated that the accuracy of GS model prediction depends on a number of other factors related to the marker type and density (Combs and Bernardo, 2013; Heffner et al., 2011a; Moser et al., 2010; Poland et al., 2012; Solberg et al., 2008), size of the TP (Calus and Veerkamp, 2007; Combs and Bernardo, 2013; Heffner et al., 2011a; VanRaden et al., 2009), trait heritability (Combs and Bernardo, 2013; Moser et al., 2010), genetic relationships between individuals in the training and validation populations (Solberg et al., 2008; VanRaden et al., 2009; Wientjes et al., 2013), population structure (de los Campos et al., 2015; de Roos et al., 2009; Guo et al., 2014b; Isidro et al., 2015), G × E (Burgueño et al., 2011; Crossa et al., 2015; Heslot et al., 2014; Jarquín et al., 2014a; Lopez-Cruz et al., 2015; Pérez-Rodríguez et al., 2015) and the statistical method used for prediction (Calus, 2010). These factors are interrelated in a complex manner (Desta and Ortiz, 2014). Moreover, most of these factors are population and environment specific and it is imperative to assess the predictive ability of different statistical models and model parameters for different breeding populations and environments. Therefore, the objectives of this study were to i) determine GS model prediction accuracy under increasing TP size and marker density, and ii) examine the effect of population structure on prediction accuracy.

## 4.2 Materials and Methods

A population of 231 hexaploid wheat varieties and advanced breeding lines were used in this study to assess the effects of TP size, marker density and population structure on genomic

prediction accuracy. Detailed descriptions of the phenotypic and genotypic data are provided in Chapter three.

### 4.2.1 Effect of training population size on genomic prediction accuracy

To determine the effect of TP size on model prediction accuracy, three TP sizes ($N_{TP}$ = 50, 100, and 200) were used with the full marker density (17,887 SNPs) to predict GEBVs using G-BLUP. Predictions were made for heading date, plant height, maturity, grain yield, test weight, thousand-kernel weight, grain protein content, falling number and sedimentation volume using each TP size. For each trait-TP size combination, prediction accuracy was assessed using a cross-validation design of five random training-validation partitions. Models were fitted using the R package BGLR, v1.0.4 (Pérez and de los Campos, 2014). In each fold, prediction accuracy was assessed by calculating Pearson's correlation between GEBVs and observed phenotypes of the lines in the validation set. The averages of the five folds were reported.

### 4.2.2 Effect of marker density on genomic prediction accuracy

The effect of marker density on prediction accuracy was assessed using five marker densities (770, 3K, 13K, 15K, and 18K SNPs). The first two marker densities were selected from 15K polymorphic SNPs, that have position on the hexaploid wheat consensus map (Wang et al., 2014a), based on genetic distances using the software MapThin, v1.11 (Howey and Cordell, 2011). The 770 SNPs were selected based on a density of one marker for every 4.5 cM distance on average (770 SNPs on a 3,535-cM map). The 3K set was based on a density of one marker for every 1.2 cM distance on average (2,817 SNPs on a 3,535-cM map). The 13K set was obtained by removing markers that were in complete LD (i.e. $r^2$ between marker scores equal to one) from the 18K set. The 15K set included 15,248 SNPs that have a position on the hexaploid wheat consensus map, and the 18K set included all 17,887 polymorphic SNPs. Separate predictions were made for heading date, plant height, maturity, grain yield, test weight, thousand-kernel weight, grain protein content, falling number and sedimentation volume using these sets of markers. The G-BLUP model was fitted in R using the package BGLR (Pérez and de los Campos, 2014). Prediction accuracy was assessed using fivefold cross-validation and Pearson's correlation, as described in chapter three.

### 4.2.3 Effect of population structure on genomic prediction accuracy

Distance-based phylogenetic analysis was performed to determine population structure using DARwin, v5.0.158 (Perrier and Jacquemoud-Collet, 2006). A dissimilarity matrix was calculated from the marker data based on Manhattan distance using a bootstrap test with 1,000 cycles of resampling. These bootstrapped dissimilarities were used to construct a phylogenetic tree based on neighbour-joning method and the weighted pair group method using average (WPGMA) clustering criteria (Saitou and Nei, 1987). The phylogenetic tree was displayed on the tree generator iTOL, v3 (Letunic and Bork, 2016). Based on the number of subpopulations identified from the tree, lines were clustered using marker-based K-means clustering (Hartigan and Wong, 1979). The function 'kmeans' was used in R (R Development Core Team, 2016), to partition the 231 lines into three groups. The effect of population structure on the prediction of grain yield was assessed by modelling genetic heterogeneity into components that are constant across groups and components that are group specific as described in de los Campos et al. (2015). Three approaches were compared using 18K markers: i) interaction model, ii) across-group model and, iii) stratified or within-group model.

**i) Interaction model:** In this model, marker effects are decomposed into components that are constant across groups ($b_{0k}$ where $k$ = 1, 2, ... $p$ refers to markers) and components that are group specific ($b_{1k}$, $b_{2k}$ and $b_{3k}$ for groups 1, 2 and 3, respectively). The marker effects for the three groups are indicated by: $\beta_{1k} = b_{0k} + b_{1k}$, $\beta_{2k} = b_{0k} + b_{2k}$ and $\beta_{3k} = b_{0k} + b_{3k}$. The regression equation can be indicated in a matrix format as:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1\mu_1 \\ 1\mu_2 \\ 1\mu_3 \end{bmatrix} + \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} b_0 + \begin{bmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} \quad (4.1)$$

Where $y_1$, $y_2$, $y_3$, and $X_1$, $X_2$, $X_3$ represent the phenotypes and genotypes of the individuals in groups 1, 2 and 3, respectively, $\mu_1$, $\mu_2$, and $\mu_3$ are group specific intercepts, $b_0$, $b_1$, $b_2$, and $b_3$ are vectors of marker effects and $\varepsilon_1$, $\varepsilon_2$, and $\varepsilon_3$ represent model residuals.

**ii) Across-group model:** This model assumes constant marker effects across groups and can be obtained by setting $b_1 = b_2 = b_3 = 0$. This approach is equivalent to fitting a model to the whole data set ignoring population structure. The regression equation in a matrix format becomes:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1\mu_1 \\ 1\mu_2 \\ 1\mu_3 \end{bmatrix} + \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} b_0 + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} \quad (4.2)$$

**iii) Stratified or within-group model:** In this model, regression of phenotypes on markers was conducted separately in each group. This can be achieved by setting $b_0$ = 0; and the regression equation in a matrix format becomes:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1\mu_1 \\ 1\mu_2 \\ 1\mu_3 \end{bmatrix} + \begin{bmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} \quad (4.3)$$

The interaction, across-group and stratified approaches were fitted using BRR and BayesB models implemented in the R package BGLR, v1.0.4 (Pérez and de los Campos, 2014). Prediction was made using 50 random training-validation partitions. In each training-validation partition, 187 lines (24 in group one, 76 in group two and 87 in group three) were used as a training set and 44 lines (6 in group one, 17 in group two and 21 in group three) for cross-validation. Prediction accuracy was assessed within each group using Pearson's correlation between predicted GEBVs and phenotypes of individuals in the validation sets. In each training-validation partition, predictions were based on 50,000 iterations obtained after discarding 5,000 samples as a burn-in.

## 4.3 Results and Discussion

### 4.3.1 Genomic prediction with increasing training population sizes

The size of the TP is the most important factor that affects genomic prediction accuracy. Increasing the TP from 50 to 100 increased the prediction accuracy for all traits on average by 11% (ranging from 3 to 18%), except for grain yield (Fig. 4-1). When the TP was increased from 100 to 200, the average prediction accuracy increased for all traits on average by 18% (Fig. 4-1). The highest increase was observed for grain yield (36%), while the lowest increase was observed for thousand-kernel weight (7%). The small increase for thousand-kernel weight was because a high prediction accuracy (r = 0.73) was already obtained using 100 lines in the TP (Fig. 4-1). Overall, increasing the TP size from 50 to 200 resulted in an increase in prediction accuracy ranging from 20 to 35% and an average increase of 28% for all traits. The benefit of increasing sample size to increase the power of a statistical test is a well-established principle in statistics. In

GS, several studies reported that increasing the TP size increases the accuracy of GS by providing more data to estimate marker effects (Asoro et al., 2011; Meuwissen et al., 2001; Muir, 2007; Saatchi et al., 2010; VanRaden et al., 2009). Meuwissen et al. (2001) showed that reducing the TP from 2200 to 500 reduced prediction accuracies by 61, 27, and 17% for least square, BLUP, and BayesB estimation methods, respectively. In winter wheat, Heffner et al. (2011a) reported that decreasing the TP size from 288 to 192 and 96 reduced the average genomic prediction accuracy by 11 and 30%, respectively. In a similar study, Bentley et al. (2014) tested the effect of TP size on the prediction of agronomic and end-use quality traits in European wheat based on bootstrap resampling of 50, 100 and 200 individuals as TP. Improved prediction accuracy was reported for flowering time, thousand-kernel weight and protein content as the TP size increased, but no marked improvement in accuracy was observed for grain yield (Bentley et al., 2014). Goddard and Hayes (2007) speculated that the accuracy of estimating marker effects will approach one as the total number of individuals with phenotypes and marker genotypes increases.

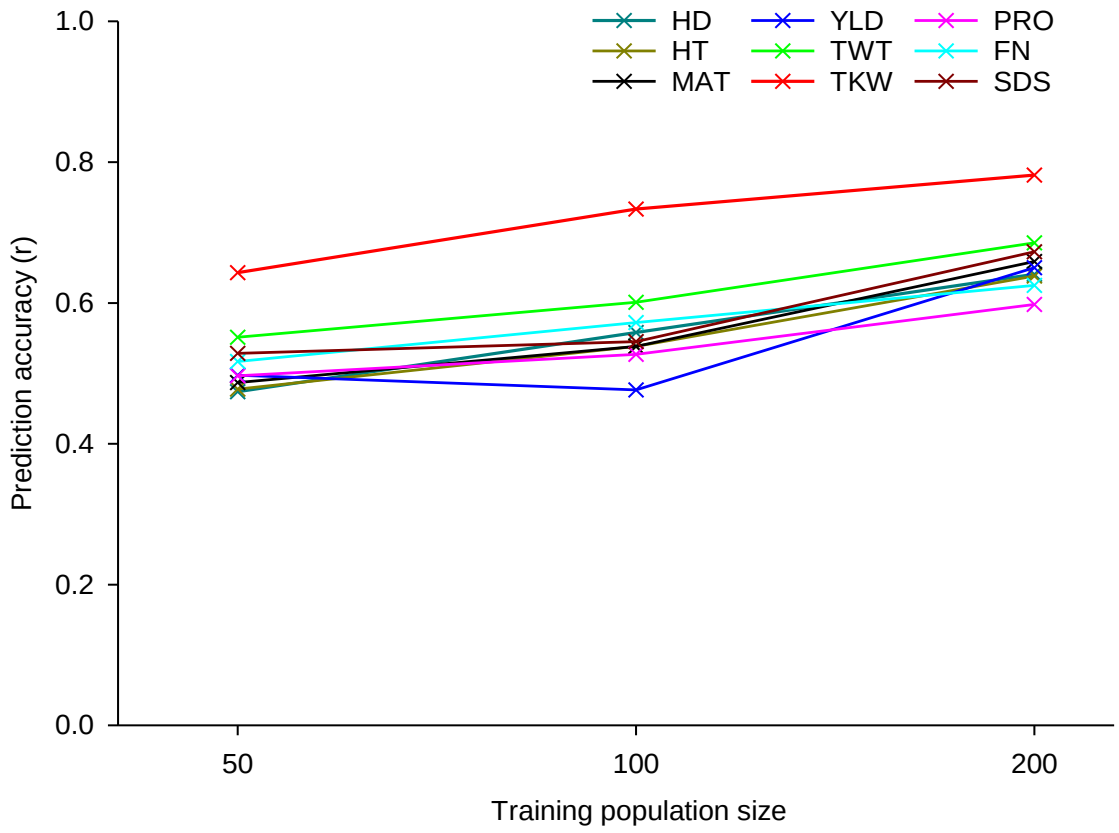



**Fig. 4-1.** Average prediction accuracy (from five training-validation partitions) based on Pearson's correlation (*r*) between GEBVs and trait phenotypes plotted against training population size. HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight, PRO: grain protein, FN: falling number, SDS: sedimentation volume.

When TP size is sufficiently large, even traits of low $h^2$ can be predicted quite accurately (Lorenz et al. 2011). Combs and Bernardo (2013) indicated that the product of $h^2$ and TP size is the key factor that determines prediction accuracy rather than $h^2$ and TP size individually. This suggests that traits with initially low $h^2$ can be evaluated with larger TP size to maximize prediction accuracy (Combs and Bernardo, 2013). Hayes et al. (2009b) showed that nearly 9,000 individuals are required to get a prediction accuracy of 0.7 for a trait with $h^2 = 0.2$, but when $h^2 = 0.8$, about 1,000 individuals are required to get similar prediction accuracy. A decrease in prediction accuracy due to reduced $h^2$ can be compensated by using a larger number of observations to estimate marker effects (de Roos et al., 2009; Saatchi et al., 2010; Solberg et al., 2008). However, the correlation of GEBVs with phenotype can never exceed the square root of $h^2$ (Charmet and Storlie, 2012).

Heritability is the upper limit of the phenotypic variance that can be explained based on genetic markers (Wray et al., 2013). In addition, genomic prediction never accounts for all phenotypic variation because the upper limit can be achieved only if all genetic variants that affect the trait are known and if their effects are estimated without error (Wray et al., 2013). Together, there is interplay between $h^2$ and TP size, both of which are desired for making accurate genomic predictions.

### 4.3.2 Genomic prediction with increasing marker density

In this study, five different marker densities were used to predict agronomic and end-use quality traits in wheat. The results showed that there was no difference in prediction accuracy for all traits when we adjusted the number of markers (Fig. 4-2). Mean accuracy for all traits ranged from 0.57 to 0.75 and 0.57 to 0.76 when genomic predictions were made using 770 and 3K evenly spaced SNPs, respectively. Similarly, prediction accuracy ranged from 0.56 to 0.77 when predictions were made using 13K, 15K or 18K SNPs indicating no improvement in accuracy with increasing marker density.

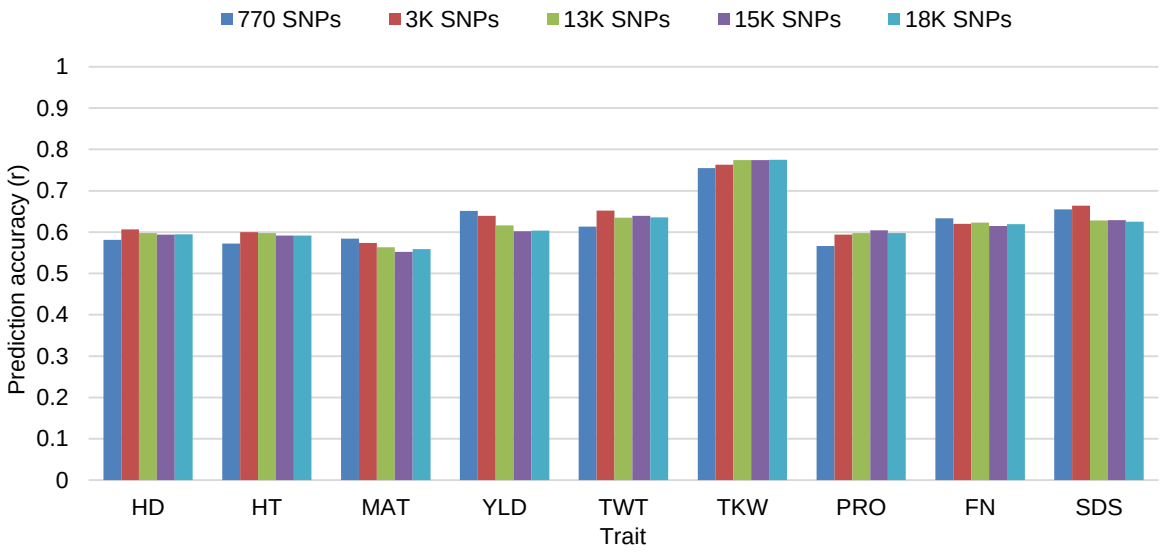



**Fig. 4-2.** Average prediction accuracy (from fivefold cross-validation) based on Pearson's correlation (*r*) between GEBVs estimated for different marker densities and trait phenotypes. HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight, PRO: grain protein, FN: falling number, SDS: sedimentation volume.

The reason why prediction accuracy did not increase when more markers were used could be because only 770 SNPs were sufficient to accurately estimate the genomic relationships among the lines. We used G-BLUP to evaluate the effect of marker density on genomic prediction accuracy. Meuwissen et al. (2016) indicated that G-BLUP is expected to yield little improvement in accuracy with increasing marker density. This is because genomic relationships can be accurately estimated using lower marker density and further increase in marker density barely improves the accuracy of estimating genomic relationships and thus GEBVs (Meuwissen et al., 2016). However, Bayesian variable selection models that attempt to identify the casual SNPs are expected to benefit more from a higher marker density (Meuwissen et al., 2016). In this study, prediction accuracies were similar among the evaluated marker densities when we used BayesB. This could be because only 770 SNPs were sufficient to capture the effects of most QTL affecting the traits. Heffner et al. (2009) indicated that the benefit of increasing marker density is to maximize the number of QTL in LD with at least one marker, which also maximizes the number of QTL whose effects will be captured by markers, and this may have already been maximized at 770 SNPs in this study. Moser et al. (2010) reported that a low-density assay of evenly spaced SNPs can provide sufficient prediction accuracies if the information content of the subset of SNPs is sufficient to estimate effects of distinct haplotypes. Bassi et al. (2016) also indicated that even marker distribution across the genome and their ability to capture important QTL underlying traits are important considerations for GS than marker number. Heffner et al. (2011a) reported that reducing marker density from 1158 to 768 and 384 resulted in a small decrease in prediction accuracy in winter wheat; however, a further reduction to 192 reduced the average GS accuracy by 10%. Asoro et al. (2011) used 1,005 DArT markers for the prediction of agronomic and quality traits in oat and reported improved prediction accuracy when the number of markers was increased from 300 to 600 and from 600 to 900. However, the increase in accuracy with marker number was dependent on traits, and for some traits accuracy reached a plateau at 600 markers (Asoro et al., 2011). In cattle, Moser et al. (2010) showed that accuracies were very sensitive when fewer than 1,000 SNPs were used, but accuracy reached a plateau when SNP density exceeded 1,000. An empirical study based on biparental and mixed populations of maize, wheat, and barley also indicated that the accuracy of GS predictions increased as the number of markers increased; however, gains in prediction accuracy began to plateau at 40 to 80% of the total marker density (Combs and Bernardo, 2013). Gains in accuracy began to plateau despite the small number of

markers used: 1213 markers (maize biparental population), 223 markers (barley biparental population), 1178 markers (barley mixed population), and 731 markers (wheat mixed population) (Combs and Bernardo, 2013). The lack of increase in prediction accuracies after a certain number of markers indicated marker saturation in the populations (Combs and Bernardo, 2013). Lorenz et al. (2011) also observed no difference in accuracy when prediction was made for plant height and grain yield using three randomly sampled marker numbers (300, 600, and 900) in barley, oat and wheat data sets. Similarly, in cattle, the accuracy of prediction for several traits was not significantly different when up to 75% of the original markers were masked (Luan et al., 2009; VanRaden et al., 2009). Overall, our results agree with other studies, which indicate that a reduced subset of evenly spaced markers can be sufficient for GS. This suggests that a reduced subset of SNP arrays could be designed for GS, which would reduce the genotyping cost and computational time for genomic predictions.

Other explanations for the minimal effect of marker density on GS prediction accuracy could be the size of the TP. When considering the TP size, there may not be enough degrees of freedom to benefit from the increase in marker number. Sample size is one of the most important factors limiting GS prediction accuracy (de los Campos et al., 2015). Muir (2007) reported that increasing the number of markers without increasing the TP size may reduce accuracy because it increases collinearity among markers. Other research indicated that increasing the number of phenotypic records has a more important effect on accuracy than marker number (VanRaden et al., 2009; Lorenz et al., 2011). Although we achieved maximum accuracies using the smallest number of SNPs, 770, including additional SNPs may improve accuracy if we also increase the TP size.

### 4.3.3 Genomic prediction accounting for population structure

The population was divided into three major clusters using neighbor-joining and a Manhattan dissimilarity matrix (Fig. 4-3). Cluster one (blue) and cluster two (pink) were more related to each other than cluster three (green). Then, marker-based K-means clustering was performed to partition the 231 wheat lines into three groups, with the aim of increasing within-group and decreasing between-group relatedness for cross-validation (Fig. 4-4). The clustering based on the K-means mostly coincided with the clustering obtained with the phylogenetic tree except for a few lines at the edges of the clusters, which could be attributed to differences between the two

approaches. There were 30, 93, and 108 wheat lines in groups one, two, and three, respectively (Fig. 4-4).

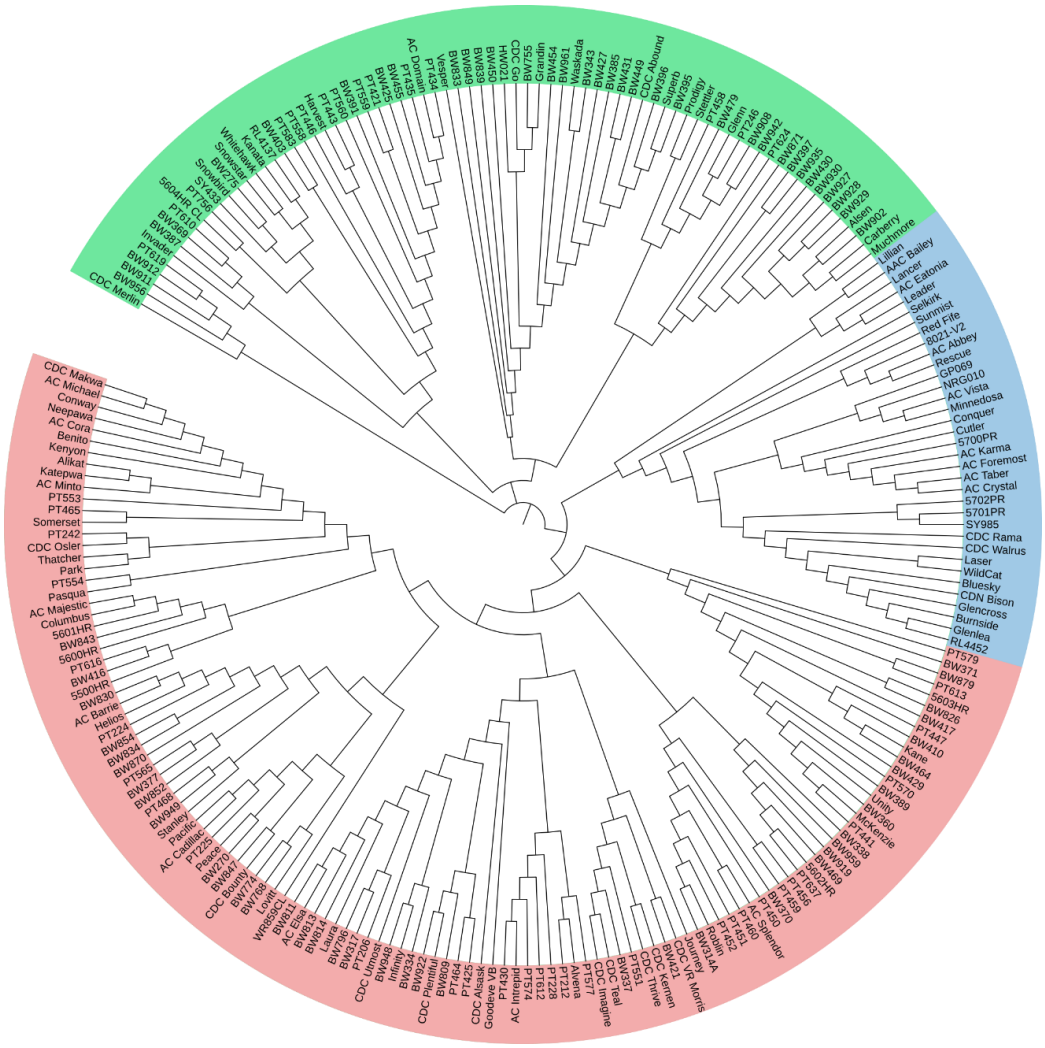



**Fig. 4-3.** A phylogenetic tree based on neighbor-joining and a Manhattan dissimilarity matrix showing clustering of the 231 wheat lines into three groups represented in different colors.

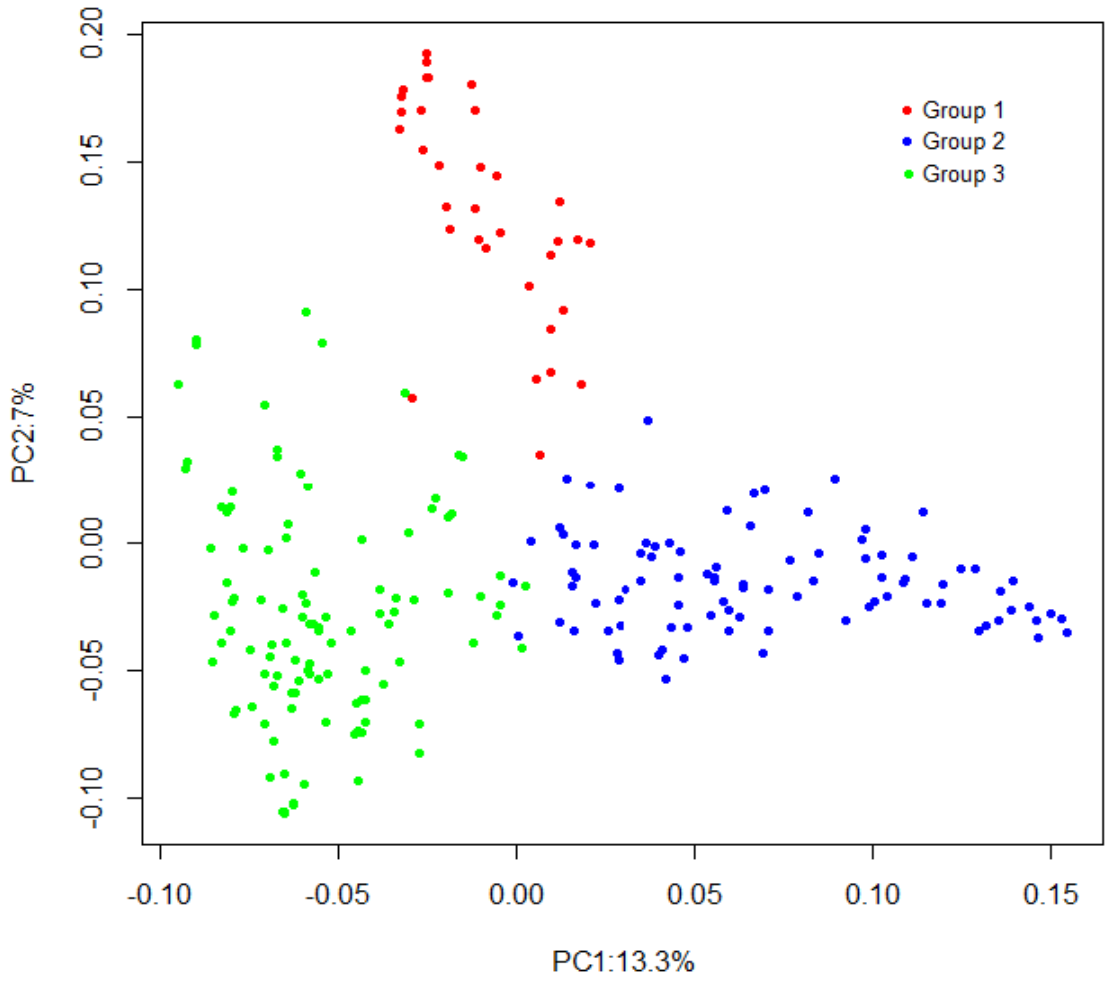



**Fig. 4-4.** First two principal components showing marker-based K-means clustering of the 231 wheat lines into three groups (Group one = 30 lines, group two = 93 lines, and group three = 108 lines).

In this study, we assessed three approaches to account for population structure; i) interaction model, ii) across-group model, and iii) stratified or within group model. The prediction accuracy of the interaction model was similar to an across-group model that ignored population structure (Fig. 4-5). On average, the interaction model consistently showed similar results to the across-group model in all three groups using both BRR and BayesB, which indicated that accounting for population structure did not improve prediction accuracy (Fig. 4-5). On the other hand, estimating marker effects separately within each group greatly reduced prediction accuracy in both methods. Most of the negative correlation in the stratified analysis could be the result of fitting nearly 18K SNPs on small number of phenotypes, which may have resulted in overfitting and poor predictive

performance. de los Campos et al. (2015) evaluated these methods using a wheat data set that had two clearly defined clusters and a pig data set with less marked differentiation between groups, similar to what we observed in our data. In both data sets, the interaction model was either the best performing method or was comparable to either the stratified or across-group analyses (de los Campos et al., 2015). In the pig data set, there was an overall advantage towards across-group analysis relative to stratified analysis, but the relative rankings of the three approaches varied between traits (de los Campos et al., 2015). Similarly, in the wheat data set differences between the stratified, across-group, and interaction models were small, but the rankings of these methods differed between environments (de los Campos et al., 2015). Overall, the method that accounted for population structure was not advantageous in this study.

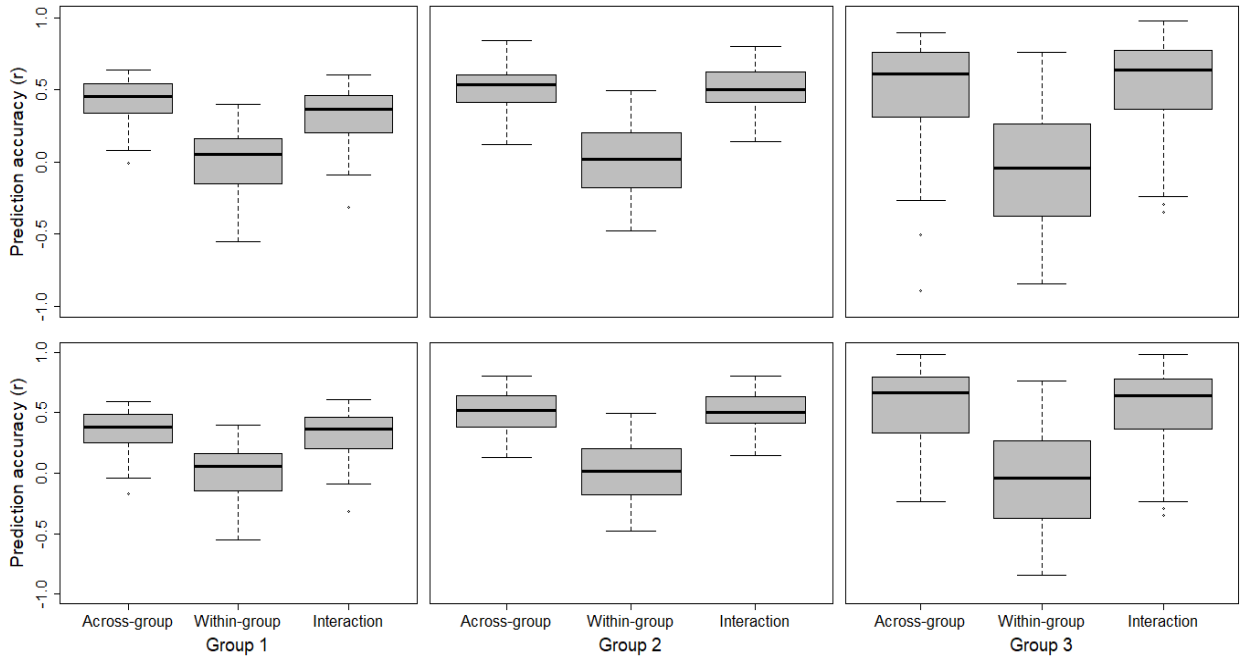



**Fig. 4-5.** Box plots showing prediction accuracy for grain yield estimated from 50 random training-validation partitions using Across-group, Within-group and Interaction BRR (top) and BayesB (bottom) models.

The results also showed that there was no difference between BRR and BayesB for all three approaches (Fig. 4-5). Similarly, de los Campos et al. (2015) reported very small differences between BRR and BayesB. BRR is the Bayesian counterpart of RR-BLUP, which includes all

markers and shrinks their effects uniformly, while BayesB considers most loci to have no effect on the trait and most markers are excluded from the model (Lorenz et al., 2011). In theory, BRR is expected to outperform BayesB for the prediction of traits, such as grain yield that are controlled by many loci with small effects, while BayesB is expected to perform well for traits controlled by few large effect QTL (Lorenz et al., 2011). The similar performance of the two methods in our study might be because of the absence of large effect QTL in the population. The other reason might be due to the small TP size used in this study. We used 231 wheat lines while de los Campos et al. (2015) used 599 and 3534 records in the wheat and pig data sets, respectively.

The effect of TP size on prediction accuracy was also observed in this study. There was an increase in the average prediction accuracy with increasing the TP size for the interaction and across-group models (Fig. 4-5). Groups one, two, and three had 30, 93 and 108 lines, respectively. In each training-validation partition, 24, 76 and 87 lines were used as TP and 6, 17 and 21 lines were used as a validation set for groups one, two, and three, respectively. Group three had the largest TP size and had the highest average prediction accuracies, which is consistent with the results above and with previous research.

### 4.4 Conclusion

In this study, we evaluated different parameters that can affect GS prediction accuracy. Increasing the TP size always improved the prediction accuracy, indicating that TP size is the most important factor that can affect genomic predictions. On the other hand, marker density did not affect prediction accuracy. This suggests that a reduced subset of evenly distributed markers across the genome can be sufficient to achieve optimal prediction accuracy. Correcting for population structure did not improve prediction accuracy. In this population, models that borrow information across subpopulations, either by considering constant marker effects or using interaction terms, are equally effective to account for population structure. However, the stratified analysis greatly reduced prediction accuracy and this method should not be considered in the future.

# 5. Modelling Genotype-by-Environment Interaction for Genomic Prediction in Wheat

## 5.1 Introduction

Genotype-by-environment interaction is an important issue when dealing with multi-environment plant breeding trials. Multi-environment plant breeding trials play an important role to assess G × E and genotype stability, and to predict the performance of untested lines (Burgueño et al., 2012). In multi-environment trials, G × E is expressed either as inconsistent responses of some genotypes relative to others due to genotypic rank change (crossover interaction) or as changes in the absolute differences between genotypes without rank changes (non-crossover interaction) (Crossa, 2012). Crossover G × E is most important in plant breeding because it affects selection decisions across environments (Crossa et al., 2004).

Genomic selection is a novel approach suggested to improve genetic gain of quantitative traits but is also affected by G × E interaction (Burgueño et al., 2012; Crossa et al., 2015; Jarquín et al., 2014a; Jarquín et al., 2017; Lopez-Cruz et al., 2015; Pérez-Rodríguez et al., 2015; Sukumaran et al., 2017). The effect of G × E on GS model prediction could be due to differential response of chromosome regions associated with phenotypic variation, its indirect effects on trait heritability, or both. The variance of G × E is part of the phenotypic variance and is included in the equation to estimate heritability of traits in a multi-environment trial.

Many of the studies on GS use data from a single environment or average LS-means across environments. However, in recent years, several GS models that incorporate G × E have been reported (Burgueño et al., 2012; Crossa et al., 2015; Dawson et al., 2013; Jarquín et al., 2014a; Jarquín et al., 2017; Lopez-Cruz et al., 2015; Pérez-Rodríguez et al., 2015; Sukumaran et al., 2017). Genotype-by-environment interaction can be incorporated into GS models using different techniques. Burgueño et al. (2011) used mixed-effect linear models that consider G × E as a random effect via factor analysis, while considering other factors such as sites as either fixed or random effects. In another approach, Heslot et al. (2014) extended factorial regression and developed a new machine learning approach that integrates ECs and crop models into the GS framework to account for G × E. Crop models are sets of equations developed by extensively studying the behavior of a few genotypes under a range of growing conditions and this information is used to reduce the daily weather variables to a few covariates per crop growth stage (Heslot et al., 2014). Similarly, Technow et al. (2015) used approximate Bayesian computation that allowed

them to incorporate crop growth models directly into the estimation of whole genome marker effects in GS. Jarquín et al. (2014a) proposed another approach to incorporate high-dimensional ECs into GS using a reaction norm model where the main and interaction effects of markers and environments or ECs were included in GS using covariance structures. Several other studies applied the reaction norm model for prediction of traits in wheat and cotton using genetic markers or pedigree relationships (Jarquín et al., 2017; Pérez-Rodríguez et al., 2015; Sukumaran et al., 2017). Recently, Lopez-Cruz et al. (2015) reported that G × E can be modelled using a M × E GS model when genomic and environmental covariate data are available. Crossa et al. (2015) reported an extension of the M × E model using priors that induce shrinkage (BRR) and variable selection (BayesB). These studies reported substantial increases in genomic prediction accuracy when multi-environment prediction models that incorporate G × E were used because these methods allow information to be borrowed across environments (Crossa et al., 2015; Jarquín et al., 2014a; Jarquín et al., 2017; Lopez-Cruz et al., 2015; Pérez-Rodríguez et al., 2015; Sukumaran et al., 2017). Thus, the objectives of this study were to i) examine GS prediction accuracy when modelling G × E, and ii) evaluate different methods of accounting for G × E when making genomic predictions.

## 5.2 Materials and Methods

The varcomp data set was used to assess the effects of modelling G × E on genomic prediction accuracy. All lines were genotyped using the wheat 90K SNP array, which generated 17,887 polymorphic SNP markers. Genomic predictions were made for grain yield and protein content. Detailed description of the phenotypic and genotypic data sets used in this study are provided in Chapter three.

### 5.2.1 Phenotypic data analysis

Phenotypic data in each environment were analyzed using ANOVA with SAS Mixed models v9.4 (SAS Institute Inc., 2015). Genotypes were considered as fixed effect, while replication and block nested in replication were considered random effects. The DDFM=Kr option was used for approximating the degrees of freedom for means. LS-means of lines in each environment were used for model prediction and validation. The phenotypic data analyses included all 100 lines in the varcomp dataset for a better estimate of (co)variances, but five lines with genotyping errors

and 14 lines from the CWAD wheat class were excluded and LS-means of the remaining 81 lines were used for this study.

### 5.2.2 Statistical models used to incorporate genotype-by-environment interaction in GS

The effect of modelling G × E on GS model prediction accuracy was assessed using two different approaches. The first approach involved modelling G × E using a M × E GS model as implemented in Lopez-Cruz et al. (2015). The second approach involved modelling G × E using a class of reaction norm models that incorporate the main and interaction effects of molecular markers and ECs using covariance functions as implemented in Jarquín et al. (2014a).

#### *5.2.2.1 Marker-by-environment interaction GS model*

In this approach, the performance of the M × E model was compared to a combined analysis (across-environment model) and a stratified (single-environment) model. These models were fitted using the G-BLUP model implemented in the R package BGLR, v1.0.4 (Pérez and de los Campos, 2014).

**i) M × E model:** For this model, marker effects are decomposed into components that are constant across environments (main effects) and components that are environment specific (interaction). This model allows information across environments to be borrowed, while permitting marker effects to vary across environments (Lopez-Cruz et al., 2015). The effects of the $k^{\text{th}}$ marker in the $j^{\text{th}}$ environment ($\beta_{jk}$) is a combination of the main effect ($b_{0k}$) common to all environments and an interaction term ($b_{jk}$) specific to the $j^{\text{th}}$ environment. The regression equation that fits the phenotype of the $i^{\text{th}}$ individual in $j^{\text{th}}$ environment is indicated by:

$$y_{ij} = \mu_j + \sum_{k=1}^{P} X_{ijk}\left(b_{0k} + b_{jk}\right) + \varepsilon_{ij} \tag{5.1}$$

where $i$ = 1,2,……,n refers to individuals; $j$ = 1,2,. . .,s refers to environments; $k$ = 1,2,...,p refers to markers; $\mu_j$ is an intercept; $X_{ijk}$ is a matrix of centered and standardized marker genotypes; $b_{0k}$ is a vector of marker effect common to all environments; $b_{jk}$ is an interaction term specific to the $j^{\text{th}}$ environment, and $\varepsilon_{ij}$ is a vector of model residuals.

The M × E model is similar to the interaction model used in Section 4.2.3 to model population structure in GS. When modelling population structure, marker effects were

decomposed between subpopulations, while in the M × E model the marker effects are decomposed between environments. The main difference between these approaches is that in the M × E model, individuals can have phenotypic records in all environments but when modelling population structure, individuals belong to only one cluster (de los Campos et al., 2015).

**ii) Across-environment model:** In this model, marker effects are assumed to be constant across environments (no interaction). This approach is equivalent to fitting a regression model using the averages of phenotypes from combined analysis of balanced data across environments (Lopez-Cruz et al., 2015). This model also allows information to be borrowed across environments, but unlike the M × E model, this is achieved by forcing marker effects to be constant across environments (Lopez-Cruz et al., 2015). The regression equation is indicated by:

$$y_{ij} = \mu_j + \sum_{k=1}^{P} X_{ijk} b_{0k} + \varepsilon_{ij} \quad (5.2)$$

where $i$ = 1,2,……,n refers to individuals; $j$ = 1,2,. . .,s refers to environments; $k$ = 1,2,...,p refers to markers; $\mu_j$ is an intercept; $X_{ijk}$ is a matrix of centered and standardized marker genotypes; $b_{0k}$ is a vector of constant marker effects across environments, and $\varepsilon_{ij}$ is a vector of model residuals.

**iii) Stratified or single-environment model:** In this model, phenotypes are regressed on markers separately in each environment using a linear model. This is a special case of the model in Equation 5.1 obtained by removing the main effects of markers, $b_0$= 0. The regression equation is indicated by:

$$y_{ij} = \mu_j + \sum_{k=1}^{P} X_{ijk} b_{jk} + \varepsilon_{ij} \quad (5.3)$$

where $i$ = 1,2,……,n refers to individuals; $j$ = 1,2,. . .,s refers to environments; $k$ = 1,2,...,p refers to markers; $\mu_j$ is an intercept; $X_{ijk}$ is a matrix of centered and standardized marker genotypes; $b_{jk}$ is a vector of marker effects specific to the $j^{th}$ environment, and $\varepsilon_{ij}$ is a vector of model residuals.

For each trait, the prediction accuracy of the three models was assessed using a total of 50 random training-validation partitions. Within each training-validation partition, 80% of the records were used as TP and the remaining 20% were used to assess the prediction accuracy. The same training-validation partitions were used to assess the accuracy of M × E, across-environment and

single-environment models. The average and the standard deviation of the 50 partitions were reported. Two cross-validation designs (CV1 and CV2) were tested to simulate real situations faced by plant breeders (Burgueño et al., 2012). Cross-validation one (CV1) involved predicting phenotypes of lines that have never been tested in any of the environments, mimicking newly developed lines. Cross-validation two (CV2) involved predicting phenotypes of lines that were evaluated in some environments but not in others, thus, mimicking incomplete field trials. For each cross-validation technique and training-validation partition, predictions were based on 50,000 iterations obtained after discarding 5,000 samples as a burn-in. Within each training-validation partition, prediction accuracy was assessed by calculating Pearson's correlation between the predicted values and phenotypes of lines in the validation set within environment.

#### *5.2.2.2 Reaction norm models that incorporate environmental covariates*

In addition to implementing the M × E model, we also tested a class of reaction norm models that were developed by Jarquín et al. (2014a) to account for G × E in GS. These models can be considered as an extension of the standard G-BLUP and can be interpreted as reaction norm models in which genetic and environmental factors are described using a linear regression on genetic markers and ECs (Jarquín et al., 2014a). The main effects of markers and ECs were modelled using techniques similar to the standard G-BLUP, while the interaction terms were included using a reaction norm model (Jarquín et al., 2014a).

Environmental covariate data were obtained from weather data for each environment. Data were obtained from a weather station located within 10 km of the experimental sites (Environment Canada, 2016). The crop growing season was divided into four months (May to August), and 13 ECs based on temperature, humidity and precipitation were obtained for each of the four months yielding a total of 52 distinct ECs for the entire growing season (Table 5-1). Environmental covariates that were not significantly correlated with each trait were removed and the remaining 43 and 48 ECs were included for the prediction of grain yield and protein content, respectively.

**Table 5-1.** Environmental covariate data used for modelling G × E using reaction norm models.

| Number | Variable name | Description |
|---|---|---|
| 1 | MinT | Minimum temperature (°C) |
| 2 | MaxT | Maximum temperature (°C) |
| 3 | AvgT | Average temperature (°C) |
| 4 | TCl20 | Number of hours with temperature < 20°C |
| 5 | TCg20 | Number of hours with temperature > 20°C |
| 6 | TCg25 | Number of hours with temperature > 25°C |
| 7 | TCg30 | Number of hours with temperature > 30°C |
| 8 | AvgRH | Average relative humidity (%) |
| 9 | RHl50 | Number of hours with RH < 50% |
| 10 | RHg50 | Number of hours with RH > 50% |
| 11 | RHg75 | Number of hours with RH > 75% |
| 12 | RHg90 | Number of hours with RH > 90% |
| 13 | TotPCP | Total precipitation (mm) |

The reaction norm model incorporates the main and interaction effects of molecular markers, environments and ECs using a covariance function. The covariance function is the cell-by-cell product of two relationship matrices, one based on markers (the GRM used in standard G-BLUP) and the other one based on ECs (Jarquín et al., 2014a). A brief description of these models is given below:

**i) Model 1 (EG):** This is a linear random effects model in which phenotypes ($y_{ij}$) are described as the sum of an overall mean ($\mu$) plus a random deviation due to the environment ($E_i$), which is a combination of site-years, plus the marker covariates using marker-derived GRM ($\mathrm{g}_j$) and a residual term ($\varepsilon_{ij}$). This model is equivalent to the standard G-BLUP with the addition of a random environmental effect (Jarquín et al., 2014a). The regression equation is indicated using the following formula:

$$y_{ij} = \mu + E_i + \mathrm{g}_j + \varepsilon_{ij} \tag{5.4}$$

with $E_i \sim N(0, \sigma_E^2)$, $\mathrm{g} \sim N\left(0, G\sigma_{\mathrm{g}}^2\right)$ and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$.

**ii) Model 2 (EGW):** This model is an extension of model 5.4 with the addition of ECs (W). The regression equation is given by:

$$y_{ij} = \mu + E_i + \mathrm{g}_j + w_{ij} + \varepsilon_{ij} \quad (5.5)$$

**iii) Model 3 (EG-G×E):** Models 5.4 and 5.5 only included the main effects of markers (G), Environment (E), and ECs without G × E. This model extends on model 5.4 by including an interaction term between the marker-based relationship matrix and environments (G × E). The regression equation is given by:

$$y_{ij} = \mu + E_i + \mathrm{g}_j + \mathrm{g}E_{ij} + \varepsilon_{ij} \quad (5.6)$$

**iv) Model 4 (EGW-G×W):** This model is an extension of model 5.5 that includes a random effect that represents interactions between the marker-based relationship matrix and the ECs (G × W). This is achieved using covariance structures. The equation of this model is given by:

$$y_{ij} = \mu + E_i + \mathrm{g}_j + w_{ij} + \mathrm{g}w_{ij} + \varepsilon_{ij} \quad (5.7)$$

**v) Model 5 (EGW-G×E):** This model is similar to model 5.7; except in this case, the interaction term is between the effects of markers and environments. Since the ECs may not fully explain the differences in environmental conditions, some portion of G × E may not be accounted for when modelling the interaction between markers and ECs (Jarquín et al., 2014a). An alternative way of modelling G × E is to include an interaction term between environments and the random effect of markers. The regression equation is indicated by:

$$y_{ij} = \mu + E_i + \mathrm{g}_j + w_{ij} + \mathrm{g}E_{ij} + \varepsilon_{ij} \quad (5.8)$$

**vi) Model 6 (EGW-G×WG×E):** This is the most comprehensive model that includes the main effects of markers, environments and ECs plus the interaction of markers with environments and ECs. The regression equation is given by:

$$y_{ij} = \mu + E_i + \mathrm{g}_j + w_{ij} + \mathrm{g}w_{ij} + \mathrm{gE}_{ij} + \varepsilon_{ij} \quad (5.9)$$

All models were fitted using the R package BGLR, v1.0.4 (Pérez and de los Campos, 2014). The prediction accuracy of these models was assessed using the two cross-validation designs (CV1 and CV2) described earlier. For each trait, CV1 and CV2 were implemented in a fivefold cross-validation design. Inferences for all models were based on 50,000 iterations obtained after discarding 5,000 samples as a burn-in. Prediction accuracy was assessed by calculating Pearson's correlation between the predicted values and phenotypes of lines in the validation set within each environment and fold.

## 5.3 Results and Discussion

### 5.3.1 Large phenotypic variations were observed among environments

Analysis of data in each environment showed that there were highly significant differences among lines for both grain yield and protein content ($P < 0.001$). Phenotypic correlations among the seven environments are indicated in Table 5-2. Overall, the pair-wise environments had strong positive correlations for both traits, except for grain yield where correlations were slightly lower for the environment pairs SWC-2012/Kernen-2012 ($r = 0.43$) and SWC-2012/SWC-2014 ($r = 0.38$) (Table 5-2). Correlations of grain yield among environments ranged from 0.38 to 0.72 while correlations of protein content ranged from 0.52 to 0.86. This shows that there was G × E for both traits. The average grain yield and protein content from two replications were assessed in each environment. For both traits, there was variability among the environments. The average grain yield was low in Kernen-2012 (3015 kg $ha^{-1}$) and SWC-2012 (3789 kg $ha^{-1}$), was moderate in SWC-2014 (4200 kg $ha^{-1}$), high in Kernen-2011, Kernen-2013 and Kernen-2014 (ranged between 4698 and 4784 kg $ha^{-1}$), and very high in SWC-2013 (5060 kg $ha^{-1}$) (Fig. 5-1A). Average grain protein content was the highest in Kernen-2012 and SWC-2014 (15.7 and 15.3%, respectively), was lower in Kernen-2013, Kernen-2014, SWC-2012 and SWC-2013 (range of 13.1 to 14.1%), and lowest in Kernen-2011 (12.6%) (Fig. 5-1B).

**Table 5-2.** Phenotypic correlations across seven environments for grain yield (upper diagonal) and protein content (lower diagonal) of 81 lines.

| Environment† | Kernen-2011 | Kernen-2012 | Kernen-2013 | Kernen-2014 | SWC-2012 | SWC-2013 | SWC-2014 |
|---|---|---|---|---|---|---|---|
| Kernen-2011 | – | 0.59 | 0.65 | 0.61 | 0.68 | 0.67 | 0.53 |
| Kernen-2012 | 0.74 | – | 0.68 | 0.49 | 0.43 | 0.61 | 0.72 |
| Kernen-2013 | 0.81 | 0.79 | – | 0.58 | 0.54 | 0.71 | 0.67 |
| Kernen-2014 | 0.66 | 0.68 | 0.67 | – | 0.55 | 0.65 | 0.53 |
| SWC-2012 | 0.80 | 0.70 | 0.79 | 0.61 | – | 0.61 | 0.38 |
| SWC-2013 | 0.60 | 0.59 | 0.66 | 0.53 | 0.61 | – | 0.70 |
| SWC-2014 | 0.86 | 0.86 | 0.80 | 0.68 | 0.83 | 0.67 | – |

†SWC: Swift Current.

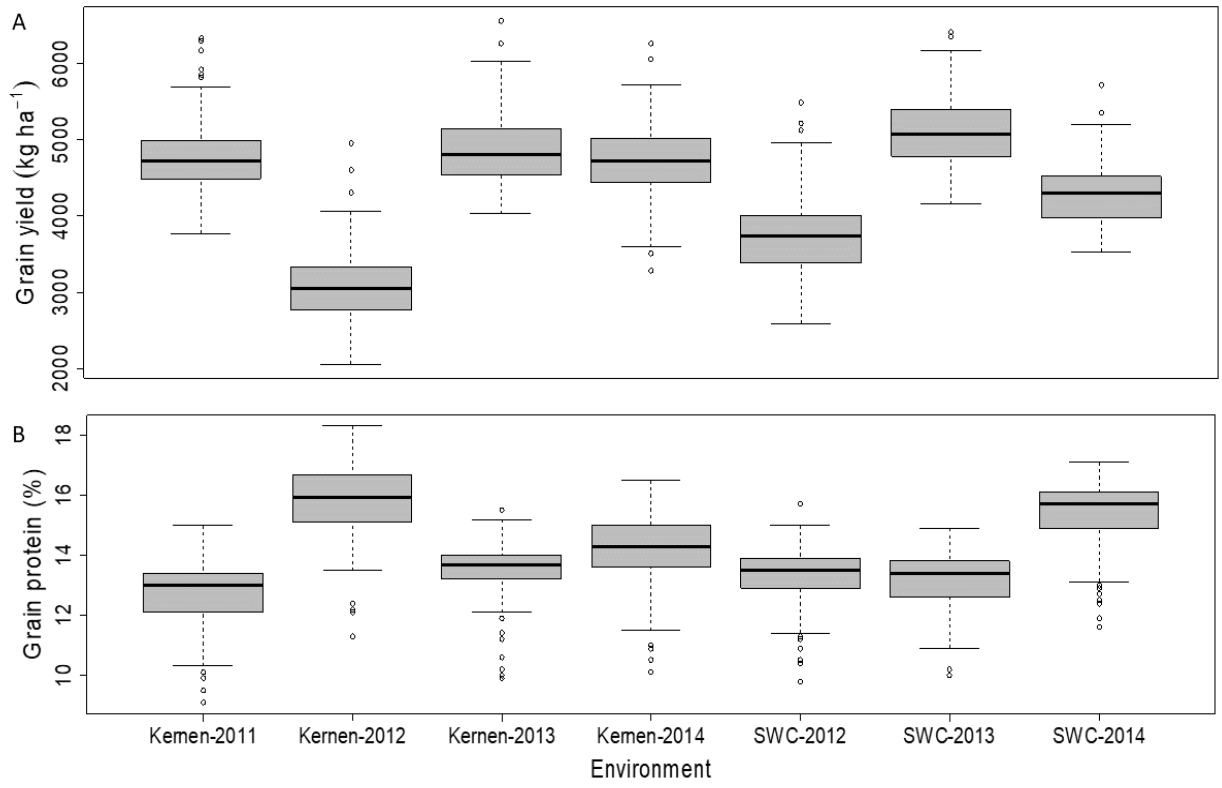



**Fig. 5-1.** Boxplots for (A) grain yield and (B) grain protein content of 81 lines across seven environments. Data within each environment are averages of two replications. SWC: Swift Current.

### 5.3.2 Genomic predictions using the marker-by-environment interaction model

The prediction accuracy of the M × E model showed no consistent improvement compared to the single-environment and across-environment models when assessing grain yield. The results indicated that prediction accuracy for grain yield was generally lower for the M × E model compared to the single-environment model in CV1, but it was 24.8 to 62.8% higher than the single-environment model in CV2 (Table 5-3). However, the M × E model showed slightly higher or similar prediction accuracy compared to the across-environment model in both cross-validation designs; this indicated that there was no benefit of modelling G × E for grain yield (Table 5-3). In a similar study, Dawson et al. (2013) also reported no difference in accuracy between models that accounted for G × E and global models that ignored G × E for the prediction of wheat grain yield based on unbalanced historic data over a period of 17 years. Similarly, no significant improvements in accuracy were observed for the M × E model when predicting grain protein content. The M × E model gave consistently lower prediction accuracy than the single-environment model for prediction of grain protein content in CV1, but the results were variable compared to the across-environment model (Table 5-4). In CV1, the M × E model showed 8.9% higher accuracy compared to the across-environment model in Kernen-2012, while it was either lower or similar to the across-environment model in the other environments (Table 5-4). However, in CV2 the accuracy of the M × E model was 12.7 to 29.4% higher compared to the single-environment model, but the results were similar to the across-environment model (Table 5-4).

For both traits, the benefit of M × E over the single-environment model varied based on the cross-validation design used, but the performance of the M × E model was similar to the across-environment model that ignored G × E. Generally, the single-environment model gave better prediction for both traits in CV1 compared to the M × E and across-environment models in all environments, except for grain yield in Kernen-2014 where the results were comparable for all three models (Tables 5-3 and 5-4). However, both the across-environment and the M × E models gave substantially higher prediction accuracy than the single-environment model in CV2. This was because both models allow borrowing of information for the same line across environments, although the mechanisms of achieving this are different. In the across-environment model, the borrowing of information is achieved by considering marker effects to be constant across environments, while in M × E model marker effects are decomposed into components that are constant across environments and others that are environment-specific (Lopez-Cruz et al., 2015).

Crossa et al. (2014) also reported that modelling G × E in GS improves prediction accuracy by borrowing information from correlated environments. The benefit of a multi-environment model over single-environment models comes from modelling genetic correlations between environments (Burgueño et al., 2012). The M × E model is based on the variance component estimation of marker main effects and effects that are environment-specific (Crossa et al., 2015; Lopez-Cruz et al., 2015). The proportion of genomic variance explained by the main effects of markers was shown to have a direct relationship with the phenotypic correlation between environments; therefore, the M × E model performs best for prediction of traits in a set of environments that have positive and similar correlation (Crossa et al., 2015; Lopez-Cruz et al., 2015). Crossa et al. (2015) showed that the M × E model gave higher prediction accuracy for traits that had positive correlations between environments (heading date and thousand-kernel weight) than for traits that had close to zero or negative correlations between environments (grain yield and test weight). In this study, the observed phenotypic correlations among the environments were high for both traits (Table 5-2). Although the prediction accuracy of the M × E model was substantially higher than the single-environment model in CV2, it had almost similar performance with the across-environment model for both traits suggesting that there was no benefit of modelling G × E in this dataset.

**Table 5-3.** Average and standard deviation of prediction accuracy based on Pearson's correlation between predicted and actual grain yield. Genomic predictions were made using 50 training-validation partitions and two cross-validation techniques (CV1 and CV2) across seven environments.

| Grain yield | | | | |
|---|---|---|---|---|
| **Environment†** | **Single-environment** | **Across-environment** | **M × E** | **% Change‡ SE; AE** |
| | **CV1** | | | |
| Kernen-2011 | 0.596(0.205) | 0.548(0.231) | 0.564(0.220) | -5.4%; 2.9% |
| Kernen-2012 | 0.596(0.164) | 0.528(0.196) | 0.553(0.181) | -7.3%; 4.7% |
| Kernen-2013 | 0.478(0.262) | 0.465(0.255) | 0.462(0.262) | -3.2%; -0.6% |
| Kernen-2014 | 0.515(0.169) | 0.523(0.226) | 0.519(0.183) | 0.8%; -0.8% |
| SWC-2012 | 0.454(0.267) | 0.399(0.250) | 0.413(0.261) | -9.0%; 3.6% |
| SWC-2013 | 0.563(0.199) | 0.531(0.226) | 0.534(0.219) | -5.1%; 0.5% |
| SWC-2014 | 0.507(0.180) | 0.489(0.198) | 0.498(0.188) | -1.8%; 1.8% |
| | **CV2** | | | |
| Kernen-2011 | 0.659(0.190) | 0.829(0.090) | 0.822(0.097) | 24.8%; -0.8% |
| Kernen-2012 | 0.587(0.160) | 0.721(0.126) | 0.731(0.127) | 24.4%; 1.3% |
| Kernen-2013 | 0.527(0.243) | 0.851(0.100) | 0.857(0.092) | 62.8%; 0.7% |
| Kernen-2014 | 0.540(0.178) | 0.737(0.104) | 0.748(0.103) | 38.5%; 1.5% |
| SWC-2012 | 0.445(0.269) | 0.640(0.166) | 0.640(0.165) | 43.9%; 0.0% |
| SWC-2013 | 0.544(0.206) | 0.843(0.089) | 0.846(0.084) | 55.4%; 0.4% |
| SWC-2014 | 0.486(0.173) | 0.739(0.111) | 0.757(0.109) | 55.7%; 2.4% |

†SWC: Swift Current.

‡Change in prediction accuracy of the M × E model compared to the prediction accuracy of the single-environment (SE) and the across-environment (AE) models.

**Table 5-4.** Average and standard deviation of prediction accuracy based on Pearson's correlation between predicted and actual grain protein. Genomic predictions were made using 50 training-validation partitions and two cross-validation techniques (CV1 and CV2) across seven environments.

| Grain protein | | | | |
|---|---|---|---|---|
| **Environment†** | **Single-environment** | **Across-environment** | **M × E** | **% Change‡ SE; AE** |
| | **CV1** | | | |
| Kernen-2011 | 0.629(0.274) | 0.627(0.292) | 0.610(0.276) | -2.9%; -2.7% |
| Kernen-2012 | 0.737(0.168) | 0.638(0.211) | 0.695(0.179) | -5.7%; 8.9% |
| Kernen-2013 | 0.669(0.269) | 0.634(0.267) | 0.642(0.271) | -3.9%; 1.3% |
| Kernen-2014 | 0.583(0.290) | 0.591(0.284) | 0.567(0.299) | -2.8%; -4.1% |
| SWC-2012 | 0.618(0.232) | 0.617(0.262) | 0.586(0.251) | -5.1%; -5.1% |
| SWC-2013 | 0.619(0.224) | 0.585(0.214) | 0.576(0.217) | -6.9%; -1.5% |
| SWC-2014 | 0.750(0.186) | 0.723(0.171) | 0.724(0.176) | -3.4%; 0.2% |
| | **CV2** | | | |
| Kernen-2011 | 0.642(0.224) | 0.835(0.099) | 0.831(0.099) | 29.4%; -0.5% |
| Kernen-2012 | 0.781(0.131) | 0.862(0.091) | 0.880(0.077) | 12.7%; 2.1% |
| Kernen-2013 | 0.789(0.186) | 0.917(0.076) | 0.932(0.062) | 18.0%; 1.6% |
| Kernen-2014 | 0.671(0.209) | 0.815(0.126) | 0.804(0.135) | 19.7%; -1.3% |
| SWC-2012 | 0.666(0.186) | 0.835(0.085) | 0.817(0.096) | 22.6%; -2.2% |
| SWC-2013 | 0.643(0.165) | 0.797(0.122) | 0.801(0.119) | 24.6%; 0.6% |
| SWC-2014 | 0.779(0.157) | 0.910(0.068) | 0.908(0.071) | 16.7%; -0.2% |

†SWC: Swift Current.

‡Change in prediction accuracy of the M × E model compared to the prediction accuracy of the single-environment (SE) and the across-environment (AE) models.

Prediction accuracy also varied depending on the cross-validation design. Prediction accuracy for grain yield was generally lower in CV1 compared to the accuracy in CV2 except for Kernen-2012, SWC-2012, SWC-2013 and SWC-2014 where the accuracy of the single-environment model was slightly lower in CV2 (Table 5-3). But for grain protein content, the accuracy in CV2 was substantially higher than the accuracy obtained in CV1 for all three methods (Table 5-4). These results agree with Burgueño et al. (2012) who reported 31, 17.5 and 21.8% higher accuracy in CV2 compared to CV1 for models based on pedigree, marker, and pedigree plus marker, respectively. Crossa et al. (2015) also reported that all predictions in CV1 were lower than in CV2 for grain yield and heading date. This is because CV2 uses information within lines across environments, while this is not possible in CV1 because the lines have not been evaluated in any environment (Burgueño et al., 2012).

Similar GS studies in wheat indicated that the performance of the GS models is highly variable and is influenced by the environment and populations used. Based on three extensive bread wheat data sets, Lopez-Cruz, et al. (2015) reported that the accuracy of the M × E model was substantially higher than that of the across-environment model, but it was either similar or greater than the single-environment model depending on the cross-validation design and correlation between environments. Lopez-Cruz et al. (2015) suggested that the M × E model is more effective when applied to subsets of environments that have positive and similar correlations. The single-environment and the M × E models performed similarly in CV1, but the M × E gave higher prediction accuracy in CV2; however, in both cross-validation designs the across-environment model was the worst performing method (Lopez-Cruz et al., 2015). Crossa et al. (2015) extended the M × E method using BRR and BayesB in a multi-parental durum wheat population characterized for grain yield, test weight, thousand-kernel weight and heading date. The reported accuracies were variable for each trait, cross-validation design and environment, but overall the M × E or single-environment models performed better for all traits in CV1 while the M × E and across-environment models were better than the single-environment model in CV2. BayesB generally gave higher prediction accuracy than BRR for most model-trait combinations (Crossa et al., 2015). Some of the differences between the results obtained in this study and the results of Lopez-Cruz et al. (2015) and Crossa et al. (2015) could be due to differences in environments, TP size, or other characteristics of the populations used in these studies. Collectively, the results of this study showed that the performance of the evaluated models was

variable based on the cross-validation design, but overall, there was no benefit of modelling G × E using the M × E model in this population.

### 5.3.3 Genomic predictions using reaction norm models

We evaluated six different reaction norm models for their prediction accuracies. The Model EG included only the main effects of environments and markers without interaction terms. Similarly, model EGW included only the main effects of environments, markers and ECs without interactions, while the other four models included interaction terms of G × E, G × W or both G × W and G × E in addition to the main effects. The results indicated that adding the main effects of the ECs to the main effects of markers and environments did not improve prediction accuracy of both grain yield and protein content (Tables 5-5 and 5-6). Similar results were reported by Pérez-Rodríguez et al. (2015), where adding the main effects of ECs resulted in no change in accuracy because ECs did not vary within the environment. Moreover, adding the interaction terms did not improve prediction accuracy in both cross-validation designs, except in CV1 for grain protein where the most comprehensive model (EGW-G×WG×E) gave 2 to 11% higher prediction accuracy when compared to models that included only the main effects (Tables 5-5 and 5-6). This suggests that most of the variability within environments was explained by the main effects of markers and environments.

Previous studies in wheat reported substantial increases in genomic prediction accuracy using the reaction norm model that included G × E (Jarquín et al., 2014a; Jarquín et al., 2017; Sukumaran et al., 2017). Jarquín et al. (2014a) reported that adding interaction terms between markers, environments and ECs resulted in 35% increase in accuracy when compared to a model that accounted only for the main effects of these terms. Recently, Jarquín et al. (2017) showed that a model that included G × E terms resulted in 16 to 82% higher accuracy when compared to a model that accounted for the main effects of markers and environments. Similarly, Sukumaran et al. (2017) obtained the highest prediction accuracy when G × E was included in the model. In cotton, Pérez-Rodríguez et al. (2015) reported up to 2.7-fold increase in prediction accuracy when interaction terms were included in the model that was based only on the main effects. In our study, including the interaction terms did not have a large effect on prediction accuracy because a high accuracy was obtained using the model that included only the main effects of markers and environments. Despite the small TP size in this study, the overall prediction accuracies were higher

**Table 5-5.** Average and standard deviation of prediction accuracy (from fivefold cross-validation) based on Pearson's correlation between predicted and actual grain yield. Genomic predictions were made using six statistical models and two cross-validation techniques (CV1 and CV2) across seven environments.

| | Grain yield | | | | | |
|---|---|---|---|---|---|---|
| **Environment†** | **EG‡** | **EGW** | **EG-G×E** | **EGW-G×E** | **EGW-G×W** | **EGW-G×WG×E** |
| | **CV1** | | | | | |
| Kernen-2011 | 0.516(0.233) | 0.516(0.235) | 0.519(0.231) | 0.516(0.235) | 0.518(0.233) | 0.519(0.228) |
| Kernen-2012 | 0.540(0.181) | 0.540(0.185) | 0.542(0.178) | 0.537(0.182) | 0.543(0.179) | 0.536(0.183) |
| Kernen-2013 | 0.439(0.306) | 0.434(0.307) | 0.436(0.303) | 0.437(0.307) | 0.438(0.301) | 0.437(0.306) |
| Kernen-2014 | 0.487(0.205) | 0.485(0.203) | 0.484(0.204) | 0.484(0.201) | 0.487(0.200) | 0.485(0.204) |
| SWC-2012 | 0.372(0.299) | 0.370(0.303) | 0.373(0.296) | 0.371(0.303) | 0.373(0.298) | 0.374(0.300) |
| SWC-2013 | 0.491(0.261) | 0.490(0.262) | 0.491(0.259) | 0.488(0.264) | 0.498(0.259) | 0.492(0.258) |
| SWC-2014 | 0.402(0.261) | 0.404(0.259) | 0.407(0.258) | 0.400(0.262) | 0.413(0.257) | 0.404(0.261) |
| | **CV2** | | | | | |
| Kernen-2011 | 0.843(0.059) | 0.843(0.060) | 0.842(0.059) | 0.840(0.059) | 0.843(0.059) | 0.842(0.059) |
| Kernen-2012 | 0.739(0.144) | 0.739(0.143) | 0.739(0.144) | 0.739(0.144) | 0.739(0.144) | 0.738(0.144) |
| Kernen-2013 | 0.847(0.050) | 0.848(0.050) | 0.847(0.049) | 0.847(0.050) | 0.847(0.050) | 0.848(0.050) |
| Kernen-2014 | 0.773(0.105) | 0.773(0.104) | 0.772(0.105) | 0.771(0.105) | 0.772(0.105) | 0.772(0.105) |
| SWC-2012 | 0.708(0.065) | 0.708(0.066) | 0.709(0.066) | 0.709(0.068) | 0.709(0.067) | 0.709(0.067) |
| SWC-2013 | 0.808(0.100) | 0.808(0.101) | 0.809(0.099) | 0.808(0.100) | 0.808(0.100) | 0.807(0.101) |
| SWC-2014 | 0.705(0.104) | 0.705(0.103) | 0.705(0.105) | 0.704(0.104) | 0.704(0.103) | 0.706(0.103) |

† SWC: Swift Current.

‡E: main effect of environments, G: main effect of marker covariates, W: main effect of environmental covariates (ECs), G×E: interactions between markers and environments, G×W: interactions between markers and ECs.

**Table 5-6.** Average and standard deviation of prediction accuracy (from fivefold cross-validation) based on Pearson's correlation between predicted and actual grain protein. Genomic predictions were made using six statistical models and two cross-validation techniques (CV1 and CV2) across seven environments.

| | Grain protein | | | | | |
|---|---|---|---|---|---|---|
| **Environment†** | **EG‡** | **EGW** | **EG-G×E** | **EGW-G×E** | **EGW-G×W** | **EGW-G×WG×E** |
| | | | **CV1** | | | |
| Kernen-2011 | 0.710(0.212) | 0.711(0.208) | 0.708(0.212) | 0.709(0.210) | 0.711(0.209) | 0.791(0.185) |
| Kernen-2012 | 0.758(0.079) | 0.760(0.076) | 0.755(0.081) | 0.756(0.080) | 0.759(0.082) | 0.774(0.085) |
| Kernen-2013 | 0.784(0.128) | 0.782(0.127) | 0.780(0.130) | 0.781(0.126) | 0.782(0.129) | 0.842(0.118) |
| Kernen-2014 | 0.771(0.076) | 0.771(0.075) | 0.770(0.073) | 0.769(0.074) | 0.771(0.076) | 0.785(0.053) |
| SWC-2012 | 0.697(0.162) | 0.696(0.160) | 0.694(0.164) | 0.695(0.158) | 0.697(0.162) | 0.751(0.097) |
| SWC-2013 | 0.666(0.138) | 0.661(0.140) | 0.662(0.143) | 0.661(0.139) | 0.662(0.139) | 0.729(0.109) |
| SWC-2014 | 0.793(0.090) | 0.793(0.087) | 0.790(0.093) | 0.793(0.091) | 0.793(0.088) | 0.813(0.098) |
| | | | **CV2** | | | |
| Kernen-2011 | 0.846(0.102) | 0.846(0.102) | 0.847(0.102) | 0.846(0.103) | 0.847(0.102) | 0.846(0.102) |
| Kernen-2012 | 0.862(0.081) | 0.862(0.081) | 0.861(0.081) | 0.861(0.081) | 0.861(0.081) | 0.861(0.081) |
| Kernen-2013 | 0.910(0.055) | 0.910(0.055) | 0.910(0.056) | 0.910(0.056) | 0.911(0.055) | 0.910(0.056) |
| Kernen-2014 | 0.833(0.082) | 0.832(0.082) | 0.832(0.083) | 0.832(0.083) | 0.832(0.082) | 0.832(0.083) |
| SWC-2012 | 0.874(0.037) | 0.875(0.038) | 0.874(0.037) | 0.874(0.037) | 0.874(0.037) | 0.874(0.037) |
| SWC-2013 | 0.627(0.345) | 0.627(0.344) | 0.627(0.345) | 0.628(0.341) | 0.626(0.345) | 0.628(0.342) |
| SWC-2014 | 0.877(0.082) | 0.876(0.082) | 0.877(0.081) | 0.876(0.082) | 0.876(0.082) | 0.876(0.082) |

† SWC: Swift Current, SK.

‡E: main effect of environments, G: main effect of marker covariates, W: main effect of environmental covariates (ECs), G×E: interactions between markers and environments, G×W: interactions between markers and ECs.

compared to the previously reported values in wheat (Jarquín et al., 2014a; Jarquín et al., 2017; Sukumaran et al., 2017). This could be because we used data generated from a balanced experimental design across two locations and four years for a total of seven environments, while the previous studies were based on diverse multi-environment trials. Jarquín et al. (2014a) implemented these models to predict grain yield for 139 wheat lines evaluated in 134 locations and eight years for a total of 340 environments. The lines used in Jarquín et al. (2014a) were obtained from different commercial breeding programs and had weaker genetic relationships. Moreover, Jarquín et al. (2014a) used data from a single plot, while we used the mean of two replications in each environment. Similarly, Jarquín et al. (2017) used a highly unbalanced historical data from 31 environments. Sukumaran et al. (2017) also used data generated from 136 environments across eight countries. Moreover, Sukumaran et al. (2017) tested the reaction norm models using pedigrees instead of molecular markers. Previous studies showed that GS models that include molecular markers have more accurate predictions compared to pedigree-based models (Arruda et al., 2015; Crossa et al., 2010). Overall, interaction terms typically improve prediction accuracies, but in our study the main effects of marker and environments were sufficient to achieve higher accuracies and there was no real benefit of including the interaction terms.

The two methods tested for cross-validation also differed in their prediction accuracies. Cross-validation one consistently gave lower prediction accuracy (ranging from 0.37 to 0.54) compared to CV2 (ranging from 0.70 to 0.84) for grain yield (Table 5-5). Similar results were observed for grain protein content, except for SWC-2013 where prediction in CV2 was lower compared to CV1 for all models (Table 5-6). This has been observed in several studies that used similar cross-validation designs (Burgueño et al., 2012; Crossa et al., 2015; Jarquín et al., 2014a; Jarquín et al., 2017; Lopez-Cruz et al., 2015; Pérez-Rodríguez et al., 2015). This indicates that including information from correlated environments is important to improve prediction accuracy.

### 5.3.4 Comparing the accuracy of marker-by-environment interaction and reaction norm models

The overall performance of the M × E model compared to the reaction norm models appear to be inconsistent and varied based on the trait and cross-validation design. Modelling G × E using the M × E model gave a slightly higher prediction accuracy compared to modelling G × E using the reaction norm models for grain yield in CV1 (Tables 5-3 and 5-5). However, the single-

environment model was superior to all other models for grain yield (Table 5-3). In CV2, the results were inconsistent when predictions were made for grain yield. In three environments (Kernen-2013, SWC-2013 and SWC-2014) the M × E model slightly outperformed the reaction norm models, while in three other environments (Kernen-2011, Kernen-2014 and SWC-2012) the reaction norm models were better, but results were similar using both approaches in Kernen-2012 (Tables 5-3 and 5-5). For grain protein, the reaction norm models gave substantially higher prediction accuracy compared to the M × E, single-environment and across-environment models using CV1 in all environments (Tables 5-4 and 5-6). In CV2, the M × E was better in four environments (Kernen-2012, Kernen-2013, SWC-2013 and SWC-2014), while the reaction norm models were better in the other three environments (Kernen-2011, Kernen-2014 and SWC-2012). Overall, the prediction accuracy of the different models that accounted for G × E varied across the two traits and cross-validation designs, indicating that model selection is an important process for GS when choosing to incorporate G × E.

## 5.4 Conclusion

In this study, we evaluated M × E and reaction norm GS models to account for G × E when making predictions for grain yield and protein content. For both traits, the single-environment model was either the best or similar to the other best model in CV1, while in CV2 the M × E and across-environment models were considerably better than the single-environment model. However, comparable results were obtained between the M × E and across-environment models in CV2. In the reaction norm models, adding the interaction terms of either G × E, G × W or both G × W and G × E did not improve prediction accuracy. This indicates that there was no advantage of modelling G × E in GS using both approaches in the population used in this study.

# 6. Genome-Wide Association Mapping of Agronomically Important Traits in Wheat

## 6.1 Introduction

Molecular plant breeding tools are used to identify genes controlling complex traits. The two most commonly used tools for dissecting complex traits in plants are linkage analysis (QTL mapping) and association mapping, also known as LD mapping (Zhu et al., 2008). Linkage disequilibrium is the non-random association of alleles at two genetic loci, which in random mating populations is mostly generated by mutation and genetic drift, and decays by recombination (Breseghello and Sorrells, 2006a). Linkage analysis with experimental populations derived from bi-parental crosses detects QTL that are segregating in the population and provides relevant information about traits that are specific to the same or genetically related populations, while results from association mapping are relevant to a wider germplasm base (Zhu et al., 2008).

Association mapping is a method that relies on gametic phase disequilibrium to study the relationship between phenotypic variation and genetic polymorphisms across a set of germplasms (Flint-Garcia et al., 2003). The main principle of this method is that LD tends to be maintained over many generations between loci that are genetically linked to one another (Neumann et al., 2011). Unlike the classical linkage analysis that uses experimental (bi-parental) mapping populations, association mapping evaluates genetic diversity across natural populations to identify polymorphisms that correlate with phenotypic variation (Flint-Garcia et al., 2003). Association mapping is further classified into two categories: 1) candidate gene association mapping, and 2) genome-wide association mapping (GWAM). Candidate gene association mapping tests the relation between DNA polymorphism of a candidate gene with the trait of interest, while GWAM surveys the whole genome for casual genetic variations using dense genome-wide markers (Rafalski, 2002; Zhu et al., 2008). The advantage of association mapping over linkage analysis is that it uses a more diverse population, and therefore examines a broader set of genetic variation for marker-trait correlations; this results in enhanced mapping resolution and broader allele coverage (Abdurakhmonov and Abdukarimov, 2008; Neumann et al., 2011; Zhu et al., 2008). Association mapping panels can also be used to study several traits within a breeding program and can save time and money since there is no need to develop a biparental mapping population for each trait (Abdurakhmonov and Abdukarimov, 2008; Neumann et al., 2011; Zhu et al., 2008). In plant breeding, germplasm bank collections, elite breeding materials, and synthetic or specialized

populations are used for association mapping (Breseghello and Sorrells, 2006a). Association mapping with pedigree-based elite breeding material is likely to identify superior alleles that have been captured by breeding practices and facilitates MAS (Zhu et al., 2008). Quantitative trait loci detected from elite breeding materials are relevant for a breeding program and can be directly utilized for MAS. However, in elite breeding materials, population structure and cryptic relationships between lines can lead to spurious associations (Pritchard and Rosenberg, 1999). Methods that account for population structure and genetic relatedness must be used to minimize false positive associations between phenotypes and unlinked loci.

Linkage analysis and association mapping are applied in plant breeding to identify moderate to large effect QTL underlying trait variation that are then used in MAS. In recent years, genomic selection has become an alternative to standard MAS (Meuwissen et al., 2001). In GS, genome-wide SNP variation is modelled without identifying loci or their association with the phenotype (Cobb et al., 2013). Genomic selection and GWAS can be performed on the same population and highly significant SNPs identified from GWAS can be fitted as fixed effects in GS models without shrinking their effects (Begum et al., 2015; Rutkoski et al., 2014; Spindel et al., 2016; Spindel et al., 2015). Thus, the objective of this study was to detect marker-trait associations in spring bread wheat for nine agronomic and end-use quality traits.

## 6.2 Materials and Methods

### 6.2.1 Plant material, phenotypic and genotypic data

This study used a mapping population of 231 spring bread wheat varieties and advanced breeding lines (Appendix A). The combined phenotypic data across all environments were used in this study. Detailed description of the plant materials and phenotypic data are provided in Section 3.2.1. For all traits, extreme values that fell beyond 1.5 times the interquartile range of a box plot were excluded from the association analysis. A total of 17,887 polymorphic SNPs with CF higher than 90% and MAF greater than 10% were generated for this population as described in Section 3.2.2. The chromosomal locations of the SNPs were determined based on the hexaploid wheat consensus map (Wang et al., 2014a). SNPs that did not have position (12.89%) and that mapped to multiple positions on the consensus map (1.86%) were excluded. Redundant markers that had squared pairwise correlations greater than 0.99 (42.94%) and co-segregating SNPs (27.02%) were removed. The remaining 2,735 SNPs were used for association analyses.

### 6.2.2 Population structure and LD analyses

The population structure of the 231 lines was evaluated using distance based hierarchical clustering, model-based Bayesian clustering, and principal component analysis (PCA). A dissimilarity matrix was calculated from the marker data based on the Euclidean distance using the function 'dist' in R (R Development Core Team, 2016). Hierarchical clustering was applied to the Euclidean distance matrix based on Ward's criterion (ward.D2) using the function 'hclust' in R (R Development Core Team, 2016). A model-based Bayesian clustering was conducted using the program STRUCTURE v2.3.4 (Pritchard et al., 2000), based on a selected subset of 581 weakly correlated markers (markers having squared pairwise correlations smaller than 0.1). Markov chain Monte Carlo cycles were repeated 100,000 times after 10,000 cycles of a burn-in period. The default setting of admixture model and correlated allele frequencies was tested with the number of subpopulations (*K*) from one to twelve. Each test included twenty independent runs. Optimal *K* was estimated based on the $\Delta K$ that is the rate of change in the log likelihood of data between consecutive *K* values. $\Delta K$ was estimated using STRUCTURE HARVESTER, v0.6.94 (Earl and vonHoldt, 2012). Data from the twenty independent runs were integrated using the *FullSearch* algorithm in CLUMPP, v1.1.2, which deals with label switching between multiple calculations using the same *K* in the analysis (Jakobsson and Rosenberg, 2007). Bar plots were made using STRUCTURE PLOT, v1.0 (Ramasamy et al., 2014). Principal component analysis was performed on centered genotype data using the function 'svd' in R (R Development Core Team, 2016). Missing marker genotypes were replaced with the numeric genotype mean for that marker to perform PCA. A two-sample *t*-test was conducted in SAS to test differences in measured phenotypes between subpopulations.

The software PLINK, v1.07 (Purcell et al., 2007) was used to estimate the LD parameter $r^2$ among loci. All 2,735 polymorphic SNPs were used for LD analysis. The LD parameter $r^2$ was estimated for all linked and unlinked loci on all chromosomes. Loci were considered linked if they were on the same chromosome and unlinked if they were on different chromosomes. The $r^2$ values of the linked loci on all chromosomes were plotted against genetic distance (cM) and a smooth non-linear regression line was drawn in R (Marroni et al., 2011). A critical value of $r^2$, as an evidence of linkage, was derived from the 95$^{th}$ percentile distribution of square root transformed unlinked $r^2$ estimates (Breseghello and Sorrells, 2006b). The intersection of the regression line

with the baseline drawn at the critical value of $r^2$ was considered as an estimate of the average LD decay in the population (Breseghello and Sorrells, 2006b).

### 6.2.3 Marker-trait association analysis

Genome-wide association mapping was conducted using MLM that combines both population structure information and the level of pairwise relatedness coefficients or kinship-matrix as covariates using the software TASSEL, v3.0 (Bradbury et al., 2007). The MLM approach was followed to test marker-trait associations because it is effective in removing the confounding effects of population structure in association analysis (Yu et al., 2006). A kinship-matrix was computed from the marker data using TASSEL. For each trait, MLM regression was conducted using two different approaches to correct for population structure. The first approach was based on the population membership coefficients (Q-matrix) obtained from the Bayesian clustering using STRUCTURE software. However, earlier studies showed that relatedness among individuals strongly affects STRUCTURE outputs (Camus-Kulandaivelu et al., 2007). STRUCTURE assumes that the marker loci are unlinked and at linkage equilibrium with one another and Hardy-Weinberg equilibrium within populations (Pritchard et al., 2000). The materials used in this study are elite breeding lines with different degrees of relatedness and the STRUCTURE output may be biased. Another common approach to account for population structure in association mapping is the use of marker derived PCs (Price et al., 2006). In this study, marker-derived PCs that describe population stratification as covariates were also included in the regression model. For each trait, marker-trait associations were tested using five different models. The first model controlled only for kinship (K). The other four models corrected for both kinship and population structure information included using the Q-matrix (QK) as well as two (2PC+K), three (3PC+K) and five PCs (5PC+K). Mixed linear model analyses were conducted using the default settings of TASSEL (optimum compression level and PD3 variance component estimation). Associations were declared significant based on a false discovery rate (FDR) of 0.2 in the MLM to control for multiple testing (Benjamini and Hochberg, 1995). The FDR was calculated for all SNPs based on the 'BH' (Benjamini and Hochberg) method using the R function 'p.adjust' (R Development Core Team, 2016). Marker probabilities with $-\log_{10}$ ($P$-value) $\geq 3.0$ were also considered significant if the region was reported in previous studies.

## 6.3 Results and Discussion

### 6.3.1 Population structure and LD decay

Based on the Bayesian clustering, the optimal number of subpopulations was estimated as $K = 2$ according to the value *of* $\Delta K$ (Fig. 6-1). Similarly, the dendrogram clustered the 231 lines into two major groups (Fig. 6-2A). The bar plot based on the population membership coefficient obtained from the Bayesian clustering mostly corresponded with the grouping based on the hierarchical clustering, but it also indicated the presence of substantial admixtures (Fig. 6-2B).

The screeplot showed that the first two PCs explained 20% of the genetic variance, indicating that population structure effects are mild in this population (Fig. 6-3). Plots of the first two PCs clustered the 231 wheat lines into two weakly differentiated groups which showed some overlaps (Fig. 6-4). Both the screeplot and plots of the first two PCs indicated that the lines used in this study did not represent a highly-structured population. Absence of clear differentiation between the two clusters could be because the population is composed of elite breeding materials that have been intercrossed frequently resulting in admixture of germplasm.

The *t*-test analysis comparing means of the two groups indicated that there were significant differences between the two subpopulations in plant height ($t = 3.98$, $P < 0.001$), maturity ($t = 5.45$, $P < 0.001$), test weight ($t = 3.44$, $P < 0.001$), thousand-kernel weight ($t = 2.73$, $P = 0.007$), grain protein ($t = 4.78$, $P < 0.001$), falling number ($t = 4.80$, $P < 0.001$) and sedimentation volume ($t = 4.26$, $P < 0.001$). However, no significant difference was observed between the two groups for heading date ($t = 1.4$, $P = 0.164$) and grain yield ($t = 1.39$, $P = 0.165$).

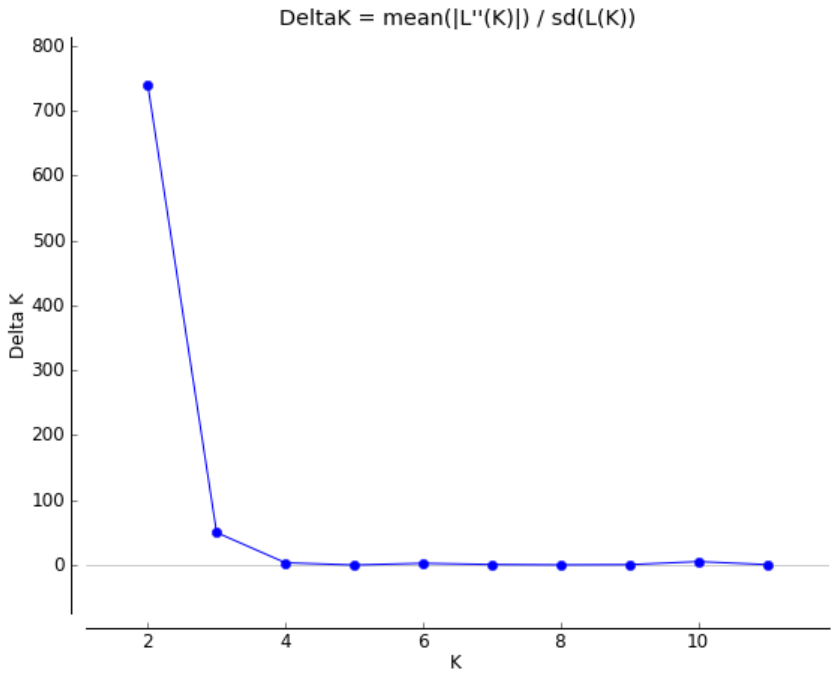



**Fig. 6-1.** The ΔK peak value at K = 2, that indicated the presence of two subpopulations based on Bayesian clustering.

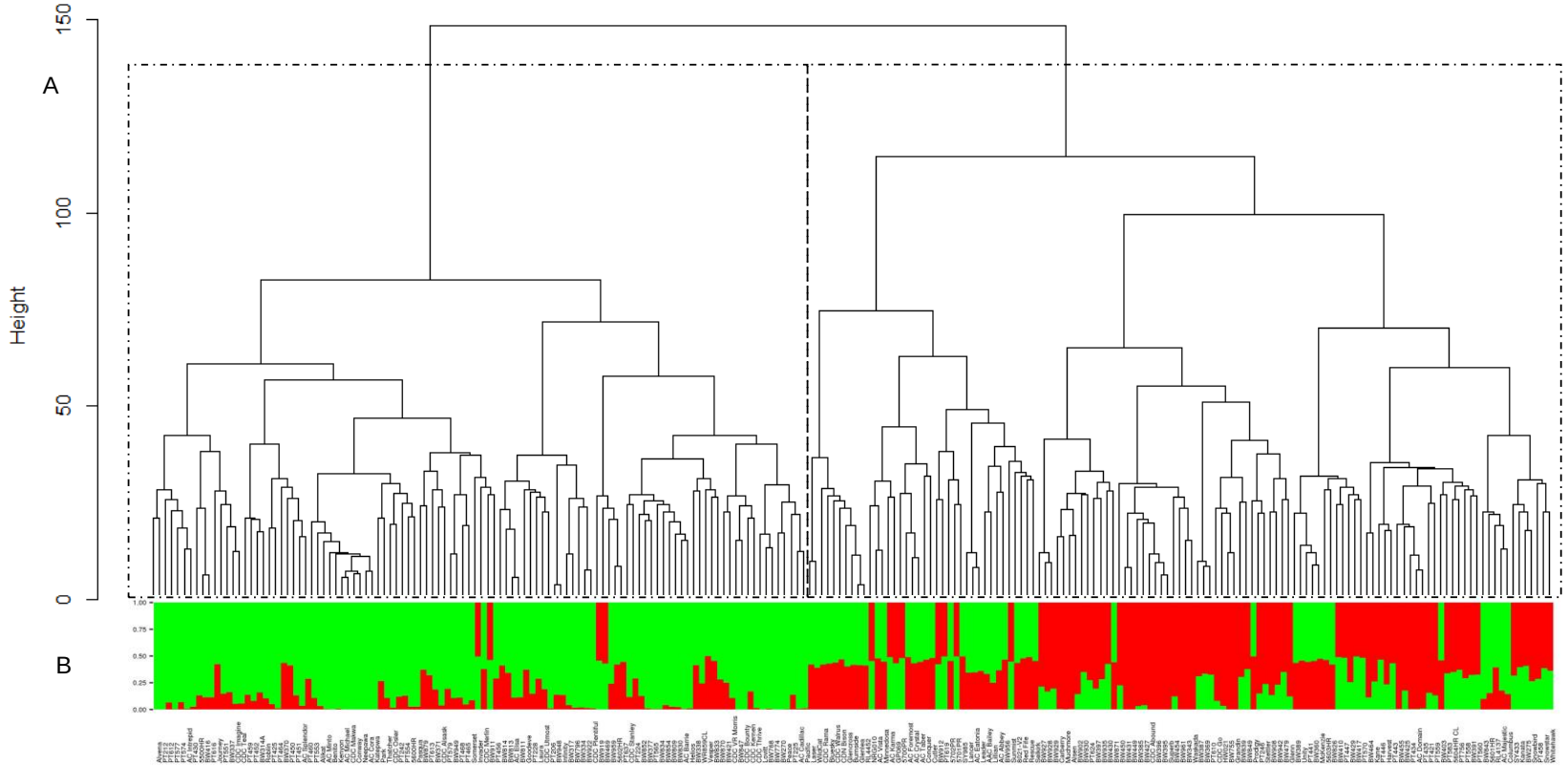



**Fig. 6-2.** Population structure analysis. (A) Hierarchical clustering using the Ward's method, (B) bar plots showing subpopulations represented by different colors based on the population membership coefficients obtained from Bayesian clustering.

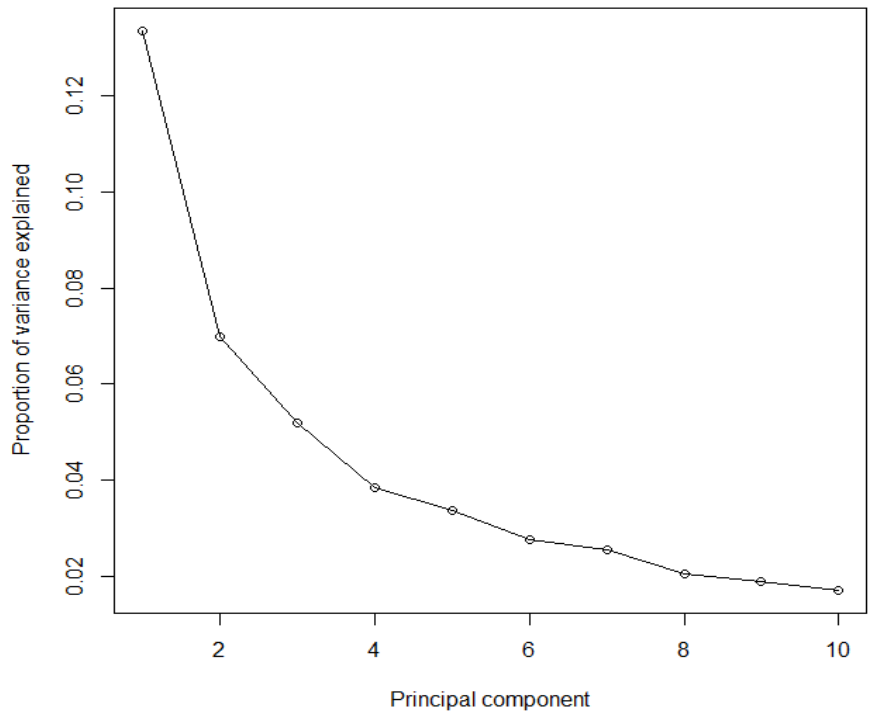



**Fig. 6-3.** Proportion of variance explained by marker based principal components.

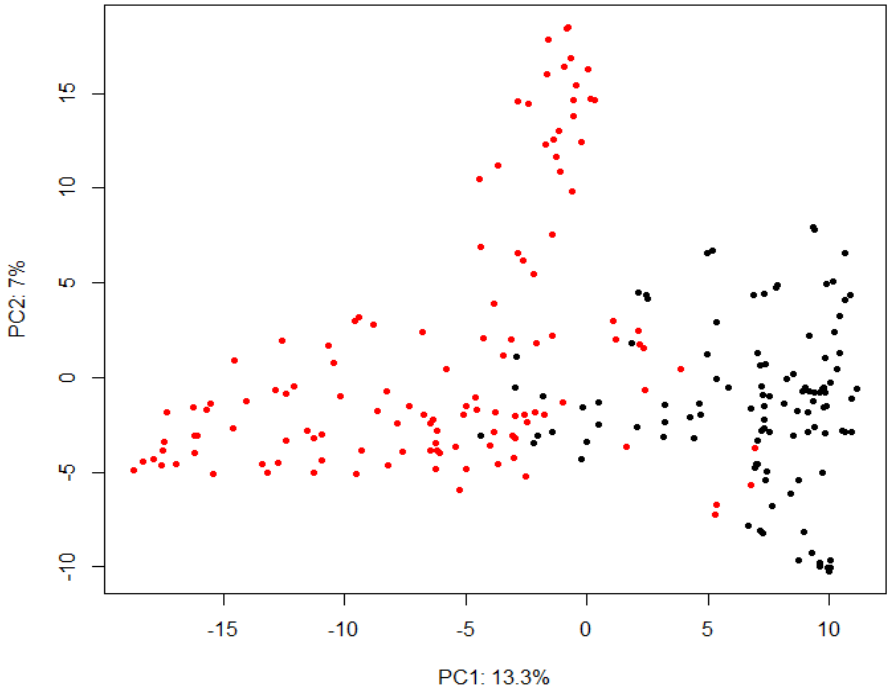



**Fig. 6-4.** First two principal components using 2,735 SNP markers run on 231 wheat lines. Each solid circle represents one line. The colors correspond to the two subpopulations identified from hierarchical clustering.

LD decay was estimated across all chromosomes using 2,735 polymorphic SNPs. A critical $r^2$ value was estimated at 0.28 based on the 95$^{th}$ percentile of the distribution of unlinked $r^2$ estimates (Fig. 6-5). On average, LD across all chromosomes decayed at a genetic distance of 1.6 cM (Fig. 6-5). This is slightly lower compared to the 2-3 cM average extent of LD previously reported for Canadian bread wheat and durum accessions (Somers et al., 2007).

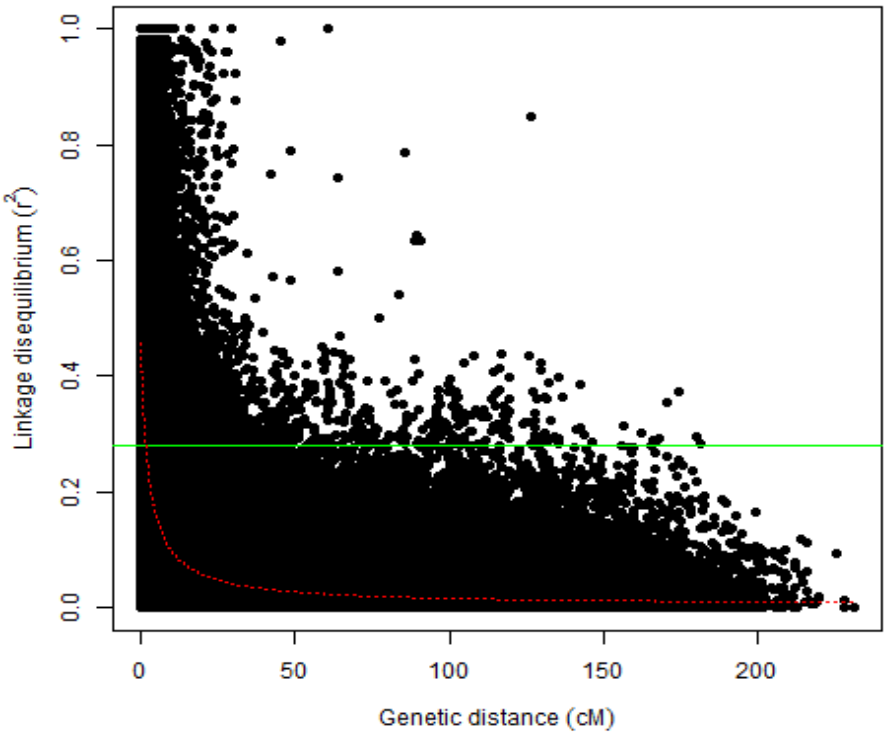



**Fig. 6-5.** Genome-wide LD ($r^2$) values plotted against genetic distance (cM) for the 231 wheat lines. The green horizontal line indicates the 95$^{th}$ percentile of the distribution of unlinked $r^2$ while the fitted curve (red line) indicates the LD decay.

### 6.3.2 Marker-trait associations

Marker-trait associations were conducted using the MLM approach that combined both population structure and the level of pairwise relatedness. To check whether each model adequately controlled population structure and pairwise relatedness, quantile-quantile (Q-Q) plots were generated for all model-trait combinations based on the observed *P*-values for all SNPs (Y-axis) and the expected distribution of *P*-values under the null hypothesis of no association (X-axis) (Fig. 6-6). For all traits, the Q-Q plots indicated that there was no difference between the model that did not account for population structure and the other models that corrected for both population structure and kinship. This indicates that the effect of population structure was not strong in this population. For all traits, the Q-Q plots were close to the diagonal line except for deviations towards the upper-right end of the diagonal suggesting that models adequately controlled false positives because of population structure or kinship. The results reported in this study were based on the MLM 5PC+K model for all traits. Although the effect of population structure was not strong, we used five PCs to account for population structure because it explained about 30% of the genetic variance.

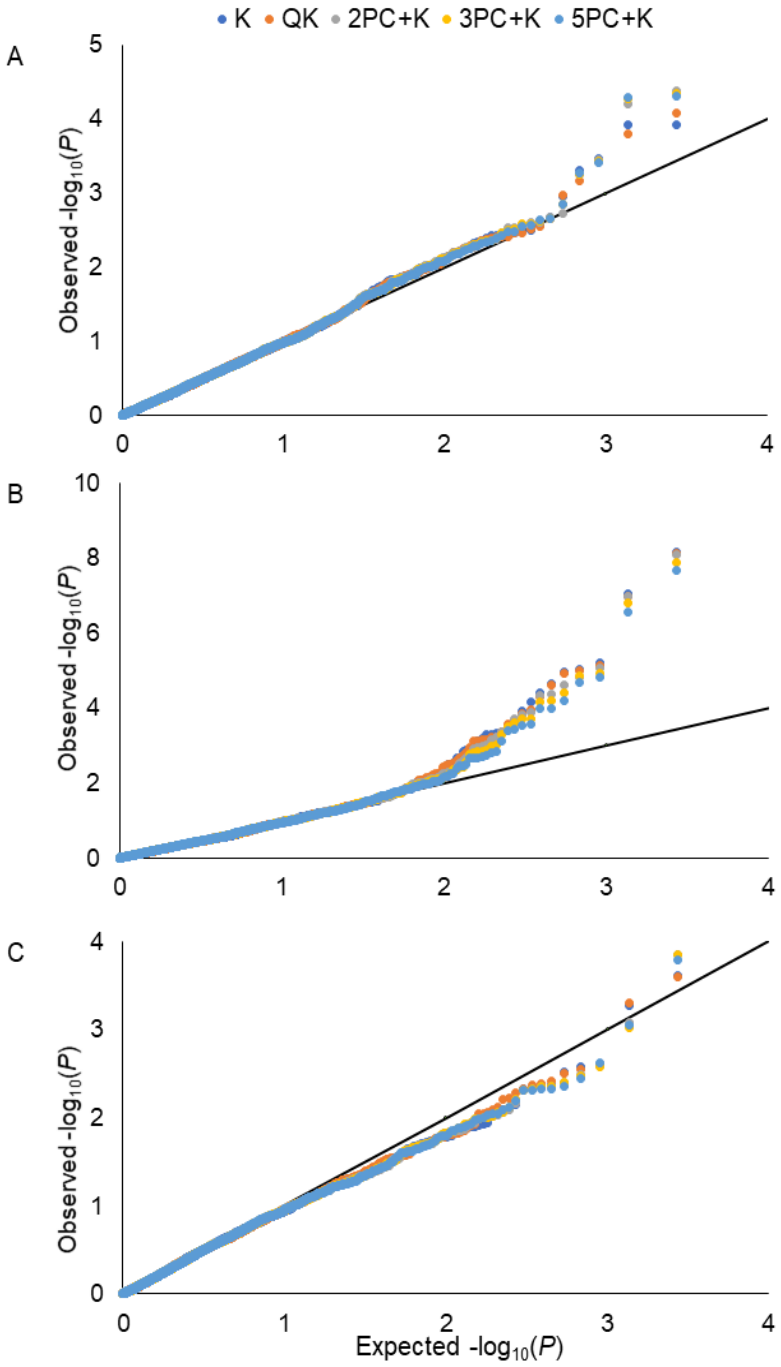



**Fig. 6-6.** Q-Q plots comparing the distribution of observed versus expected *P*-values for association analyses using five statistical models. (A) heading date, (B) plant height, (C) maturity, (D) grain yield, (E) test weight, (F) thousand-kernel weight, (G) grain protein, (H) falling number, (I) sedimentation volume.

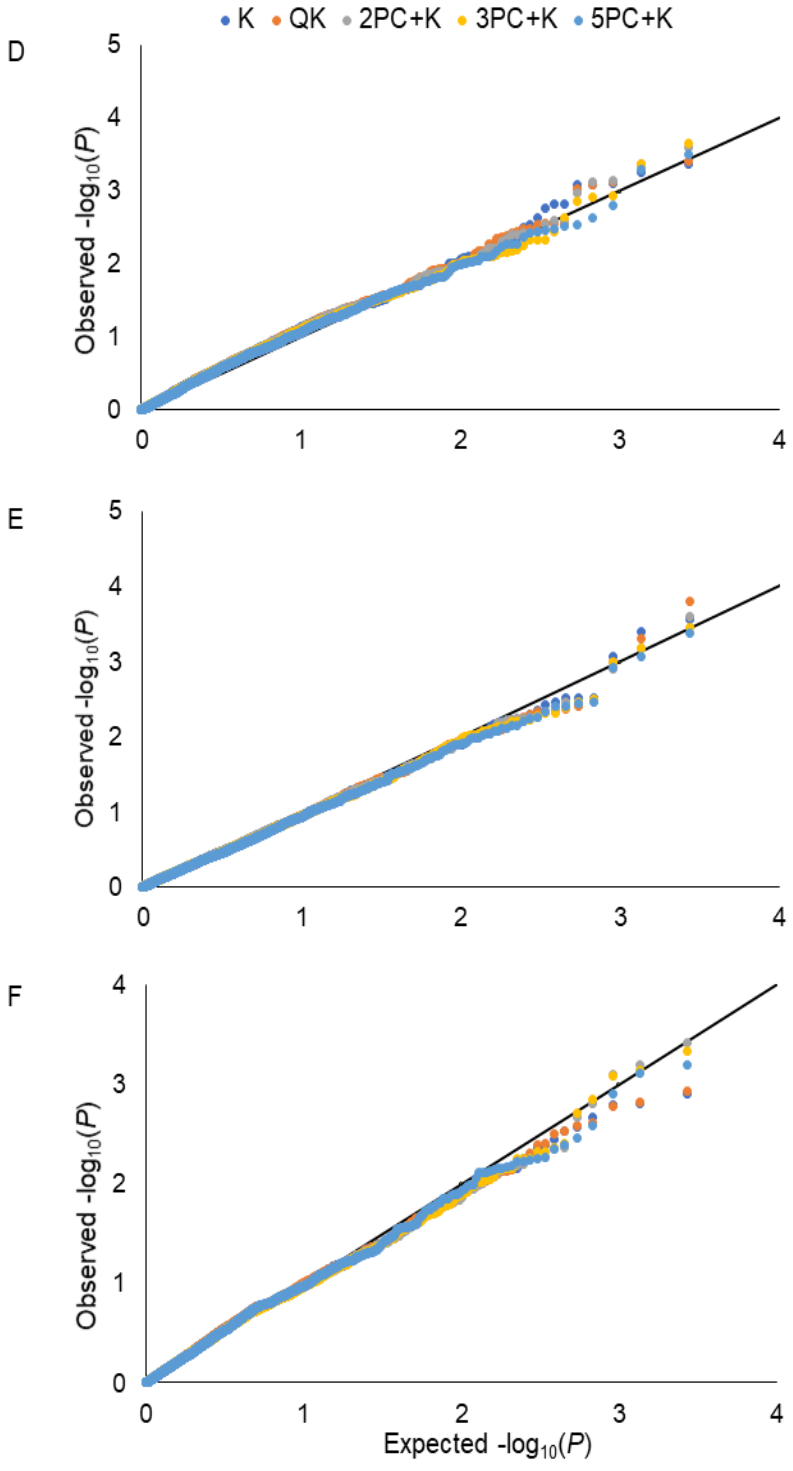



**Fig. 6-6.** Continued

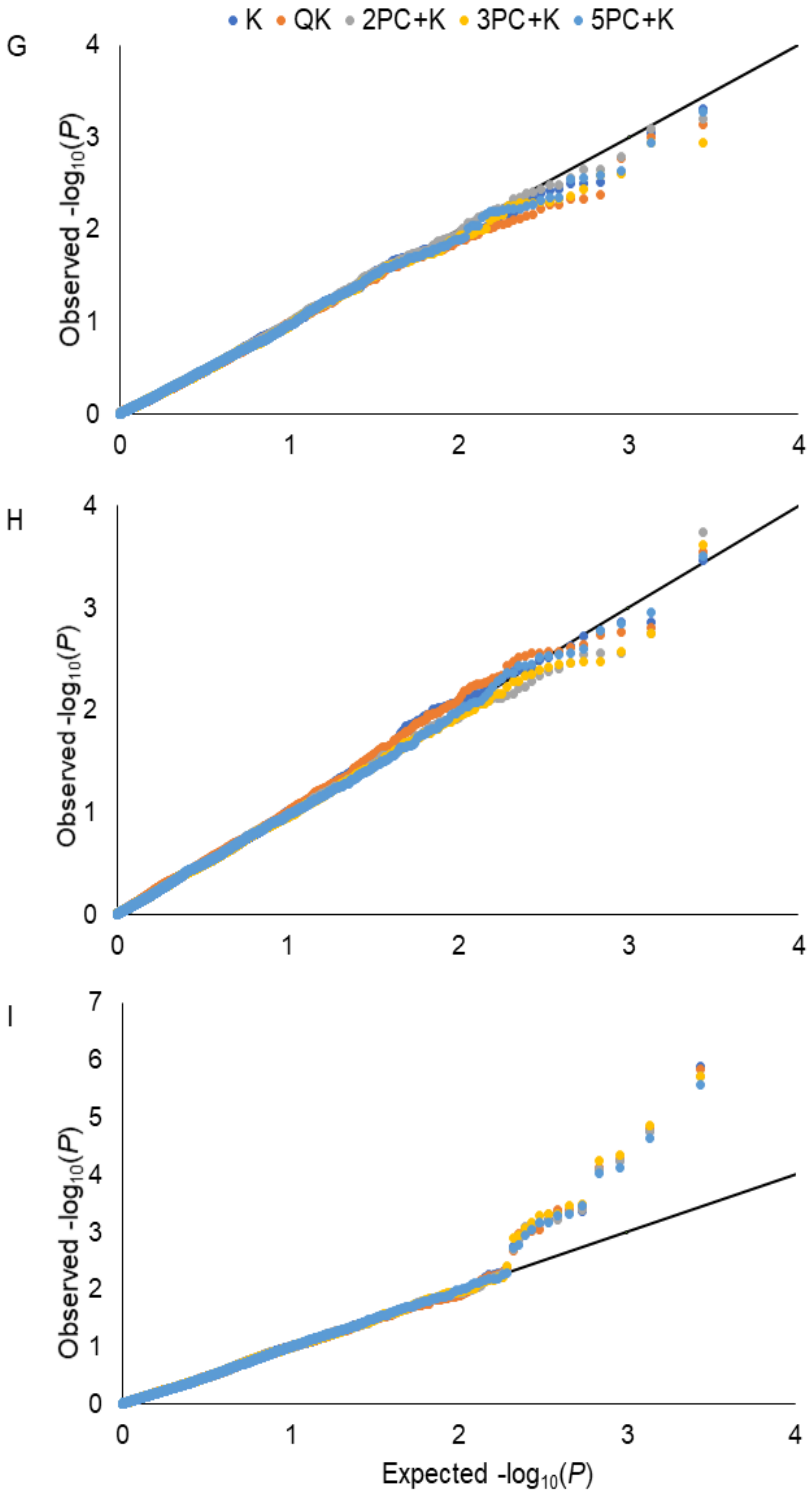



**Fig. 6-6.** Continued

We used elite cultivars and advanced breeding material to identify marker-trait associations for important agronomic and end-use quality traits. A total of 34 significant marker-trait associations were detected for eight traits at a significance threshold of $-\log_{10}$ (*P*-value) ≥ 3.0 (Table 6-1; Appendix D). Twelve markers were significantly associated with plant height, ten markers were associated with sedimentation volume, four markers were associated with heading date, three markers were associated with thousand-kernel weight, two markers were associated with maturity, and one marker was associated with each of grain yield, test weight, and falling number.

Twelve SNPs on three chromosomes were significantly associated with plant height (Table 6-1; Appendix D). Nine of these SNPs were located on chromosome 4B (39.9 – 72.5 cM), two SNPs on chromosome 2A (109.5 – 110.1 cM) and one SNP on chromosome 5B (115.7 cM). Each of these SNPs explained 6 to 15% of the phenotypic variance across environments. The genomic regions on 4B and 5B corresponded to genes or QTLs that were reported in wheat, but no plant height QTL was reported previously on 2A. There are several reduced height (*Rht*) genes that affect plant height in wheat. The two major genes *Rht-B1b* (*Rht1*) and *Rht-D1b* (*Rht2*) are located on the short arm of chromosomes 4B and 4D, respectively (Börner et al., 1997). Moreover, *Rht-B1c* (*Rht3*) was localized in bread wheat near the centromere on the short arm of chromosome 4B (Börner et al., 1997). Additional marker analyses could confirm that these genes correspond to the region identified on chromosome 4B in this study. A prominent QTL for plant height was reported on the long arm of chromosome 5B in wheat based on a DH population derived from a cross between 'RL4452' and 'AC Domain' (McCartney et al., 2005).

A total of ten significant marker-trait associations were detected for sedimentation volume (Table 6-1; Appendix D). Sedimentation volume is a measure of bread making quality. In wheat, bread making quality is most dependent on the quality and quantity of wheat protein (Bushuk et al., 1969). Glutenin and gliadins are major components of the wheat protein responsible for bread making quality. The ten SNPs were localized to three chromosomes 1A (19.3 and 111.6 cM), 1B (121.7 – 125.3 cM), and 5D (102.9 – 105.4 cM). Each of these SNPs explained from 5 to 10% of the phenotypic variance. In wheat, the genes which are responsible for encoding glutenin and gliadins subunits are located on homoeologous regions on chromosomes one and six (Payne, 1987). Genes controlling the high molecular weight glutenin subunits *Glu-A1*, *Glu-B1*, and *Glu-D1* are located on the long arms of chromosomes 1A, 1B and 1D (Payne, 1987). Similarly, genes

responsible for coding low molecular weight glutenin subunits are located on the short arms of chromosomes 1A, 1B and 1D while genes for α- and β-gliadins occur on the short arms of chromosomes 6A, 6B, and 6D (Payne, 1987). Huang et al. (2006) reported a QTL for sedimentation volume that explained 8.8% of the phenotypic variance on the short arm of chromosome 5D based on 185 DH lines derived from a cross between a Canadian wheat variety 'AC Karma' and a breeding line '87E03-S2B1'. Similarly, a QTL that explained 17.3% of the variance in sedimentation volume was reported on 5DS (Kunert et al., 2007). The SNPs associated with sedimentation volume on chromosomes 1A and 1B could be associated with the previously described genes and QTL, while the SNP on 5D could be associated with a different QTL. Further research is required to validate the identified SNPs associated with sedimentation volume.

The four SNPs that were associated with heading date were located on chromosomes 2D (19.3 cM), 5B (97.28 and 110.56 cM) and 5D (80.68 cM) (Table 6-1; Appendix D). Each of these SNPs explained 6 to 8% of the phenotypic variance and together explained 28% of the variation in heading date across environments. Flowering time in wheat is controlled by three groups of genes, namely vernalization response (*Vrn* genes), photoperiod response (*Ppd* genes), and earliness *per se* genes (*Eps* genes) (Snape et al., 2001; Worland, 1996). The most important vernalization response genes *VRN-A1, VRN-B1,* and *VRN-D1* were mapped previously in collinear regions on the long arm of group five chromosomes (Dubcovsky et al., 1998; Law et al., 1976), while the photoperiod response genes *Ppd-A1*, *Ppd-B1,* and *Ppd-D1* were mapped to homologous positions on the short arm of group two chromosomes (Scarth and Law, 1984; Worland et al., 1998). Earliness *per se* genes were also reported on several chromosomes in wheat (Worland, 1996). A previous study reported earliness *per se* QTL on chromosome 5B based on Canadian spring wheat cultivars (Kamran et al., 2013). The SNPs identified in this study may correspond to these previously described genes.

Three SNPs in two genomic regions were associated with thousand-kernel weight. Two SNPs were located on chromosome 6A (85.1 – 85.7 cM) and one SNP on 7B (78.3 cM). Recently, Zou et al. (2017) reported QTL for thousand-kernel weight on chromosomes 6A (79 cM) and 7B (158 cM) based on RILs derived from a cross between 'Attila' and 'CDC Go' spring wheat cultivars. Major QTL associated with thousand-kernel weight that explained 16.1 to 22.4% of the variance were also reported in spring bread wheat on chromosome 6A (Simons et al., 2012).

**Table 6-1.** Association of markers with eight agronomic and end-use quality traits detected with a mixed linear model. Associations that passed the FDR threshold are indicated in bold.

| Trait† | SNP ID | Chromosome | Position (cM) | *P*-value | -log$_{10}$*P*-value | Marker $R^2$(%) |
|---|---|---|---|---|---|---|
| **HT** | **BS00022896_51** | **2A** | **109.52** | **4.05E-04** | **3.4** | **6** |
| **HT** | **BS00012320_51** | **2A** | **110.13** | **1.46E-05** | **4.8** | **9** |
| **HT** | **RAC875_c12959_869** | **4B** | **39.93** | **7.42E-04** | **3.1** | **7** |
| **HT** | **Tdurum_contig64772_417** | **4B** | **50.85** | **1.02E-04** | **4.0** | **7** |
| **HT** | **BobWhite_rep_c49034_132** | **4B** | **55.55** | **2.13E-05** | **4.7** | **10** |
| **HT** | **Tdurum_contig33737_157** | **4B** | **55.96** | **2.70E-07** | **6.6** | **12** |
| **HT** | **IAAV971** | **4B** | **57.49** | **2.11E-08** | **7.7** | **15** |
| **HT** | **Excalibur_c56787_95** | **4B** | **58.10** | **3.84E-04** | **3.4** | **6** |
| **HT** | **Kukri_c11415_1074** | **4B** | **68.45** | **2.87E-04** | **3.5** | **6** |
| **HT** | **Kukri_c17224_278** | **4B** | **71.29** | **1.03E-04** | **4.0** | **8** |
| **HT** | **wsnp_Ra_c22026_31453420** | **4B** | **72.53** | **6.25E-05** | **4.2** | **7** |
| **HT** | **BS00022673_51** | **5B** | **115.69** | **2.62E-04** | **3.6** | **6** |
| **SDS** | **CAP12_c3074_192** | **1A** | **19.34** | **3.54E-04** | **3.5** | **6** |
| SDS | BobWhite_c26569_190 | 1A | 111.55 | 7.04E-04 | 3.2 | 5 |
| **SDS** | **BS00077498_51** | **1B** | **121.71** | **9.58E-05** | **4.0** | **7** |
| SDS | RFL_Contig16_132 | 1B | 122.14 | 5.24E-04 | 3.3 | 7 |
| **SDS** | **BS00035267_51** | **1B** | **122.38** | **2.32E-05** | **4.6** | **8** |
| **SDS** | **BS00068077_51** | **1B** | **122.52** | **7.62E-05** | **4.1** | **9** |
| SDS | wsnp_JD_c14411_14148961 | 1B | 125.07 | 6.82E-04 | 3.2 | 7 |
| SDS | BobWhite_c14362_86 | 1B | 125.26 | 4.82E-04 | 3.3 | 6 |
| **SDS** | **BS00000020_51** | **5D** | **102.91** | **2.73E-06** | **5.6** | **10** |
| SDS | Excalibur_c49805_63 | 5D | 105.38 | 9.31E-04 | 3.0 | 6 |
| **HD** | **wsnp_CAP12_c812_428290** | **2D** | **19.03** | **3.81E-05** | **4.4** | **8** |
| HD | wsnp_Ex_c13485_21225504 | 5B | 97.28 | 3.25E-04 | 3.5 | 6 |
| **HD** | **BS00065128_51** | **5B** | **110.56** | **3.36E-05** | **4.5** | **8** |
| HD | Excalibur_c15835_86 | 5D | 80.68 | 8.60E-04 | 3.1 | 6 |
| TKW | BS00063096_51 | 6A | 85.07 | 7.96E-04 | 3.1 | 7 |
| TKW | Jagger_c8913_220 | 6A | 85.66 | 8.89E-04 | 3.1 | 6 |
| TKW | BobWhite_c7082_196 | 7B | 78.31 | 4.06E-04 | 3.4 | 7 |
| MAT | wsnp_CAP12_c812_428290 | 2D | 19.03 | 2.47E-04 | 3.6 | 6 |
| MAT | Tdurum_contig8072_1192 | 5B | 75.79 | 5.40E-04 | 3.3 | 5 |
| YLD | BS00062731_51 | 5B | 104.55 | 3.20E-04 | 3.5 | 6 |
| TWT | BS00000592_51 | 5B | 93.44 | 8.46E-04 | 3.1 | 5 |
| FN | Excalibur_c15048_488 | 2D | 37.62 | 3.06E-04 | 3.5 | 6 |

†HT: plant height, SDS: sedimentation volume, HD: heading date, TKW: thousand-kernel weight, MAT: maturity, YLD: grain yield, TWT: test weight, FN: falling number.

The SNPs associated with maturity were also in similar regions with previously reported genes or QTL. Maturity was associated with two SNPs located on 2D (19.3 cM) and 5B (75.8 cM). Each SNP explained 6% of the phenotypic variance across environments. In wheat, chromosomes 2D and 5B are known to carry the photoperiod response (*Ppd-D1*) and vernalization response (*VRN-B1*) genes, respectively (Worland, 1996). These genes control important phases of growth and development (tillering, stem extension, heading, flowering and ripening) in wheat (Košner and Pánková, 1998). Recently, Zhou et al. (2017) reported a maturity QTL on 2D (confidence interval of 0-14.5 cM) flanked by *Ppd-D1* gene and on 5B (56 cM). The SNP marker '*wsnp_CAP12_c812_428290*' was significantly associated with both heading date and maturity on chromosome 2D (Table 6-1). This indicates that either the same QTL is controlling heading date and maturity (pleiotropy) or QTL controlling these traits are closely linked on chromosome 2D.

Only one SNP was found associated with each of grain yield, test weight and falling number. SNPs on chromosome 5B (104.6 cM and 93.4 cM) were associated with grain yield and test weight, respectively. These SNPs explained a small proportion of the phenotypic variance (6 and 5%, respectively), highlighting the quantitative nature of these traits. Previous studies reported QTL for grain yield and test weight in spring bread wheat on the short and long arms of chromosome 5B, respectively (Perez-Lara et al., 2016; Simons et al., 2012). Similarly, one SNP on 2D (37.6 cM) was associated with falling number that explained 6% of the phenotypic variance. Kunert et al. (2007) reported a QTL for falling number on 2D approximately 14 cM away from the SNP detected in this study. An unexpected result from this study was that no marker-trait association was detected for grain protein content. This could be because of the narrow variation in grain protein content among the lines used in this study. Sampling additional wheat lines with more variation in protein content would be required to identify significant associations.

## 6.4 Conclusion

In this study, 34 significant marker-trait associations were detected for eight agronomic and end-use quality traits. The genetic effects of most of the identified QTL were relatively small, explaining less than 10% of the phenotypic variance, highlighting the quantitative nature of the evaluated traits. The identified marker-trait associations would need to be validated in different genetic backgrounds to be useful for MAS. However, these markers can be fitted as fixed effects in GS to enhance genomic prediction accuracy.

# 7. Mapping of Quantitative Trait Loci Associated with Agronomically Important Traits in Wheat

## 7.1 Introduction

Canada is the sixth largest producer of wheat in the world, much of which is exported for its high-quality end-use purposes. Canadian wheat varieties are grouped into different market classes based on their end-use quality parameters (Canadian Grain Commission, 2015). Wheat breeding programs in Canada target the release of cultivars that meet the necessary standards for agronomic performance, resistance to biotic factors, and quality attributes for cultivar registration and growth (DePauw et al., 2011). Many of the agronomically important traits that are of interest to wheat breeders are quantitatively inherited. Quantitative traits have a wide range of distributions and are controlled by multiple small-effect genes that are influenced by the environment, interaction among genes, and interactions between genes and the environment. Understanding the genetic control of these traits in wheat is an important step for marker assisted breeding. QTL mapping involves the identification of genomic regions that control quantitative traits in segregating populations developed from homozygous inbred parents (Doerge, 2002). Experimental mapping populations such as F2 populations, backcross populations, RILs, and doubled haploids are commonly used for QTL mapping in plants (Collard et al., 2005; Flint-Garcia et al., 2003). QTL mapping has been used to dissect the genetic basis of variation of complex traits and identify loci with large phenotypic effects. Genetic dissection of a quantitative trait will succeed only when some of the QTLs segregating in the population have relatively large effects (Lander and Botstein, 1989). When markers that are linked to QTL with large effects are identified they can be used for MAS. However, MAS has limited application for traits controlled by many QTL with small effects.

Genomic selection has emerged as an alternative to the standard MAS for improving complex traits that are controlled by many QTL with small effects (Meuwissen et al., 2001). Genomic selection uses genome-wide markers without significance testing to predict the breeding value of individuals in a breeding population (Meuwissen et al., 2001). This ensures that all major and minor effect QTL underlying traits are utilized for MAS. Genomic prediction models, such as ridge regression, treat genome-wide markers as random effects and shrink their effects uniformly without distinguishing between known major genes or QTL and unknown QTL with minor effects (Bernardo, 2014). This may result in underestimation of the effects of large effect QTL and affect the genetic gain from GS (Bernardo, 2014). Recently, a new GS approach was suggested to model

known major genes or QTL as fixed effects and unknown small effect QTL as random effects to enhance genomic predictions (Bernardo, 2014; Rutkoski et al., 2014; Spindel et al., 2016; Spindel et al., 2015; Zhang et al., 2014). Therefore, the objective of this study was to identify QTL controlling important agronomic traits in RILs developed from three-way crosses of spring bread wheat varieties. Markers significantly associated with QTL identified from this study will be used to enhance the accuracy of genomic predictions in this population.

## 7.2 Materials and Methods

### 7.2.1 Plant material and phenotypic data

This study used a mapping population of 304 RILs developed from three-way crosses (CDC Plentiful//Pasteur/CDC Utmost) made at the CDC, University of Saskatchewan. We also used these lines as SC in GS study (Chapter eight). Pasteur is a short-statured, later maturing but high yielding general-purpose wheat cultivar from Wiersum Plant Breeding in the Netherlands. CDC Utmost and CDC Plentiful are standard height, early to medium maturing and high yielding cultivars from the CDC, University of Saskatchewan. The first cross was made in a controlled environment facility (Phytotron) between Pasteur and CDC Utmost during early fall of 2011 and the second cross was made to CDC Plentiful during the winter of 2011/2012. The F1 generation was grown in the Phytotron during summer 2012. Seed from F1 was bulked and the F2 generation was grown in a greenhouse at the University of Saskatchewan. The F3 generation was advanced through single seed decent. Two seeds from each F2 spike were bulked and random samples were planted at a seeding rate of eight seeds per 3.78-liter pot. The F3 spikes were threshed individually and each were planted (F4 generation) on single hill plots at the Seed Farm during spring 2013. A total of 2,712 F4 entries were grown using unreplicated hills with the three parents and 'AC Barrie' grown as replicated check cultivars. A single spike was harvested and threshed individually from each F4 plant and the F4:F5 generation was grown under field conditions in a winter nursery during the winter of 2013/2014. The population was then advanced to the F6, F7, and F8 generations for field trials and phenotyping.

Five independent field trials were conducted across two research sites for the F4:F6, F4:F7 and F4:F8 generations. In the F4:F6 generation, 506 entries were randomly selected, and each were grown at Kernen on a 2.48 $m^2$ plot area with four seeded rows during the spring/summer of 2014. The field experiments were arranged in 11 blocks, each containing 50 plots (46 entries and four

check cultivars), for a total of 550 plots. In the F4:F7 generation, 322 entries were randomly selected and advanced to the F8 generation. The F4:F7 and F4:F8 generations were grown both at Kernen and in Rosthern on a 4.25 $m^2$ plot area, with five seeded rows during the spring/summer of 2015 and 2016. Seeding rates were 300 seeds per $m^2$. The field experiments were arranged in seven blocks, each containing 50 plots (46 entries and four check cultivars), for a total of 350 plots at each environment (site-year). The lines were arranged in an augmented randomized complete block design (ARCBD), where the four check cultivars were randomly assigned to plots within each block and unreplicated entries were randomly arranged in the remaining plots (Federer, 1961). In summary, five field trials were conducted; the F4:F6 generation at Kernen in 2014, as well as the F4:F7 and F4:F8 generations at Kernen and Rosthern in both 2015 and 2016. Traits including heading date, plant height, maturity, grain yield, test weight and thousand-kernel weight were measured as described in Section 3.2.1.

### 7.2.2 Phenotypic data analysis

The phenotypic data were analyzed using analysis of variance (ANOVA) with SAS Mixed models, v9.4 (SAS Institute Inc., 2015). Phenotypic data were analyzed separately in each environment (Kernen 2014, Kernen 2015, Kernen 2016, Rosthern 2015, and Rosthern 2016) and then combined across environments. For the separate analysis, lines (entries plus check cultivars) were considered as fixed effects and block was considered as a random effect. For the combined analysis, environment, block nested in environment and line-by-environment interactions were considered as random effects. To control for block-to-block heterogeneity, trait values of entries were adjusted relative to the four check cultivars repeated in each block using the LSMEANS procedure in SAS (Wolfinger et al., 1997). The phenotypic data analyses included the 506 entries in the F4:F6 generation, 322 entries in the F4:F7 and F4:F8 generations, and the four check cultivars for a better estimate of (co)variances but only 304 lines that were evaluated across all generations and have marker data were used for QTL analyses. Broad-sense heritability ($H^2$) and Pearson correlation coefficients among the traits were estimated as described in Section 3.2.3.

### 7.2.3 Genotypic data

Genomic DNA was extracted from fresh leaves of one-week-old seedlings in the F4:F7 generation and all lines were genotyped using the same methods described in Section 3.2.2.

Genotype calling was performed using the GenomeStudio Polyploid Clustering Module v1.0 (Illumina, San Diego, CA), and erroneous lines and markers were filtered from the analysis (Appendix E). A total of 16,115 polymorphic SNPs with call frequency (CF) > 90% and minor allele frequency (MAF) > 20% were obtained. For QTL analyses, a thinned subset of 1,219 evenly spaced SNPs (mean genetic distance of 2.9 cM between adjacent SNPs) was selected based purely on genetic distances, using the software MapThin, v1.11 (Howey and Cordell, 2011). The chromosomal positions of the SNP markers were determined based on the hexaploid wheat consensus genetic linkage map (Wang et al., 2014a).

### 7.2.4 QTL analyses

QTL analyses were performed using LS-means of phenotypes in each environment and averaged (combined) across all environments. QTL mapping was performed with the inclusive composite interval mapping (ICIM) procedure using QTL IciMapping v4.1 (Meng et al., 2015). Analyses of the additive effects at individual QTL (ICIM-ADD) were performed with a critical logarithm of odds (LOD) threshold estimated for each trait based on 1,000 permutations at a significance level of 0.05. Mapping parameters of 1 cM walking distance and deletion of missing phenotypes were applied. Significant QTL were displayed on individual chromosome maps using MapChart v2.3 (Voorrips, 2002).

## 7.3 Results and Discussion

### 7.3.1 Wide variations were observed in measured phenotypes

Analysis of variance indicated that there were highly significant differences among the lines for all evaluated traits ($P < 0.001$). The frequency distributions of phenotypes averaged over the five environments also indicated that there were large differences between Pasteur and the two locally adapted parents for heading date, maturity, and grain yield while the differences were moderate for all the other traits (Fig. 7-1). Transgressive segregations in both directions of the distribution were observed for all traits (Fig. 7-1). Heading date among the lines varied from 53 to 62 days with a mean of 56(±2) days, while time to maturity varied from 89 to 100 days with a mean of 94(±2) days across environments. Plant height ranged from 77.7 to 105.6 cm with a mean of 93(±5.4) cm. There was also a wide variation among the lines for grain yield, which ranged from 2348 to 5465 kg $ha^{-1}$ with a mean of 4278 (±458) kg $ha^{-1}$. Test weight ranged from 75.8 to

82.5 kg $hL^{-1}$ with a mean of 79.7(±1.2) kg $hL^{-1}$ and thousand-kernel weight varied from 28.1 to 41.7 g with a mean of 35.2(±2.5) g across the environments. Moderate to high estimates of broad sense heritability were obtained for grain yield (0.55), plant height (0.55), maturity (0.64), test weight (0.71), heading date (0.75), and thousand-kernel weight (0.78) across all environments. Correlation coefficients between the evaluated traits indicated a very strong positive correlation between heading date and maturity (0.87) (Table 7-1). In contrast, there were weak positive correlations between heading date and grain yield (0.12) or test weight (0.22), while the correlation was negative between heading date and thousand-kernel weight (-0.19). Plant height was moderately correlated with heading date (0.41), maturity (0.35), and thousand-kernel weight (0.29). Similarly, test weight was positively correlated with maturity (0.31) and grain yield (0.35) (Table 7-1).

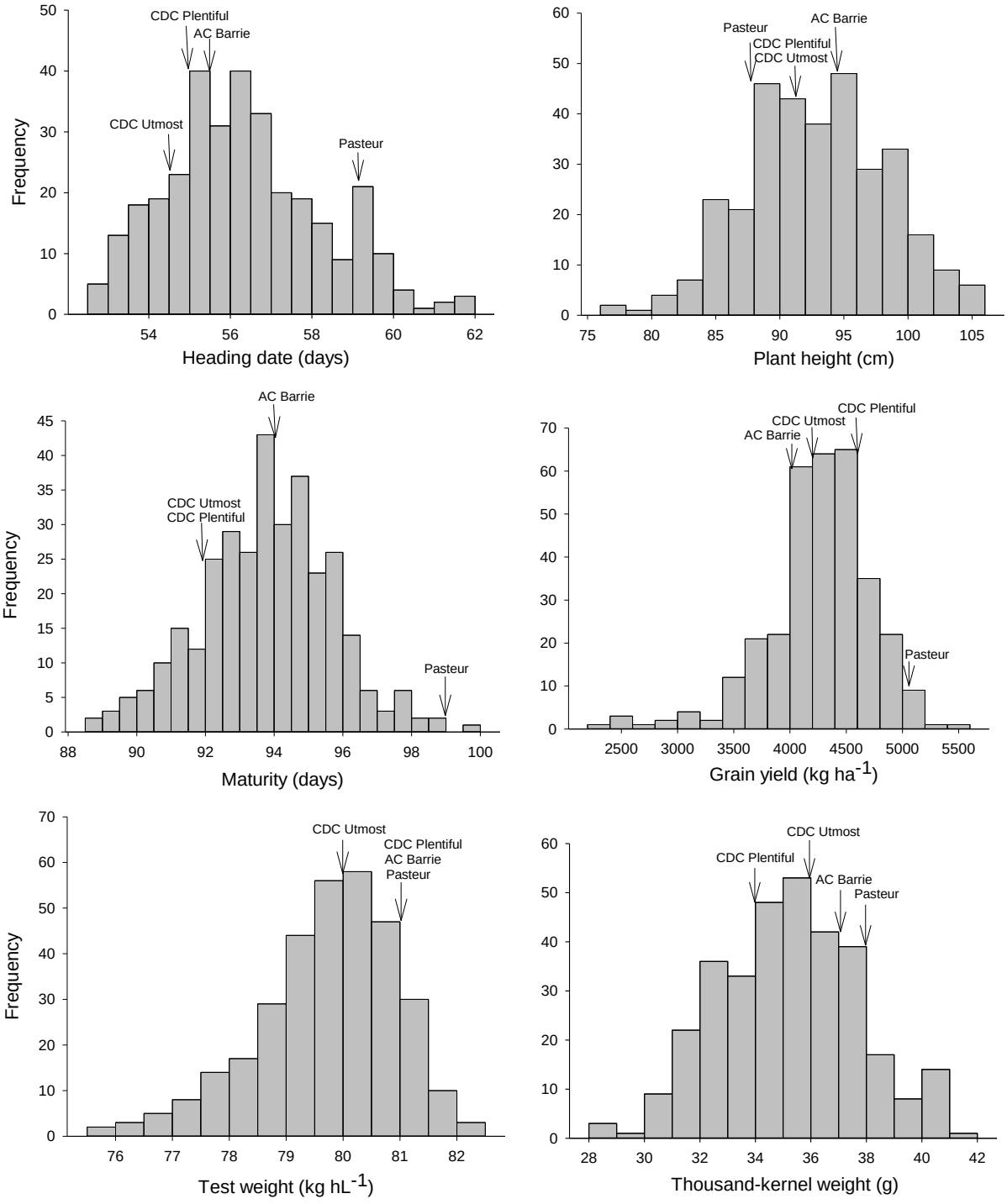



**Fig. 7-1.** Frequency distributions of six agronomic traits measured in the mapping population (selection candidates). Data were averaged across F6 to F8 generations and all environments. The values for the check cultivars (AC Barrie, CDC Utmost, CDC Plentiful, and Pasteur) are indicated with arrows.

**Table 7-1.** Correlations among adjusted traits of the selection candidates averaged across five environments.

| | | | Pearson's correlations, N = 304 | | | | |
|---|---|---|---|---|---|---|---|
| **Trait†** | **Mean** | **SD‡** | **HD** | **HT** | **MAT** | **YLD** | **TWT** |
| HD | 56.3 | 1.9 | 1 | | | | |
| HT | 93.0 | 5.4 | 0.41*** | 1 | | | |
| MAT | 93.8 | 1.9 | 0.87*** | 0.35*** | 1 | | |
| YLD | 4278.0 | 457.8 | 0.12* | 0.09NS | 0.05NS | 1 | |
| TWT | 79.7 | 1.2 | 0.22*** | 0.01NS | 0.31*** | 0.35*** | 1 |
| TKW | 35.2 | 2.5 | -0.19*** | 0.29*** | -0.11NS | 0.06NS | -0.04NS |

*, **, ***, Significant at the 0.05, 0.01 and 0.001 probability level, respectively, and *NS* not significant

†HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight

‡SD: Standard deviation of the trait

### 7.3.2 Twenty-three QTL were identified for six agronomic traits

The number of markers used for QTL mapping varied from 23 on 4D to 96 on 5B, with an average of 58 markers per chromosome. The total map length across the 21 chromosomes spanned 3,526 cM. The range of genetic distance between adjacent SNPs varied from 0.04 to 52.3 cM with a mean of 2.9 cM (Appendix F). Using these data, a total of 23 QTL were identified based on the average phenotypic data across the five environments (Fig. 7-2). Six QTL were identified for heading date and test weight, four QTL for maturity, three QTL for thousand-kernel weight, and two QTL for plant height and grain yield (Fig. 7-2).

The QTL associated with heading date were mapped at 128.6 cM on chromosome 2B (*QHD.usw-2B*), 19.9 cM on 2D (*QHd.usw-2D*), 106.6 cM on 4A (*QHd.usw-4A*), 90 cM on 4B (*QHd.usw-4B*), 162 cM on 5B (*QHd.usw-5B*), and 97 cM on 7D (*QHd.usw-7D.1*) (Fig. 7-2 and Table 7-2). Each of these QTL explained from 1.8 to 19.2% of the phenotypic variance and together explained 49.4% of the variance in heading date across environments (Table 7-2). *QHd.usw-2D* was also detected in each of the five environments and explained from 9.5 to 22.8% of the phenotypic variance in each environment. Similarly, *QHd.usw-4A* and *QHd.usw-7D.1* were detected in all environments except Rosthern 2015 and explained from 2.9 to 4.7% and 8.8 to 15.5% of the phenotypic variance in each environment, respectively. *QHd.usw-2B* and *QHd.usw-5B* were detected in Kernen 2016, each explaining 6.2 and 5.1% of the phenotypic variance. *QHd.usw-4B* was also detected in Kernen 2016 and Rosthern 2016 and explained 2.4 and 9.4% of the phenotypic variance in each environment, respectively. Moreover, four environment specific QTL were detected on chromosome 1B (*QHd.usw-1B.1* and *QHd.usw-1B.2*), 7A (*QHd.usw-7A*), and 7D (*QHd.usw-7D.2*) (Appendix G). Each of these QTL explained from 2.4 to 6.6% of the phenotypic variance in each environment.

The QTL for test weight were mapped at 124 cM on chromosome 2A (*QTwt.usw-2A*), 120 cM on 2B (*QTwt.usw-2B.2*), 59 cM on 4B (*QTwt.usw-4B*), 71.9 cM on 6A (*QTwt.usw-6A*), 72 cM on 6B (*QTwt.usw-6B*), and 70 cM on 7B (*QTwt.usw-7B*) (Fig. 7-2 and Table 7-2). Each of these QTL accounted for 3.3 to 7.9% of the phenotypic variance and the six QTL together explained 34.6% of the variance in test weight across the five environments. For the analyses of individual environments, *QTwt.usw-2A*, *QTwt.usw-6A*, and *QTwt.usw-6B* were detected in one environment and explained from 3.5 to 9.4% of the phenotypic variance. *QTwt.usw-2B.2* was detected in two

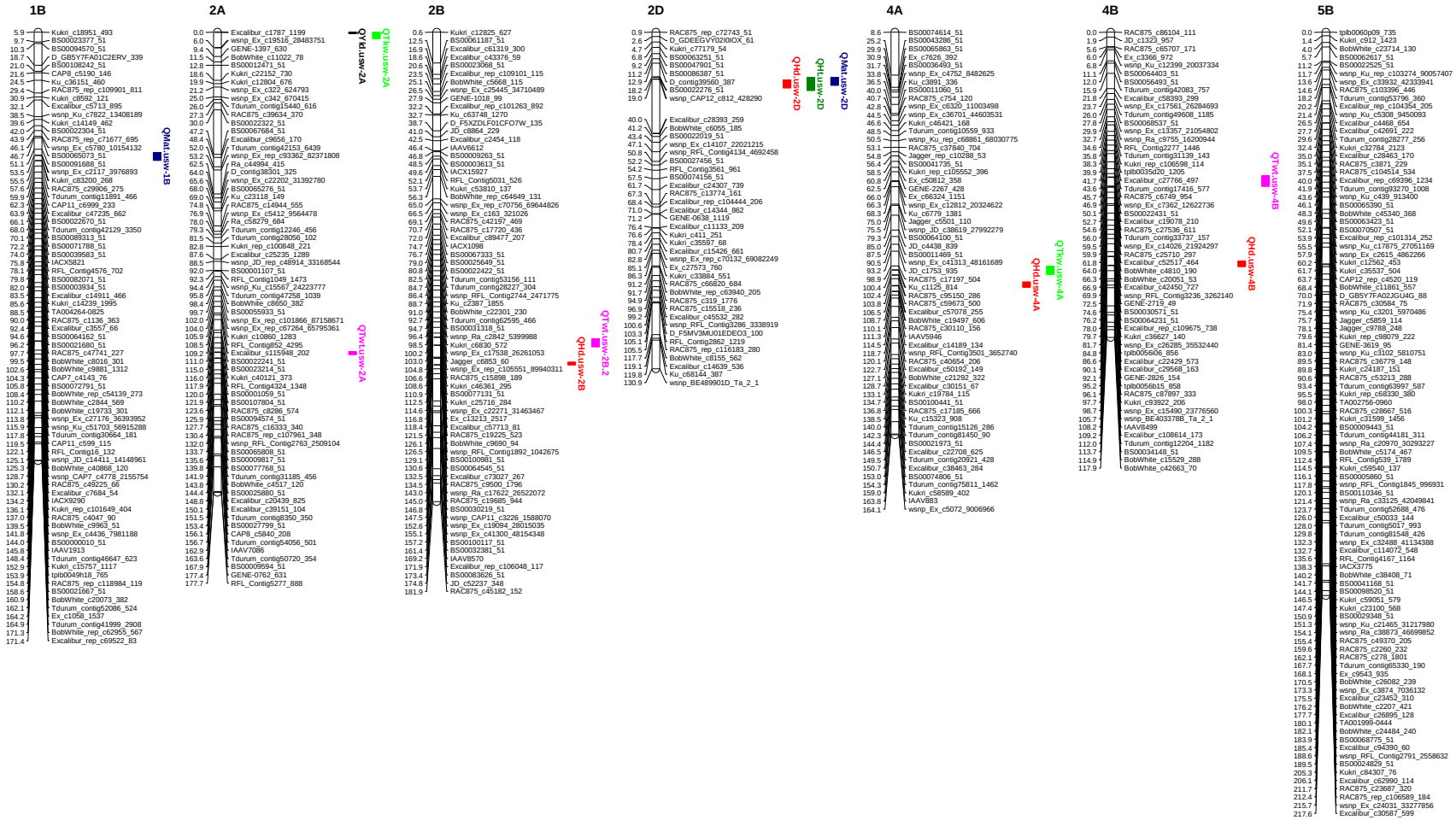



**Fig. 7-2.** The distribution of QTL identified for six agronomic traits based on the combined data across five environments. QTL are displayed on the right side of each chromosome with vertical bars showing the QTL confidence interval defined by 1-LOD drop. The QTL identified for each trait are indicated in different colors.

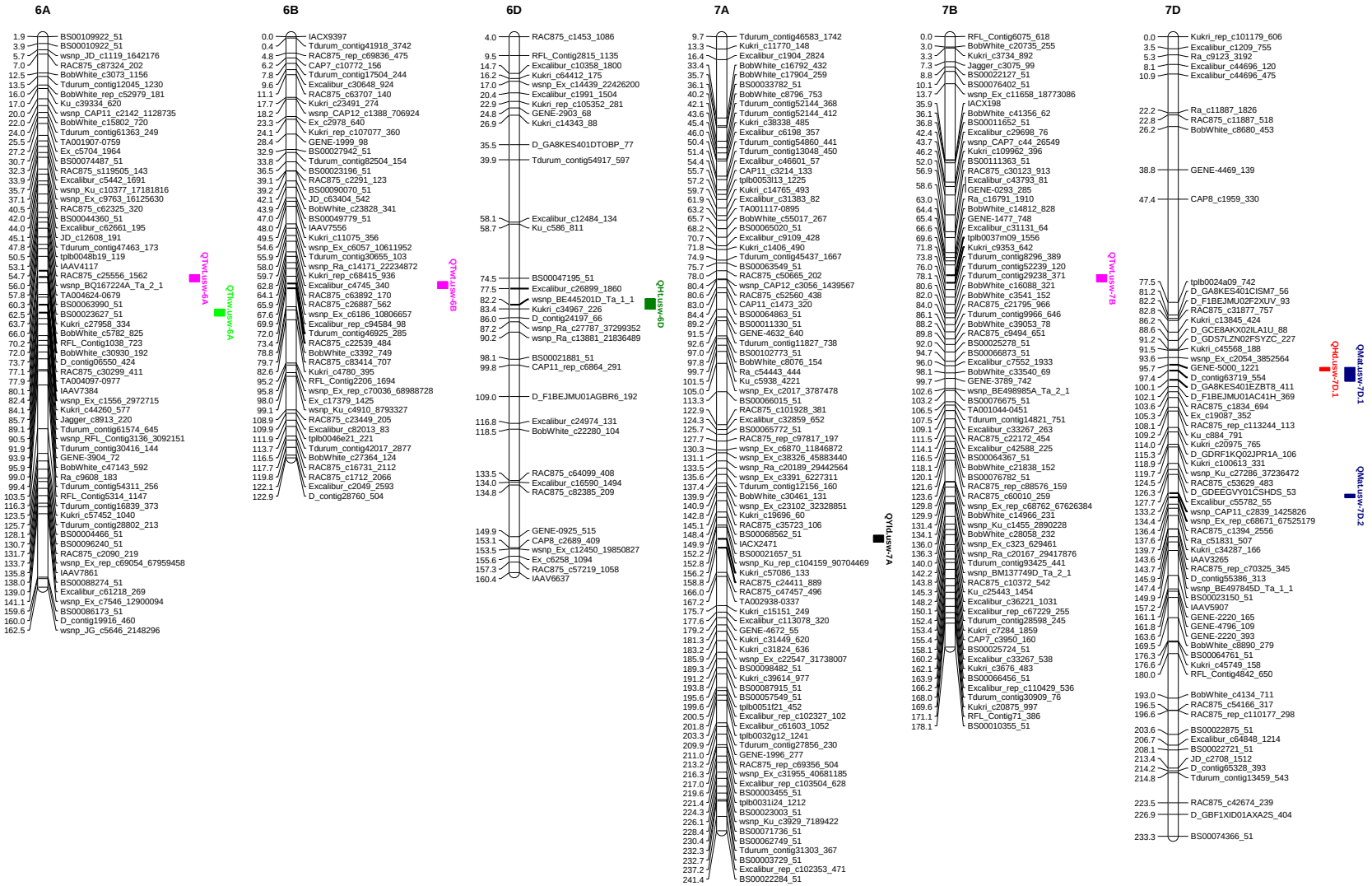



**Fig. 7-2.** Continued

**Table 7-2.** Summary of QTL identified for six agronomic traits based on 304 RILs evaluated across five environments. QTL analyses were conducted using LS-means of each environment and averaged (combined) across all environments.

| QTL | Trait† | Environment | Chromosome | Position (cM) | Confidence Interval | Left Marker | Right Marker | LOD | $R^2$ (%) | Add‡ |
|---|---|---|---|---|---|---|---|---|---|---|
| *QHd.usw-2B* | HD | Kernen-16 | 2B | 128.6 | 128.1 – 129.1 | wsnp_RFL_Contig1892_1042675 | BS00100981_51 | 6.9 | 6.2 | -0.7 |
| *QHd.usw-2B* | HD | Combined | 2B | 128.6 | 128.1 – 129.1 | wsnp_RFL_Contig1892_1042675 | BS00100981_51 | 10.9 | 8.4 | -0.7 |
| *QHd.usw-2D* | HD | Kernen-14 | 2D | 20.9 | 18.4 – 24.4 | wsnp_CAP12_c812_428290 | Excalibur_c28393_259 | 15.8 | 11.5 | 0.8 |
| *QHd.usw-2D* | HD | Kernen-15 | 2D | 20.9 | 19.4 – 23.4 | wsnp_CAP12_c812_428290 | Excalibur_c28393_259 | 29.9 | 21.1 | 1.3 |
| *QHd.usw-2D* | HD | Kernen-16 | 2D | 20.9 | 18.4 – 24.4 | wsnp_CAP12_c812_428290 | Excalibur_c28393_259 | 19.5 | 11.5 | 1.0 |
| *QHd.usw-2D* | HD | Rosthern-15 | 2D | 22.9 | 20.4 – 24.4 | wsnp_CAP12_c812_428290 | Excalibur_c28393_259 | 32.3 | 22.8 | 1.5 |
| *QHd.usw-2D* | HD | Rosthern-16 | 2D | 19.9 | 18.4 – 22.4 | wsnp_CAP12_c812_428290 | Excalibur_c28393_259 | 19.5 | 9.5 | 1.0 |
| *QHd.usw-2D* | HD | Combined | 2D | 19.9 | 19.4 – 22.4 | wsnp_CAP12_c812_428290 | Excalibur_c28393_259 | 32.9 | 19.2 | 1.1 |
| *QHd.usw-4A* | HD | Kernen-14 | 4A | 105.6 | 104.1 – 106.1 | RAC875_c59673_500 | Excalibur_c57078_255 | 7.6 | 4.7 | -0.6 |
| *QHd.usw-4A* | HD | Kernen-15 | 4A | 105.6 | 104.1 – 106.1 | RAC875_c59673_500 | Excalibur_c57078_255 | 5.7 | 2.9 | -0.6 |
| *QHd.usw-4A* | HD | Kernen-16 | 4A | 106.6 | 105.1 – 107.1 | Excalibur_c57078_255 | BobWhite_c19497_606 | 9.1 | 4.1 | -0.7 |
| *QHd.usw-4A* | HD | Rosthern-16 | 4A | 104.6 | 103.1 – 106.1 | RAC875_c59673_500 | Excalibur_c57078_255 | 9.6 | 4.4 | -0.8 |
| *QHd.usw-4A* | HD | Combined | 4A | 106.6 | 105.1 – 107.1 | Excalibur_c57078_255 | BobWhite_c19497_606 | 9.2 | 4.0 | -0.6 |
| *QHd.usw-4B* | HD | Kernen-16 | 4B | 104 | 100.5 – 107.5 | wsnp_Ex_c15490_23776560 | wsnp_BE403378B_Ta_2_1 | 4.9 | 2.4 | -0.5 |
| *QHd.usw-4B* | HD | Rosthern-16 | 4B | 98 | 97.5 – 98.5 | Kukri_c93922_206 | wsnp_Ex_c15490_23776560 | 6.8 | 9.4 | 1.0 |
| *QHd.usw-4B* | HD | Combined | 4B | 90 | 88.5 – 90.5 | Excalibur_c22429_573 | Excalibur_c29568_163 | 4.2 | 1.8 | -0.3 |
| *QHd.usw-5B* | HD | Kernen-16 | 5B | 162 | 161.5 – 162.5 | RAC875_c2260_232 | RAC875_c278_1801 | 10.2 | 5.1 | -0.7 |
| *QHd.usw-5B* | HD | Combined | 5B | 162 | 161.5 – 162.5 | RAC875_c2260_232 | RAC875_c278_1801 | 8.2 | 4.2 | -0.6 |
| *QHd.usw-7D.1* | HD | Kernen-14 | 7D | 94 | 92.5 – 94.5 | wsnp_Ex_c2054_3852564 | GENE-5000_1221 | 11.3 | 12.0 | -0.9 |
| *QHd.usw-7D.1* | HD | Kernen-15 | 7D | 98 | 96.5 – 100.5 | D_contig63719_554 | D_GA8KES401EZBT8_411 | 16.6 | 8.8 | 1.0 |
| *QHd.usw-7D.1* | HD | Kernen-16 | 7D | 97 | 96.5 – 98.5 | GENE-5000_1221 | D_contig63719_554 | 26.9 | 14.5 | 1.3 |
| *QHd.usw-7D.1* | HD | Rosthern-16 | 7D | 94 | 92.5 – 94.5 | wsnp_Ex_c2054_3852564 | GENE-5000_1221 | 20.8 | 15.5 | -1.3 |
| *QHd.usw-7D.1* | HD | Combined | 7D | 97 | 96.5 – 97.5 | GENE-5000_1221 | D_contig63719_554 | 23.1 | 11.8 | 1.0 |

**Table 7-2.** Continued

| QTL | Trait† | Environment | Chromosome | Position (cM) | Confidence Interval | Left Marker | Right Marker | LOD | $R^2$ (%) | Add‡ |
|---|---|---|---|---|---|---|---|---|---|---|
| *QTwt.usw-2A* | TWT | Kernen-16 | 2A | 106 | 105.5 – 107.5 | Kukri_c10860_1283 | RFL_Contig852_4295 | 6.0 | 3.5 | -0.3 |
| *QTwt.usw-2A* | TWT | Combined | 2A | 124 | 123.5 – 124.5 | RAC875_c8286_574 | BS00094574_51 | 6.2 | 5.8 | -0.4 |
| *QTwt.usw-2B.2* | TWT | Kernen-14 | 2B | 114.6 | 114.1 – 115.1 | wsnp_Ex_c22271_31463467 | Ex_c13213_2517 | 5.1 | 4.2 | 0.2 |
| *QTwt.usw-2B.2* | TWT | Rosthern-16 | 2B | 114.6 | 114.1 – 115.1 | wsnp_Ex_c22271_31463467 | Ex_c13213_2517 | 7.3 | 5.7 | 0.4 |
| *QTwt.usw-2B.2* | TWT | Combined | 2B | 120.6 | 119.1 – 122.1 | Excalibur_c57713_81 | RAC875_c19225_523 | 4.8 | 3.3 | 0.3 |
| *QTwt.usw-4B* | TWT | Kernen-15 | 4B | 65 | 64.5 – 66.5 | BobWhite_c4810_190 | BobWhite_c20051_53 | 5.2 | 5.5 | 0.3 |
| *QTwt.usw-4B* | TWT | Kernen-16 | 4B | 49 | 47.5 – 50.5 | wsnp_Ex_c7362_12622736 | BS00022431_51 | 16.0 | 10.4 | -0.5 |
| *QTwt.usw-4B* | TWT | Rosthern-15 | 4B | 62 | 61.5 – 62.5 | Excalibur_c52517_464 | BobWhite_c4810_190 | 5.3 | 4.2 | -0.4 |
| *QTwt.usw-4B* | TWT | Rosthern-16 | 4B | 59 | 55.5 – 59.5 | Tdurum_contig33737_157 | wsnp_Ex_c14026_21924297 | 13.4 | 10.4 | 0.5 |
| *QTwt.usw-4B* | TWT | Combined | 4B | 59 | 55.5 – 59.5 | Tdurum_contig33737_157 | wsnp_Ex_c14026_21924297 | 13.3 | 7.9 | 0.4 |
| *QTwt.usw-6A* | TWT | Rosthern-16 | 6A | 82.9 | 82.4 – 83.4 | wsnp_Ex_c1556_2972715 | Kukri_c44260_577 | 8.7 | 9.4 | -0.5 |
| *QTwt.usw-6A* | TWT | Combined | 6A | 71.9 | 71.4 – 73.4 | RFL_Contig1038_723 | BobWhite_c30930_192 | 9.3 | 7.2 | 0.4 |
| *QTwt.usw-6B* | TWT | Kernen-16 | 6B | 73 | 71.5 – 75.5 | Tdurum_contig46925_285 | RAC875_c22539_484 | 8.9 | 5.3 | 0.3 |
| *QTwt.usw-6B* | TWT | Combined | 6B | 72 | 71.5 – 73.5 | Tdurum_contig46925_285 | RAC875_c22539_484 | 9.2 | 5.3 | 0.3 |
| *QTwt.usw-7B* | TWT | Kernen-16 | 7B | 74 | 73.5 – 74.5 | Tdurum_contig8296_389 | Tdurum_contig52239_120 | 7.1 | 6.5 | 0.4 |
| *QTwt.usw-7B* | TWT | Rosthern-15 | 7B | 71 | 69.5 – 71.5 | tplb0037m09_1556 | Kukri_c9353_642 | 7.7 | 4.5 | -0.5 |
| *QTwt.usw-7B* | TWT | Kernen-16 | 7B | 52 | 50.5 – 52.5 | BS00111363_51 | RAC875_c30123_913 | 6.9 | 4.0 | -0.4 |
| *QTwt.usw-7B* | TWT | Combined | 7B | 70 | 69.5 – 71.5 | tplb0037m09_1556 | Kukri_c9353_642 | 8.5 | 5.1 | -0.4 |
| *QMat.usw-1B* | MAT | Rosthern-15 | 1B | 53.9 | 53.4 – 55.4 | wsnp_Ex_c2117_3976893 | Kukri_c83200_268 | 11.4 | 5.2 | 0.8 |
| *QMat.usw-1B* | MAT | Combined | 1B | 53.9 | 52.4 – 55.4 | wsnp_Ex_c2117_3976893 | Kukri_c83200_268 | 4.4 | 3.0 | 0.5 |
| *QMat.usw-2D* | MAT | Kernen-14 | 2D | 21.9 | 20.4 – 26.4 | wsnp_CAP12_c812_428290 | Excalibur_c28393_259 | 10.0 | 8.1 | 1.1 |
| *QMat.usw-2D* | MAT | Kernen-15 | 2D | 19.9 | 18.4 – 23.4 | wsnp_CAP12_c812_428290 | Excalibur_c28393_259 | 14.3 | 13.6 | 1.1 |
| *QMat.usw-2D* | MAT | Kernen-16 | 2D | 18.9 | 18.4 – 21.4 | BS00022276_51 | wsnp_CAP12_c812_428290 | 10.9 | 4.5 | 0.6 |

**Table 7-2.** Continued

| QTL | Trait† | Environment | Chromosome | Position (cM) | Confidence Interval | Left Marker | Right Marker | LOD | $R^2$ (%) | Add‡ |
|---|---|---|---|---|---|---|---|---|---|---|
| *QMat.usw-2D* | MAT | Rosthern-15 | 2D | 18.9 | 18.4 – 22.4 | BS00022276_51 | wsnp_CAP12_c812_428290 | 13.5 | 7.6 | 0.8 |
| *QMat.usw-2D* | MAT | Rosthern-16 | 2D | 18.9 | 18.4 – 23.4 | BS00022276_51 | wsnp_CAP12_c812_428290 | 7.3 | 11.8 | 0.6 |
| *QMat.usw-2D* | MAT | Combined | 2D | 18.9 | 18.4 – 21.4 | BS00022276_51 | wsnp_CAP12_c812_428290 | 17.6 | 14.7 | 0.9 |
| *QMat.usw-7D.1* | MAT | Kernen-14 | 7D | 94 | 93.5 – 94.5 | wsnp_Ex_c2054_3852564 | GENE-5000_1221 | 11.2 | 11.7 | -1.3 |
| *QMat.usw-7D.1* | MAT | Kernen-15 | 7D | 97 | 96.5 – 100.5 | GENE-5000_1221 | D_contig63719_554 | 10.3 | 9.1 | 1.0 |
| *QMat.usw-7D.1* | MAT | Rosthern-15 | 7D | 104 | 103.5 – 104.5 | RAC875_c1834_694 | Ex_c19087_352 | 13.0 | 6.2 | 0.8 |
| *QMat.usw-7D.1* | MAT | Combined | 7D | 98 | 96.5 – 100.5 | D_contig63719_554 | D_GA8KES401EZBT8_411 | 14.7 | 10.7 | 0.9 |
| *QMat.usw-7D.2* | MAT | Kernen-16 | 7D | 132 | 131.5 – 132.5 | Excalibur_c55782_55 | wsnp_CAP11_c2839_1425826 | 6.2 | 6.8 | -0.8 |
| *QMat.usw-7D.2* | MAT | Rosthern-15 | 7D | 134 | 133.5 – 135.5 | wsnp_CAP11_c2839_1425826 | wsnp_Ex_rep_c68671_67525179 | 6.0 | 2.9 | -0.5 |
| *QMat.usw-7D.2* | MAT | Combined | 7D | 134 | 133.5 – 134.5 | wsnp_CAP11_c2839_1425826 | wsnp_Ex_rep_c68671_67525179 | 4.6 | 3.4 | -0.4 |
| *QTkw.usw-2A* | TKW | Kernen-14 | 2A | 1 | 0 – 2.5 | Excalibur_c1787_1199 | wsnp_Ex_c19516_28483751 | 7.2 | 6.4 | -0.9 |
| *QTkw.usw-2A* | TKW | Kernen-16 | 2A | 1 | 0 – 2.5 | Excalibur_c1787_1199 | wsnp_Ex_c19516_28483751 | 9.5 | 9.1 | -0.9 |
| *QTkw.usw-2A* | TKW | Combined | 2A | 1 | 0 – 2.5 | Excalibur_c1787_1199 | wsnp_Ex_c19516_28483751 | 6.8 | 7.3 | -0.7 |
| *QTkw.usw-4A* | TKW | Kernen-15 | 4A | 100.6 | 99.1 – 102.1 | Ku_c1125_814 | RAC875_c95150_286 | 6.9 | 9.3 | -0.9 |
| *QTkw.usw-4A* | TKW | Kernen-16 | 4A | 100.6 | 99.1 – 102.1 | Ku_c1125_814 | RAC875_c95150_286 | 6.2 | 4.3 | -0.7 |
| *QTkw.usw-4A* | TKW | Rosthern-15 | 4A | 103.6 | 103.1 – 105.1 | RAC875_c95150_286 | RAC875_c59673_500 | 9.1 | 6.8 | 1.2 |
| *QTkw.usw-4A* | TKW | Combined | 4A | 100.6 | 99.1 – 102.1 | Ku_c1125_814 | RAC875_c95150_286 | 7.8 | 5.9 | -0.8 |
| *QTkw.usw-6A* | TKW | Kernen-14 | 6A | 81.9 | 81.4 – 83.4 | IAAV7384 | wsnp_Ex_c1556_2972715 | 5.2 | 8.5 | 1.0 |
| *QTkw.usw-6A* | TKW | Kernen-16 | 6A | 82.9 | 81.4 – 83.4 | wsnp_Ex_c1556_2972715 | Kukri_c44260_577 | 8.5 | 8.6 | 0.9 |
| *QTkw.usw-6A* | TKW | Rosthern-15 | 6A | 81.9 | 81.4 – 82.4 | IAAV7384 | wsnp_Ex_c1556_2972715 | 6.4 | 8.7 | 1.3 |
| *QTkw.usw-6A* | TKW | Rosthern-16 | 6A | 81.9 | 81.4 – 83.4 | IAAV7384 | wsnp_Ex_c1556_2972715 | 6.9 | 11.4 | 1.3 |
| *QTkw.usw-6A* | TKW | Combined | 6A | 81.9 | 81.4 – 83.4 | IAAV7384 | wsnp_Ex_c1556_2972715 | 3.8 | 7.8 | 0.8 |

**Table 7-2.** Continued

| QTL | Trait† | Environment | Chromosome | Position (cM) | Confidence Interval | Left Marker | Right Marker | LOD | $R^2$ (%) | Add‡ |
|---|---|---|---|---|---|---|---|---|---|---|
| *QHt.usw-2D* | HT | Kernen-14 | 2D | 19.9 | 18.4 – 24.4 | wsnp_CAP12_c812_428290 | Excalibur_c28393_259 | 8.7 | 11.2 | 2.5 |
| *QHt.usw-2D* | HT | Kernen-15 | 2D | 18.9 | 18.4 – 22.4 | BS00022276_51 | wsnp_CAP12_c812_428290 | 6.9 | 7.8 | 2.4 |
| *QHt.usw-2D* | HT | Kernen-16 | 2D | 18.9 | 18.4 – 26.4 | BS00022276_51 | wsnp_CAP12_c812_428290 | 5.0 | 4.9 | 1.7 |
| *QHt.usw-2D* | HT | Rosthern-15 | 2D | 20.9 | 18.4 – 26.4 | wsnp_CAP12_c812_428290 | Excalibur_c28393_259 | 8.0 | 10.4 | 2.4 |
| *QHt.usw-2D* | HT | Combined | 2D | 19.9 | 18.4 – 23.4 | wsnp_CAP12_c812_428290 | Excalibur_c28393_259 | 11.5 | 11.2 | 2.0 |
| *QHt.usw-6D* | HT | Kernen-14 | 6D | 83 | 80.5 – 83.5 | wsnp_BE445201D_Ta_1_1 | Kukri_c34967_226 | 6.1 | 6.9 | 2.3 |
| *QHt.usw-6D* | HT | Rosthern-15 | 6D | 83 | 81.5 – 83.5 | wsnp_BE445201D_Ta_1_1 | Kukri_c34967_226 | 4.9 | 5.1 | 2.0 |
| *QHt.usw-6D* | HT | Combined | 6D | 82 | 80.5 – 83.5 | Excalibur_c26899_1860 | wsnp_BE445201D_Ta_1_1 | 6.6 | 6.1 | 1.7 |
| *QYld.usw-2A* | YLD | Kernen-16 | 2A | 0 | 0 – 1.5 | Excalibur_c1787_1199 | wsnp_Ex_c19516_28483751 | 13.7 | 18.0 | 284.0 |
| *QYld.usw-2A* | YLD | Rosthern-16 | 2A | 0 | 0 – 1.5 | Excalibur_c1787_1199 | wsnp_Ex_c19516_28483751 | 5.3 | 5.3 | 169.6 |
| *QYld.usw-2A* | YLD | Combined | 2A | 0 | 0 – 0.5 | Excalibur_c1787_1199 | wsnp_Ex_c19516_28483751 | 9.5 | 10.4 | 168.7 |
| *QYld.usw-7A* | YLD | Rosthern-16 | 7A | 155.7 | 155.2 – 157.2 | wsnp_Ku_rep_c104159_90704469 | Kukri_c57086_133 | 6.8 | 11.5 | -248.9 |
| *QYld.usw-7A* | YLD | Combined | 7A | 156.7 | 155.2 – 157.2 | Kukri_c57086_133 | RAC875_c24411_889 | 5.4 | 10.6 | -170.0 |

†HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight

‡Add: Additive effect.

environments and accounted for 4.2 and 5.7% of the phenotypic variance in each environment. *QTwt.usw-4B* was detected in four out of the five environments and explained from 4.2 to 10.4% of the phenotypic variance, while *QTwt.usw-7B* was detected in three environments and accounted for 4 to 6.5% of the phenotypic variance in each environment (Table 7-2). Moreover, several environment specific QTL were identified for test weight on chromosomes 1D (*QTwt.usw-1D*), 2B (*QTwt.usw-2B.1*), 2D (*QTwt.usw-2D*), 3A (*QTwt.usw-3A*), 3D (*QTwt.usw-3D.1* and *QTwt.usw-3D.2*), and 7D (*QTwt.usw-7D*), which explained from 3.4 to 9.2 of the phenotypic variance in each environment (Appendix G).

The QTL associated with maturity were mapped at 53.9 cM on chromosome 1B (*QMat.usw-1B*), 18.9 cM on 2D (*QMat.usw-2D*), 98 cM (*QMat.usw-7D.1*) and 134 cM (*QMat.usw-7D.2*) on 7D (Fig. 7-2 and Table 7-2). Each of these QTL explained 3 to 14.7% of the phenotypic variance across all environments and together explained 31.7% of the variation in maturity across environments (Table 7-2). *QMat.usw-2D* and *QMat.usw-7D.1* were the most stable and together explained 25.4% of the variance in maturity across environments. *QMat.usw-2D* was detected in all five environments and accounted for 4.5 to 13.6% of the phenotypic variance in each environment (Table 7-2). Two QTL were detected on chromosome 7D separated by 36 cM between the QTL peaks. *QMat.usw-7D.1* was detected in three environments (Kernen 2014, Kernen 2015 and Rosthern 2015) and explained from 6.2 to 11.7% of the phenotypic variance in each environment, while *QMat.usw-7D.2* was detected in two environments (Kernen 2016 and Rosthern 2015) and explained 6.8 and 2.9% of the phenotypic variance, respectively. *QMat.usw-1B* was detected in only one environment and explained 5.2% of the phenotypic variance. Moreover, several environment specific QTL were identified on 3A (*QMat.usw-3A*), 5A (*QMat.usw-5A*), 5B (*QMat.usw-5B*), 7A (*QMat.usw-7A*), and 7B (*QMat.usw-7B*), each explaining from 3.7 to 9.3% of the phenotypic variance in each environment (Appendix G).

For thousand-kernel weight, the three QTL were mapped at 1 cM on chromosome 2A (*QTkw.usw-2A*), 100.6 cM on 4A (*QTkw.usw-4A*), and 81.9 cM on 6A (*QTkw.usw-6A*) (Fig. 7-2 and Table 7-2). Each QTL explained from 5.9 to 7.8% of the phenotypic variance and the three QTL together explained 21% of the variation in thousand-kernel weight across the five environments. Within individual environments, *QTkw.usw-2A* was detected in two environments at the same confidence interval and explained 6.4 and 9.1% of the phenotypic variance in each environment. *QTkw.usw-4A* was detected in three environments and accounted for 4.3 to 9.3% of

the phenotypic variance. *QTkw.usw-6A* was detected in all environments except Kernen 2015 and explained from 8.5 to 11.4% of the phenotypic variance in each environment. Moreover, several environment specific QTL were detected on 1A (*QTkw.usw-1A*), 1B (*QTkw.usw-1B*), 2D (*QTkw.usw-2D*), 5A (*QTkw.usw-5A*), 5B (*QTkw.usw-5B.1* and *QTkw.usw-5B.2*), 5D (*QTkw.usw-5D*), and 7A (*QTkw.usw-7A*), which explained from 3.3 to 7.9% of the phenotypic variance in each environment (Appendix G).

Two independent QTL for plant height were mapped at 19.9 cM on 2D (*QHt.usw-2D*), and 82 cM on 6D (*QHt.usw-6D*), which explained 11.2 and 6.1% of the phenotypic variance across the five environments, respectively (Fig. 7-2 and Table 7-2). *QHt.usw-2D* was also detected in all environments, except Rosthern 2016, and accounted for 4.9 to 11.2% of the phenotypic variance in each environment. *QHt.usw-6D* was detected in two environments (Kernen 2014 and Rosthern 2016) and accounted for 6.9 and 5.1% of the phenotypic variance, respectively (Table 7-2). Moreover, two environment specific QTL were detected for plant height on 1D (*QHt.usw-1D*) and 3B (*QHt.usw-3B*), each explaining 4.5 and 4.7% of the phenotypic variance, respectively (Appendix G).

When average grain yield across the five environments was considered, QTL were identified on the distal end of 2AS (*QYld.usw-2A*) and at 156.7 cM on 7A (*QYld.usw-7A*), which explained 10.4 and 10.6% of the phenotypic variance across the five environments, respectively (Fig. 7-2 and Table 7-2). Together, the two QTL explained 21% of the variation in yield across environments. *QYld.usw-2A* was detected in two out of the five environments (Kernen 2016 and Rosthern 2016) and accounted for 18 and 5.3% of the phenotypic variance, respectively, while *QYld.usw-7A* was detected only in one environment (Rosthern 2016) and explained 11.5% of the phenotypic variance. Moreover, four environment specific QTL were identified for grain yield on 1A (*QYld.usw-1A*), 1D (*QYld.usw-1D*), 2B (*Qyld.usw-2B*), and 5A (*Qyld.usw-5A*), each explaining from 2.1 to 9.8% of the phenotypic variance in each environment (Appendix G).

### 7.3.3 Coincident QTL and trait relationships

Three chromosomal regions on chromosome 2A (0 – 2.5 cM), 2D (18.4 – 23.4 cM) and 7D (96.5 – 100.5 cM) co-localized QTL for two to three traits. The region on 2A harboured QTL for grain yield (*QYld.usw-2A*) and thousand-kernel weight (*QTkw.usw-2A*). These QTL were flanked by the same set of markers in the combined data as well as in two individual environments (Table

7-2). The region on chromosome 2D (18.4 – 23.6 cM) harboured the strongest QTL for heading date (*QHd.usw-2D*), maturity (*QMat.usw-2D*) and plant height (*QHt.usw-2D*), which explained 19.2, 14.7 and 11.2% of the variation in these traits, respectively. *QHd.usw-2D* and *QMat.usw-2D* were detected in all five environments, while *QHt.usw-2D* was detected in four out of the five environments. The region on 7D (96.5 – 100.5 cM) also harboured QTL for heading date (*QHd.usw-7D.1*) and maturity (*QMat.usw-7D.1*) that explained 11.8 and 10.7% of the variation in these traits, respectively. *QHd.usw-7D.1* and *QMat.usw-7D.1* were detected in the combined data as well as four and three individual environments, respectively.

Some traits with co-localized QTL were correlated with each other across environments, which suggests that the genetic mechanisms underlying these traits are the same or are tightly linked (Table 7-1). For example, heading date, maturity, and plant height were correlated with each other (r = 0.87 for heading date and maturity, r = 0.41 for heading date and plant height, and r = 0.35 for maturity and plant height) and shared a major QTL on chromosome 2D. Heading date and maturity shared a second stable QTL on chromosome 7D. For these correlated traits with a common QTL, either the same causal polymorphism underlies the identified QTL (pleiotropy) or the QTL underlying the traits are linked. In contrast, grain yield and thousand-kernel weight shared a QTL on chromosome 2A but there was no correlation between these traits (r = 0.06), which may indicate that the QTL controlling these traits are different.

### 7.3.4 Relationship between QTL from this study and known QTL

In this study, QTL were mapped on chromosomes 2B, 2D, 4A, 4B, 5B, and 7D for heading date and on 1B, 2D, and 7D for maturity. Previously, QTL for flowering time were reported in wheat on 1B, 1D, 2B, 2D, 3B, 4A, 5A, 6B, 7A, 7B, and 7D (Lin et al., 2008; Perez-Lara et al., 2016; Sourdille et al., 2000; Sourdille et al., 2003). Similarly, various QTL for maturity were mapped on 1B, 3B, 4A, 4D, 5A, 5B, 6B, 7A, 7B, and 7D based on Canadian spring wheat mapping populations (Cuthbert et al., 2008; McCartney et al., 2005). The QTL identified in this study may be related to flowering time and maturity genes or QTL identified from previous studies. The genetic control of flowering time in wheat is complex and is known to be controlled by vernalization response (*Vrn* genes), photoperiod response (*Ppd* genes), and earliness *per se* genes (*Eps* genes), which act together to determine the exact time of flowering and adaptation of the genotype for flowering under particular environmental conditions (Snape et al., 2001; Worland,

1996). These genes also have pleiotropic effects on other aspects of growth and developmental phases, such as tillering, stem elongation, heading, and ripening (Košner and Pánková, 1998; Snape et al., 2001). The most important vernalization response genes *VRN-A1*, *VRN-B1*, and *VRN-D1* were mapped to homologous positions on the long arms of group five chromosomes (Dubcovsky et al., 1998; Law et al., 1976), while the photoperiod response genes *Ppd-A1*, *Ppd-B1*, and *Ppd-D1* were mapped to homologous positions on the short arms of group two chromosomes (Scarth and Law, 1984; Worland et al., 1998). The earliness *per se* genes are located on several chromosomes in wheat and are known to determine flowering time independent of environmental stimuli (Worland, 1996). In this study, stable QTL were detected for heading date and maturity on the short arm of 2D which is known to carry the *Ppd-D1* gene. Moreover, a second stable QTL that explained 2.9 to 4.7% of the variance was also identified for heading date on chromosome 4A (*QHd.usw-4A*). Recently, a study based on a mapping population derived from a cross between 'CDC Teal' and 'CDC Go' reported a stable QTL for heading date on 4A, 11 cM away from the QTL reported in this study, suggesting that these QTL may be the same (Chen et al., 2015). QTL for heading date and maturity were also reported on 7D, less than 5 cM distance from the heading date (*QHd.usw-7D.1*) and maturity (*QMat.usw-7D.1*) QTL reported in this study, suggesting that these QTL may be the same (Cuthbert et al., 2008). Similarly, a QTL that explained 25.7% of the phenotypic variance was reported for maturity on the short arm of chromosome 7D based on a Canadian spring wheat mapping population (McCartney et al., 2005). Further study is required to determine the relationship between the QTL from this study and the genes and QTL reported in previous studies.

The plant height QTL identified in this study may also be related to known genes or QTL. There are several reduced height (*Rht*) genes that affect plant height in wheat. The three major *Rht* genes that are important in commercial wheat cultivars are *Rht-B1b* (*Rht1*), *Rht-D1b* (*Rht2*) and *Rht8* (Börner et al., 1997; Ellis et al., 2002; Worland et al., 1998; Worland, 1996). The *Rht-B1b* and *Rht-D1b* are insensitive to exogenous gibberellic acid (GA) and mapped near the centromere on the short arm of chromosomes 4B and 4D, respectively (Börner et al., 1997; Ellis et al., 2002), while *Rht8* is GA-responsive and is located on the short arm of chromosome 2D (Worland et al., 1998; Worland, 1996). The *Rht8* and *Ppd-D1* genes are closely linked in wheat (Worland et al., 1998). The *Ppd-D1* gene is also known to have pleiotropic height reducing effects (Worland et al., 1998; Worland et al., 2001). The co-localized QTL for heading date, maturity and plant height on

chromosome 2D might be due to the pleiotropic effect of the *Ppd-D1* gene. In this study, no QTL affecting plant height was detected on chromosomes 4B and 4D, but the two QTL detected on chromosomes 2D and 6D together explained 17.3% of the variation in height. Several QTL were reported for plant height on chromosomes 2D, 4B, 4D, 5B, 7A, 7B (Cuthbert et al., 2008). Additional research will be required to confirm the association between the QTL in this study with known height reducing genes and QTL.

Grain yield is a highly quantitative trait that is influenced by genetic factors, environmental factors and their interactions. Quantitative trait loci were reported for grain yield on almost every chromosome in wheat (Asif et al., 2015; Chen et al., 2015; Cuthbert et al., 2008; Kamran et al., 2013; McCartney et al., 2005; Perez-Lara et al., 2016; Zou et al., 2017). In this study, two QTL were detected for grain yield on chromosomes 2A and 7A based on the combined data. McCartney et al. (2005) reported a QTL for grain yield at 43 cM on chromosome 2A, while in this study QTL for grain yield was located on the distal end of 2AS. Zou et al. (2017) also reported a QTL for grain yield on 7AS, while the QTL identified in this study is on 7AL. This suggests that both QTL identified for grain yield may be different from the ones reported previously.

Some of the QTL detected for test weight were mapped to genomic regions that were not reported previously. In this study, QTL were detected for test weight on chromosomes 2A, 2B, 4B, 6A, 6B, and 7B based on the combined data. Numerous QTL were reported for test weight in wheat on 1A, 1B, 1D, 2B, 2D, 3B, 3D, 4D, 5A, 5B, 5D, 6B, 7B, and 7D based on Canadian spring wheat mapping populations (Asif et al., 2015; Chen et al., 2015; McCartney et al., 2005; Zou et al., 2017). Recently, Zou et al. (2017) reported a QTL for test weight on 2B, 11.6 cM away from the QTL reported in this study. McCartney et al. (2005) also reported QTL for test weight on 2B and 6B (7 - 24 cM). Most of the QTL detected for test weight were not in genomic regions reported previously, which indicates that these QTL may be new. Each of these QTL explained a small proportion of the phenotypic variance (3.3 to 7.9%), which suggests the need for further validation. QTL for thousand-kernel weight were detected on 2A, 4A, and 6A, some of which may be related to previously reported QTL. Previous studies in wheat have reported a QTL for thousand-kernel weight on 4A within 20 cM distance from the QTL detected in this study (Asif et al., 2015; Zou et al., 2017). Similarly, Zou et al. (2017) reported a QTL for thousand-kernel weight on 6A, less than 3 cM from the QTL detected in this study. McCartney et al. (2005) also reported a QTL for

thousand-kernel weight on 2A (45 - 71 cM), while the QTL detected in this study was located on the distal end of 2AS.

### 7.4 Conclusion

In this study, we used RILs to identify several QTL controlling important agronomic traits. A region on chromosomes 2D harboured stable QTL associated with heading date, maturity and plant height. Similarly, a region on chromosome 7D was consistently associated with heading date and maturity. Moreover, stable QTL, which were detected in more than two environments, were identified for heading date on chromosome 4A, test weight on chromosomes 4B and 7B, and thousand-kernel weight on chromosomes 4A and 6A. Markers linked to the QTL identified from this study can be used for MAS to transfer desirable alleles into elite germplasm. Markers significantly associated with these QTL will also be used to enhance genomic predictions in the SC.

# 8. Accuracy of Genomic and Phenotypic Predictions in Wheat Based on Different Cross-Validation Techniques

## 8.1 Introduction

In previous chapters, numerous GS models and model parameters were assessed for their ability to make accurate predictions for different traits in different environments. Genomic selection normally involves two steps (Meuwissen, 2009). The first step involves estimating the effects of genome-wide markers based on the genotypic and phenotypic information of a TP. The second step involves predicting GEBVs of the SC from genotypic data by multiplying the marker scores of each line by the marker effects estimated from the TP. Selection decisions will then be based on GEBVs predicted using genome-wide markers (Jannink et al., 2010). Prediction accuracy is usually assessed using a different set of individuals, commonly called a validation population that has phenotypic and genotypic data. This can be achieved by calculating GEBVs for individuals in the validation population and model prediction accuracy is reported as the correlation between GEBVs and actual phenotypes of the individuals in the validation population.

Several statistical models have been developed to estimate the genetic values of individuals that have been genotyped but not phenotyped (de los Campos et al., 2009a; de los Campos et al., 2009b; Gianola et al., 2006; Gianola and van Kaam, 2008; Meuwissen et al., 2001; Park and Casella, 2008; Whittaker et al., 2000). Many of the studies on GS to date focused on evaluating the prediction accuracy of these models either by systematically partitioning the same population into training and validation sets or using a k-fold cross-validation technique. The later technique involves randomly dividing the same population into ‘k’ mutually exclusive groups of approximately equal sizes. Then marker effects are estimated using the k-1 groups and the remaining group is used for validation. This process is repeated ‘k’ times so that each group is subsequently used as a validation set. These methods are useful to compare different statistical models and model parameters, but genomic predictions based on independent populations need to be evaluated for practical application of GS. In previous chapters, we used cross-validation methods that partitioned individuals of the same population into training and validation data sets, but this approach may have limited application in a breeding program because inferences are made on known populations that have already been phenotyped.

Ideally, data that is routinely generated in a breeding program can be used to train a GS model that can be used to estimate GEBVs of individuals within breeding populations. However,

such an application is often constrained by low levels of genetic relatedness between the TP and the SC. Bassi et al. (2016) indicated that using a distantly related TP reduces the accuracy of prediction because allelic combinations within the SC will not be represented adequately. The degree of genetic relationship between the training and validation populations is an important factor that affects the accuracy of GS prediction (Clark et al., 2012; Habier et al., 2010; Habier et al., 2007; Hayes et al., 2009; Riedelsheimer et al., 2013). Genetic relationships are influenced by generations of descent or population stratification (Asoro et al., 2011). Several studies showed that when unrelated training populations are used to make predictions, accuracies are often close to zero (Charmet et al., 2014; Clark et al., 2012; Crossa et al., 2014; Riedelsheimer et al., 2013; Windhausen et al., 2012). However, successful implementation of GS in crop breeding programs largely depends on its potential to accurately estimate GEBVs of individuals that are distantly related to the TP, which GS models often fail to achieve with sufficient accuracy.

Cross-validation techniques that do not resemble the actual application of GS in a breeding program may lead to inaccurate results. If the validation set is more closely related to the TP than to the SC, then the prediction accuracy will be overestimated (Wray et al., 2013). Cryptic relatedness is another factor that can inflate prediction accuracies even when known close relatives are excluded (Wray et al., 2013). Therefore, the design of the TP and cross-validation must resemble the ways genomic predictions will be used in practice (Daetwyler et al., 2013). Habier et al. (2007) showed that GS models utilize genetic relationships among individuals as well as information from LD between markers and QTL. However, there are still unanswered questions related to what measure of genetic relationship is appropriate and the extent of genetic relationship that is sufficient to design an optimal TP to get an acceptable level of accuracy for practical application of GS. Moreover, it is not known whether including parents in a TP composed of diverse cultivars is sufficient to create a level of relatedness that can lead to accurate genomic prediction in progenies. Therefore, the objectives of this study were to i) evaluate the effects of TP composition, cross-validation technique, and genetic relationship between the TP and SC on GS accuracy, and ii) compare genomic and phenotypic prediction accuracy.

## 8.2 Materials and Methods

### 8.2.1 Plant material and phenotypic data

The TP was composed of 231 hexaploid wheat varieties and advanced breeding lines. The plant materials, field experiments and phenotypes measured for this population are described in detail in Section 3.2.1. The SC were composed of 304 RILs, which were developed from three-way crosses (CDC Plentiful//Pasteur/CDC Utmost). The plant materials, field experiments and phenotypes measured for this population are described in detail in Section 7.2.1. For both populations, traits including heading date, plant height, maturity, grain yield, test weight, and thousand-kernel weight were measured. The phenotypic data were analyzed using ANOVA, as described in Chapters three and seven.

### 8.2.2 Genotypic data

A total of 17,887 and 16,115 polymorphic SNPs were obtained for the TP and SC, respectively. About 9K SNP markers were common to the two populations. The common set of markers was used to make across-population genomic predictions. Missing marker genotypes were imputed using the function 'A.mat' in the rrBLUP package as described previously (Endelman, 2011).

### 8.2.3 Statistical modelling and genomic prediction

Four different prediction scenarios were evaluated in this study: across-population genomic prediction, within-population genomic prediction, across-year genomic prediction for combined locations, and across-year genomic predictions within locations. The first prediction scenario involved across-population genomic predictions where GEBVs for the SC were predicted based on marker effects estimated from the TP. The effect of genetic relationship between the TP and the SC on model prediction accuracy was evaluated by excluding parents from the TP, including two parents in the TP, and including the two parents with 50 and 100 random lines from the SC in the TP. Moreover, the effect of genetic relatedness on genomic prediction accuracy was investigated by separating the SC into two groups based on their genomic relationship (kinship) to the parents included in the TP. Genomic relationship among lines was estimated based on marker genotypes using the EMMA algorithm within GAPIT (Lipka et al., 2012). The two groups were

classified as closely and distantly related to the two parents (CDC Utmost and CDC Plentiful) that were included in the TP and prediction was made separately in each group. For each training-validation combination, two statistical models, G-BLUP and BayesB, were used to estimate GEBVs of the SC. A total of 9,187 polymorphic SNPs that were common between the TP and SC were used to make across-population genomic predictions.

The second prediction scenario involved within-population genomic predictions using a fivefold cross-validation. The overall means of the SC across generations (F6 to F8) and environments were used for this analysis. In this case, the 304 lines were randomly divided into five groups. The four groups were combined to estimate marker effects and GEBV were predicted for the remaining group. This was repeated until each group was used as a validation set. Predictions were made using RR-BLUP, GS + de novo GWAS, G-BLUP, BayesB, BL and RKHS-KA models. These models were chosen because they have different assumptions that are appropriate for a range of genetic architectures. In the SC, 16,115 polymorphic SNPs were used to test within-population genomic prediction accuracies. To test significant differences in prediction accuracy among the evaluated models, the cross-validation results were analyzed using a one-way ANOVA with PROC MIXED procedure in SAS (SAS Institute, 2015), using fold as blocking factor. The LSMEANS procedure was used in SAS to determine differences between statistical models.

The third scenario consisted of making a single genomic prediction across years for combined locations. For this analysis, we performed different across-year tests: the SC evaluated in 2014 (F4:F6 generation) were used to predict phenotypes of the SC evaluated in 2015 (F4:F7 generation) and 2016 (F4:F8 generation), the SC evaluated in 2015 were used to predict phenotypes of SC evaluated in 2016, and SC from 2014 and 2015 were used simultaneously to predict phenotypes of SC evaluated in 2016. The Co-op population and SC were evaluated in similar environments in 2014 and 2015 and were combined and used as TP to make similar predictions as before. Genomic predictions were made across years by merging the Co-op and SC data in 2014 to predict SC in 2015 and 2016, the 2015 Co-op and SC were used to predict SC in 2016, and finally the Co-op and SC in 2014 and 2015 were used simultaneously to predict SC in 2016. The 2014 Co-op data are means of two replications from one location, 2014 SC data are adjusted means from one location, while the 2015 Co-op and SC, and 2016 SC data are LS-Means from two locations (Kernen and Rosthern).

The fourth prediction scenario consisted of a single genomic prediction across years separately for each location. The SC evaluated at Kernen in 2014 were used to predict the phenotypes of the SC evaluated at Kernen in 2015 and 2016, the SC evaluated at Kernen in 2015 were used to predict phenotypes of SC evaluated at Kernen in 2016, and the SC from Kernen 2014 and 2015 were used simultaneously to predict phenotypes of SC evaluated at Kernen in 2016. Similarly, the SC evaluated at Rosthern in 2015 were used to predict phenotypes of SC evaluated at Rosthern in 2016. The Co-op and SC data sets were combined as described in the third prediction scenario and used to make genomic prediction across years separately in each location. In prediction scenarios three and four, 16,115 SNP markers were used for predictions involving SC only, while 9,187 SNP markers common to both the SC and Co-op populations were used to make predictions involving combined SC and Co-op populations. Predictions in scenarios three and four were made using a reaction norm model (Equation 5.4) that incorporates the main effects of molecular markers and environments using covariance functions as implemented in Jarquín et al. (2014a).

The G-BLUP, BayesB, BL, RKHS-KA, and reaction norm models were fitted using the R package BGLR (Pérez and de los Campos, 2014). The default settings of BGLR and number of iterations were used as described in Chapter three. In prediction scenarios three and four, phenotypic prediction accuracy ($r_P$) was calculated as a correlation between observed phenotypes of the SC in the environments used for training and validating the model. The RR-BLUP and GS + de novo GWAS models were fitted using the R package rrBLUP (Endelman, 2011). For the GS + de novo GWAS model, single marker regression was conducted in each fold using the LS-means of each trait across environments using Windows QTL Cartographer, v2.5.011 (Wang et al. 2012). Single marker regression was performed based on a subset of 1,219 evenly spaced SNPs that were used for QTL analyses in Chapter seven. The *P*-values from the single marker regression were sorted from low to high and multiple testing correction was performed for all SNPs using FDR methods described in Chapter six. Up to three significant markers were fitted as fixed effects as described in Chapter three (Appendix H), while all the remaining markers were fitted as random effects.

## 8.3 Results and Discussion

### 8.3.1 Across-population genomic prediction

Genomic predictions were made across populations by estimating marker effects using the TP and predicting GEBVs for the SC. When none of the parents were included in the TP, prediction accuracy was very low for all traits; accuracies of 0.15 and 0.22 for heading date, 0.05 and -0.09 for plant height, 0.15 and 0.27 for maturity, 0.05 and 0.08 for grain yield, 0.19 and 0.18 for test weight, and 0.24 and 0.27 for thousand-kernel weight in G-BLUP and BayesB, respectively (Table 8-1). Including two of the parents in the TP did not consistently improve prediction accuracies across all traits in either G-BLUP or BayesB (Table 8-1). BayesB showed slightly higher accuracy than G-BLUP for all traits except plant height and test weight. Including 50 random lines from the SC in the TP resulted in 1.2 to 4.4-fold increase in accuracy for all traits compared to when only the parents were included (Table 8-1). The largest increase was observed for plant height where accuracy increased approximately from 0 to 0.3 in both methods. A further increase of the number of SC in the TP to 100 resulted in 1.8 to 4.6-fold increase in accuracy for all traits compared to when only the parents were included (Table 8-1). The low prediction accuracy when the SC were not added to the TP could be because of the distant relationship between these populations. When the populations are distantly related, they might have different QTL or different markers in LD with shared QTL. When the data from the populations are combined, markers and QTL that are shared between populations can be used to make more accurate predictions. BayesB was the best performing model for all traits when the number of SC in the TP increased from 50 to 100. For heading date, prediction accuracy in BayesB was more than twice that obtained with G-BLUP, while for maturity there was nearly 50% increase in accuracy in BayesB compared to G-BLUP (Table 8-1). One reason BayesB outperformed G-BLUP was because some traits, such as heading date and maturity, have large effect QTL in this population (Table 7-2). Lorenz et al. (2011) reported that variable selection methods, such as BayesB should be preferred over methods that induce homogenous shrinkage, such as RR-BLUP or G-BLUP for traits that are controlled by few QTL with large effects. Another explanation for the improved performance of BayesB is that it uses information from marker-QTL LD better than G-BLUP and is expected to yield higher accuracies when the TP and SC are distantly related (Habier et al., 2007). Habier et al. (2007) indicated that the accuracy of GEBVs depends on both genetic relationships among individuals and LD between markers and QTL. In the absence of close relationships, prediction accuracy is

driven by distant relationships that will be useful when there is strong LD in the population (Clark et al., 2012). The assumption of GS is that all QTL are in LD with one or more nearby markers (Meuwissen et al., 2001). However, when the populations are distantly related, markers that are in LD with QTL in one population may not be in LD in another population (de Roos et al., 2008; Goddard and Hayes, 2007). Estimation of marker effects across populations requires not only high LD in each population, but the same linkage phases between markers and QTLs in each population (Goddard and Hayes, 2007). This shows that the marker-QTL LD phase between the TP and SC is an important factor that affects the accuracy of prediction.

**Table 8-1.** Prediction accuracy based on Pearson's correlations between GEBVs estimated using two statistical models and trait phenotypes. Predictions were made for the selection candidates using a training population with varying degrees of genetic relationships.

| Training population† | Model | HD‡ | HT | MAT | YLD | TWT | TKW |
|---|---|---|---|---|---|---|---|
| Without parents | G-BLUP | 0.153 | 0.052 | 0.146 | 0.052 | 0.194 | 0.243 |
| | BayesB | 0.216 | -0.090 | 0.266 | 0.079 | 0.182 | 0.276 |
| With two parents | G-BLUP | 0.156 | 0.072 | 0.162 | 0.069 | 0.208 | 0.209 |
| | BayesB | 0.194 | -0.070 | 0.208 | 0.078 | 0.192 | 0.228 |
| With two parents and 50 SC | G-BLUP | 0.199 | 0.311 | 0.296 | 0.162 | 0.355 | 0.478 |
| | BayesB | 0.242 | 0.308 | 0.340 | 0.181 | 0.359 | 0.501 |
| With two parents and 100 SC | G-BLUP | 0.287 | 0.309 | 0.351 | 0.324 | 0.467 | 0.590 |
| | BayesB | 0.594 | 0.323 | 0.519 | 0.332 | 0.480 | 0.602 |

†SC: selection candidates

‡HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight.

Previous studies reported higher prediction accuracy when each selection candidate had at least one highly related line in the TP (Daetwyler et al., 2014). In this study, two of the three parents of the SC (CDC Utmost and CDC Plentiful) were included in the TP. The third parent (Pasteur), which has European origin and is more distantly related than the Canadian wheat lines, was not included in the TP. To investigate the effect of genetic relatedness on genomic prediction accuracy, the SC were divided into two groups based on their genomic relationship to the parents (Fig. 8-1). The first group (Group A) was composed of 121 lines that were closely related to Pasteur but distantly related to the two parents that were included in the TP. The second group (Group B) was composed of 183 lines that were closely related to the two parents in the TP. Predictions were made for the two groups separately using the marker effects estimated from the TP that included the parents. Accuracy was slightly higher when predictions were made for lines closely related to the parents in TP for all traits, except grain yield (Table 8-2). Overall, the prediction accuracy was very low for both groups even though two of the parents were included in the TP to enhance genomic relationships. This indicates that including parents in the TP may not ensure accurate prediction of GEBVs for progenies. Similar results have been reported by Windhausen et al. (2012), where prediction accuracies were close to zero even for crosses of lines included in the TP.

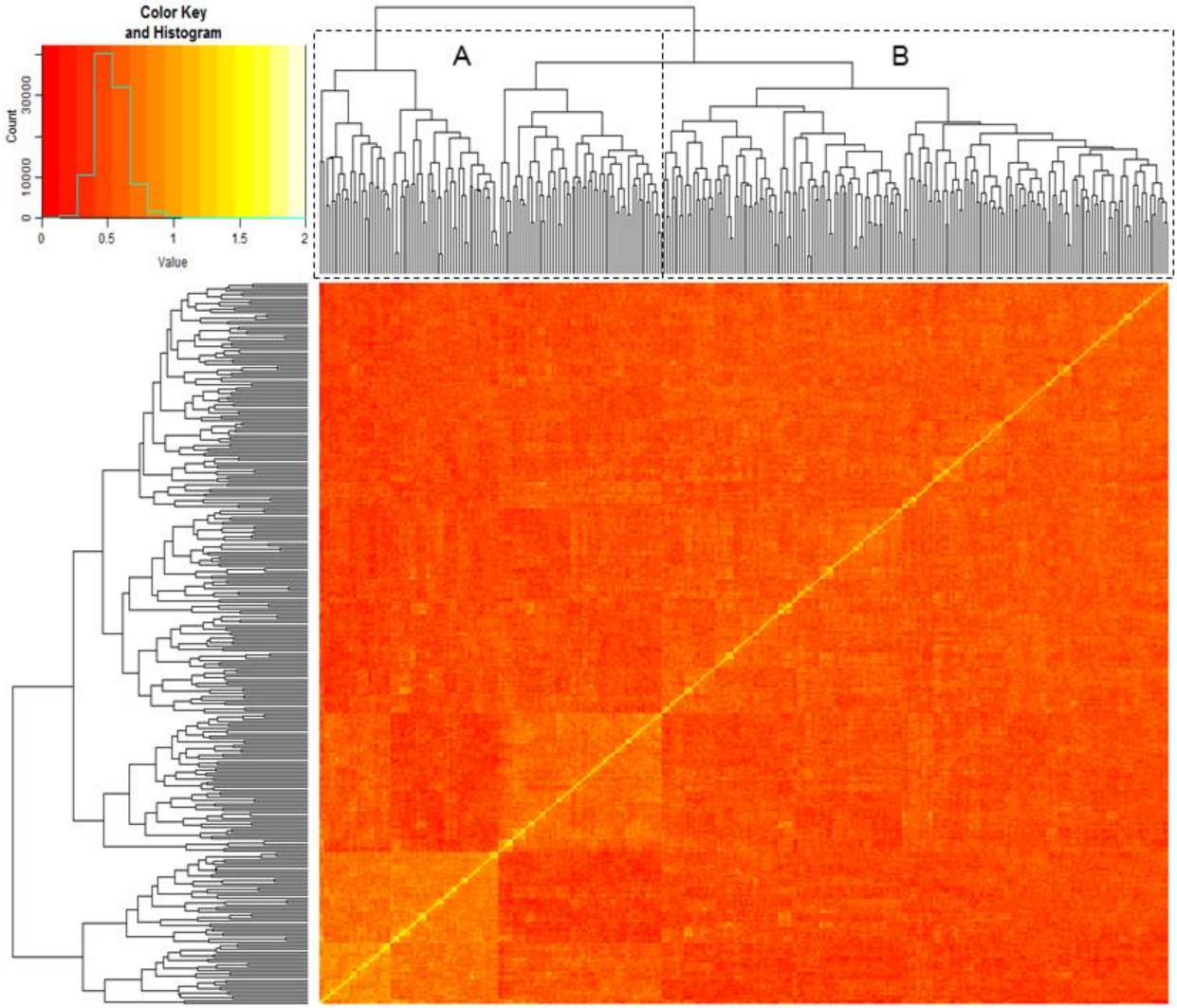



**Fig. 8-1.** Heat map and dendrogram of a genomic relationship matrix estimated using the EMMA algorithm based on 16K SNPs among the 304 wheat lines and three parents. Color codes show groups of lines based on their genomic relationships. Both rows and columns represent the lines. (A) Lines that clustered with Pasteur or none of the parents, (B) lines that clustered with CDC Utmost and CDC Plentiful.

**Table 8-2.** Prediction accuracy based on Pearson's correlations between GEBVs estimated using two statistical models and trait phenotypes. Predictions were made for six traits in two groups of selection candidates.

| | Group A† | | Group B‡ | |
|---|---|---|---|---|
| **Trait§** | **G-BLUP** | **BayesB** | **G-BLUP** | **BayesB** |
| HD | 0.142 | 0.197 | 0.164 | 0.223 |
| HT | -0.071 | -0.124 | 0.143 | -0.026 |
| MAT | 0.083 | 0.136 | 0.198 | 0.289 |
| YLD | 0.127 | 0.141 | -0.002 | -0.003 |
| TWT | 0.079 | 0.056 | 0.302 | 0.260 |
| TKW | 0.164 | 0.188 | 0.293 | 0.327 |

†Selection candidates distantly related to parents included in the training population.

‡Selection candidates closely related to parents included in the training population.

§ HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight.

The across-population genomic prediction accuracies obtained in this study were close to zero even when parents were included in the TP. This indicates that marker effects estimated in one population do not predict GEBVs accurately in a different population unless the two populations are highly related. Several studies reported accuracies close to zero when unrelated populations were used to make genomic predictions (Charmet et al., 2014; Crossa et al., 2014; Riedelsheimer et al., 2013; Windhausen et al., 2012). Negligible prediction accuracies have been reported even when unrelated subpopulations were used to make predictions (Crossa et al., 2014). Studies that included parents, siblings, or other related data sets in the TP indicated that the best accuracies are achieved when TP and SC are closely related (Daetwyler et al., 2014; Riedelsheimer et al., 2013; Riedelsheimer et al., 2013; Zhao et al., 2013; Zhao et al., 2013). Based on five biparental DH maize populations developed from crosses involving four parents, Riedelsheimer et al. (2013) reported mean accuracies of zero or negative values when prediction was made for individuals in biparental families using a model trained based on data from unrelated biparental families. However, using half-sib and full-sib families in the training set improved prediction accuracy to 0.25 and 0.59, respectively. Daetwyler et al. (2014) indicated that the TP should contain at least one line that is highly related to a SC to achieve accurate genomic prediction. Zhao et al. (2013) also reported that on average the accuracy of grain yield prediction in hybrid wheat decreased by 44% when the training and validation sets were not related versus when there was at

least one common parent. On the other hand, based on two bi-parental hybrid rye populations that share one common parent, Wang et al. (2014b) showed that accuracy was substantially lower when one population was used as TP to estimate GEBVs of another population, but accuracy increased when both populations contributed to the training and validation sets. In this study, including two of the parents in the TP did not improve the accuracy of prediction in the SC, indicating that the genetic relationship created between the TP and SC was not sufficient. Meuwissen (2009) indicated that sufficient accuracies can be achieved for unrelated individuals if the size of the TP or marker density is substantially high. Bassi et al., (2016) also suggested that a TP of at least 50 individuals that are full-sibs of the SC, 100 individuals for half-sibs, and at least 1000 individuals for a less related TP are required to achieve accuracies above 0.5. However, we were unable to achieve higher prediction accuracies when using the TP with or without the parents of the SC, indicating that across-population predictions using the data from this study may be impractical or might require more population or marker data.

### 8.3.2 Within-population genomic prediction

In contrast to the low accuracies observed when making across-population genomic predictions, we observed reasonably high accuracies for within-population genomic predictions (Tables 8-1 and 8-3). Prediction accuracies based on fivefold cross-validation within the SC ranged from 0.44 (in RKHS-KA for heading date) to 0.75 (in BayesB for heading date) for all model-trait combinations (Table 8-3). The higher accuracy when predictions were made within the same population is likely because the allele frequency and marker-QTL LD phase is similar between the training and validation sets. Using two hexaploid wheat DH populations Thavamanikumar et al. (2015) also reported higher prediction accuracies based on tenfold cross-validation in each population compared to those obtained based on independent cross-validation.

In addition to observing improved accuracies for within-population genomic predictions, we also observed differences in accuracies among the models when predictions were made for traits with major effect QTL. There were significant differences among the evaluated models for heading date ($P < 0.0001$) and maturity ($P = 0.0002$). Prediction accuracy obtained in BayesB was significantly higher than accuracies of all the other models while the GS + de novo GWAS model was the second best performing model for heading date (Table 8-3). Similarly, prediction accuracies of BayesB and GS + de novo GWAS models were significantly higher than the

accuracies of all the other models for maturity, with BayesB showing a 14% higher accuracy than GS + de novo GWAS. Accuracy obtained in BayesB was 16 to 71% and 14 to 47% higher than the other models for heading date and maturity, respectively. The high accuracies obtained in BayesB were also observed when assessing across-population genomic predictions and may be due to the presence of few large effect QTLs underlying these traits (Table 7-2). Spindel et al. (2016) reported that the GS + de novo GWAS model was the most accurate and outperformed RR-BLUP, BL, RKHS, Random forest, and multiple linear regression models, but the authors did not include BayesB, which is a variable selection method recommended for the prediction of traits that are controlled by few QTL with large effects. In this study, no significant difference was observed among the evaluated models for all the other traits. The similar performance of GS models across most traits is likely because these traits are controlled by many minor effect QTL and most models are reported to achieve similar prediction accuracies in these conditions (Clark et al., 2011).

**Table 8-3.** Average and standard deviation of prediction accuracy (from fivefold cross-validation) based on Pearson's correlation between GEBVs estimated using six statistical models and trait phenotypes of the selection candidates.

| Trait† | RR-BLUP | GS + de novo GWAS | G-BLUP | BayesB | BL | RKHS-KA |
|---|---|---|---|---|---|---|
| HD | 0.507(0.06)$^{C}$ | 0.644(0.08)$^{B}$ | 0.498(0.06)$^{C}$ | 0.750(0.06)$^{A}$ | 0.496(0.07)$^{C}$ | 0.438(0.07)$^{C}$ |
| HT | 0.467(0.07)$^{A}$ | 0.469(0.11)$^{A}$ | 0.474(0.07)$^{A}$ | 0.508(0.08)$^{A}$ | 0.452(0.09)$^{A}$ | 0.487(0.07)$^{A}$ |
| MAT | 0.545(0.07)$^{B}$ | 0.652(0.09)$^{A}$ | 0.540(0.07)$^{B}$ | 0.741(0.05)$^{A}$ | 0.539(0.08)$^{B}$ | 0.503(0.07)$^{B}$ |
| YLD | 0.489(0.06)$^{A}$ | 0.503(0.07)$^{A}$ | 0.489(0.06)$^{A}$ | 0.501(0.06)$^{A}$ | 0.483(0.07)$^{A}$ | 0.514(0.06)$^{A}$ |
| TWT | 0.538(0.15)$^{A}$ | 0.557(0.12)$^{A}$ | 0.539(0.15)$^{A}$ | 0.585(0.14)$^{A}$ | 0.540(0.15)$^{A}$ | 0.536(0.14)$^{A}$ |
| TKW | 0.627(0.08)$^{A}$ | 0.640(0.07)$^{A}$ | 0.630(0.07)$^{A}$ | 0.643(0.07)$^{A}$ | 0.631(0.07)$^{A}$ | 0.655(0.07)$^{A}$ |

$^{ABC}$ Within each row values followed by the same letter were not significantly different at the 0.05 probability level.

†HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight.

### 8.3.3 Genomic prediction across years for combined locations

In plant breeding, GS can be applied to make predictions across generations of the same cross. Accuracies obtained in this study were very high for all traits when genomic predictions were made across years (Table 8-4). When the SC in 2014 (F4:F6 generation) was used to predict SC in 2015 (F4:F7 generation) and SC in 2016 (F4:F8 generation), accuracies for all traits ranged from 0.56 to 0.76 and 0.65 to 0.84, respectively (Table 8-4). When the SC in 2015 was used to predict SC in 2016, accuracies ranged from 0.62 to 0.85. Combining data from 2014 and 2015 resulted in higher accuracy (ranged from 0.68 to 0.90) for all traits except for plant height. Our results agree with Wang et al. (2014b), who also reported that limiting the number of locations or years in field testing for the TP reduced the accuracy of GS predictions. Overall, GS can make accurate predictions across generations, and including more sample years improves accuracies.

In contrast, we observed similar or slightly lower accuracies when combining population data to make across-year genomic predictions. Accuracies declined by 0.003 to 0.07 for all traits when the Co-op and SC data sets were combined to make across-year genomic predictions for the SC. This shows that combining different populations to increase the size of the TP may not be advantageous to make prediction for lines in subsequent generations. These results agree with Charmet et al. (2014) who reported that prediction accuracies did not improve when unrelated populations from different breeding programs were merged to increase TP size. Mixing different populations may reduce LD because the phase of LD varies across populations (Goddard, 2012). The reduced accuracy when combining different populations in the TP could be because only LD that is persistent across those populations is utilized in the prediction equations (Calus, 2010). Moreover, combining multiple related or unrelated populations into one TP may reduce prediction accuracy because of intense population structure in the TP (Riedelsheimer et al., 2013).

Phenotypic prediction accuracies were computed as a correlation between the phenotypes of the SC from the respective environments used to train and validate the GS models. Phenotypic prediction accuracies ranged from 0.76 to 0.88 (heading date), 0.66 to 0.72 (plant height), 0.70 to 0.82 (maturity), 0.62 to 0.77 (grain yield), 0.73 to 0.81 (test weight), and 0.81 to 0.91 (thousand-kernel weight) (Table 8-4). The deviation of these values from one suggests the presence of G × E. The ratio of genomic prediction accuracy to phenotypic prediction accuracy ($r_{GS}/r_P$) ranged from 0.86 (test weight) to 1.03 (maturity) with a mean ratio of 0.96, indicating that accuracies obtained from genomic and phenotypic predictions across years are highly comparable. Heffner et al.

(2011a) also reported similar ratios ranging from 0.84 (heading date) to 1.09 (grain yield) with a mean ratio of 0.95 when predictions were made across years. Similarly, Zhong et al. (2009) reported comparable accuracies between genomic and phenotypic selection methods. Altogether, we observed that genomic and phenotypic prediction accuracies were similar, and that accurate predictions can be made within the same population across generations.

**Table 8-4.** Across-year genomic and phenotypic prediction accuracies based on combined data from two sites. Predictions were made for six traits evaluated across two sites (Kernen and Rosthern) and three years for a total of five environments.

| Training population† | Validation set | HD‡ | HT | MAT | YLD | TWT | TKW |
|---|---|---|---|---|---|---|---|
| | | Genomic prediction accuracy ($r_{GS}$)§ | | | | | |
| SC 2014 | SC 2015 | 0.751 | 0.700 | 0.699 | 0.562 | 0.689 | 0.761 |
| SC 2014 | SC 2016 | 0.772 | 0.647 | 0.713 | 0.665 | 0.705 | 0.843 |
| SC 2015 | SC 2016 | 0.853 | 0.618 | 0.795 | 0.644 | 0.756 | 0.835 |
| SC 2014 and 2015 | SC 2016 | 0.870 | 0.683 | 0.800 | 0.715 | 0.807 | 0.899 |
| Co-op + SC 2014 | SC 2015 | 0.687 | 0.684 | 0.661 | 0.553 | – | 0.758 |
| Co-op + SC 2014 | SC 2016 | 0.699 | 0.638 | 0.669 | 0.651 | – | 0.839 |
| Co-op + SC 2015 | SC 2016 | 0.807 | 0.583 | 0.759 | 0.632 | 0.735 | 0.819 |
| Co-op + SC 2014 and 2015 | SC 2016 | 0.836 | 0.662 | 0.781 | 0.690 | – | 0.893 |
| | | Phenotypic prediction accuracy ($r_P$)¶ | | | | | |
| SC 2014 | SC 2015 | 0.764 | 0.723 | 0.709 | 0.620 | 0.804 | 0.813 |
| SC 2014 | SC 2016 | 0.775 | 0.673 | 0.695 | 0.698 | 0.738 | 0.888 |
| SC 2015 | SC 2016 | 0.859 | 0.661 | 0.799 | 0.685 | 0.786 | 0.860 |
| SC 2014 and 2015 | SC 2016 | 0.881 | 0.717 | 0.818 | 0.767 | 0.807 | 0.912 |
| | | $r_{GS}/r_P$# | | | | | |
| SC 2014 | SC 2015 | 0.983 | 0.968 | 0.986 | 0.906 | 0.857 | 0.936 |
| SC 2014 | SC 2016 | 0.996 | 0.961 | 1.026 | 0.953 | 0.955 | 0.949 |
| SC 2015 | SC 2016 | 0.993 | 0.935 | 0.995 | 0.940 | 0.962 | 0.971 |
| SC 2014 and 2015 | SC 2016 | 0.988 | 0.953 | 0.978 | 0.932 | 1.000 | 0.986 |

†SC: selection candidates

‡HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight.

§Pearson's correlations between predicted values and trait phenotypes.

¶Pearson's correlations between phenotypes of the SC across years.

#$r_{GS}/r_P$ is the ratio of genomic prediction accuracy to phenotypic prediction accuracy.

Missing cells indicate that there was no data for the Co-op population in that year.

### 8.3.4 Genomic prediction across years within locations

One of the desired uses of GS is to make genomic predictions across years for a single location. When predictions were made across years separately in each location, both genomic and phenotypic prediction accuracies were lower compared to the prediction accuracy of combined analyses of locations (Table 8-5). At Kernen, when the SC in 2014 was used as TP to predict SC in 2015 and 2016, accuracies for all traits ranged from 0.50 to 0.74 and 0.57 to 0.76, respectively. When the SC in 2015 was used as TP to predict SC in 2016, accuracies for all traits ranged from 0.49 to 0.81 and 0.43 to 0.74 at Kernen and Rosthern, respectively (Table 8-5). Combining data from 2014 and 2015 resulted in higher accuracy (ranging from 0.59 to 0.82) at Kernen, suggesting that genomic predictions across years were better when data from multiple locations or years were combined as opposed to using data from a single location or year. However, increasing the TP size by including the Co-op data set did not improve prediction accuracy in the SC.

Phenotypic prediction accuracies ranged from 0.73 to 0.84 (heading date), 0.49 to 0.68 (plant height), 0.63 to 0.74 (maturity), 0.44 to 0.57 (grain yield), 0.66 to 0.76 (test weight), and 0.75 to 0.83 (thousand-kernel weight) at Kernen, while at Rosthern phenotypic correlations between the SC in 2015 and 2016 were 0.72 (heading date), 0.47 (plant height), 0.64 (maturity), 0.61 (grain yield), 0.63 (test weight), and 0.77 (thousand-kernel weight) (Table 8-5). These accuracies varied depending on the trait and whether data were combined between years. The ratio of genomic prediction accuracy to phenotypic prediction accuracy ($r_{GS}/r_P$) at Kernen ranged from 0.86 (test weight) to 1.10 (grain yield) with a mean ratio of 0.99, while the ratio ranged from 0.90 (grain yield) to 1.03 (heading date) with a mean ratio of 0.96 at Rosthern; this suggests that genomic and phenotypic prediction accuracies were similar within test sites. These results indicate that accurate genomic predictions can be made using data from subsequent breeding stages and GS can be applied successfully to advance generations of a cross. However, it is important to note that in this study across-year predictions were made for inbred lines (F6 generation and later) and results might differ in segregating early generations. Moreover, no selection was made when advancing generations of the SC and selection that would typically take place in a practical breeding situation might affect marker-QTL LD and the accuracy of predictions.

**Table 8-5.** Across-year genomic and phenotypic prediction accuracies in each site. Predictions were made for six traits evaluated across two sites (Kernen and Rosthern) and three years for a total of five environments.

| Training population† | Validation set | HD‡ | HT | MAT | YLD | TWT | TKW |
|---|---|---|---|---|---|---|---|
| **Kernen** | | | | | | | |
| | Genomic prediction accuracy ($r_{GS}$)§ | | | | | | |
| SC 2014 | SC 2015 | 0.736 | 0.569 | 0.633 | 0.500 | 0.655 | 0.707 |
| SC 2014 | SC 2016 | 0.750 | 0.623 | 0.652 | 0.574 | 0.641 | 0.757 |
| SC 2015 | SC 2016 | 0.807 | 0.492 | 0.714 | 0.486 | 0.688 | 0.736 |
| SC 2014 and 2015 | SC 2016 | 0.819 | 0.610 | 0.743 | 0.586 | 0.729 | 0.808 |
| Co-op + SC 2014 | SC 2015 | 0.679 | 0.560 | 0.593 | 0.496 | – | 0.698 |
| Co-op + SC 2014 | SC 2016 | 0.681 | 0.610 | 0.616 | 0.566 | – | 0.759 |
| Co-op + SC 2015 | SC 2016 | 0.768 | 0.477 | 0.679 | 0.484 | 0.671 | 0.727 |
| Co-op + SC 2014 and 2015 | SC 2016 | 0.769 | 0.605 | 0.729 | 0.576 | – | 0.805 |
| | Phenotypic prediction accuracy ($r_P$)¶ | | | | | | |
| SC 2014 | SC 2015 | 0.729 | 0.566 | 0.64 | 0.528 | 0.764 | 0.761 |
| SC 2014 | SC 2016 | 0.748 | 0.67 | 0.628 | 0.540 | 0.663 | 0.800 |
| SC 2015 | SC 2016 | 0.802 | 0.493 | 0.715 | 0.443 | 0.690 | 0.752 |
| SC 2014 and 2015 | SC 2016 | 0.840 | 0.681 | 0.743 | 0.569 | 0.721 | 0.829 |
| | $r_{GS}/r_P$# | | | | | | |
| SC 2014 | SC 2015 | 1.009 | 1.005 | 0.988 | 0.947 | 0.858 | 0.930 |
| SC 2014 | SC 2016 | 1.003 | 0.930 | 1.038 | 1.063 | 0.967 | 0.946 |
| SC 2015 | SC 2016 | 1.006 | 0.997 | 0.999 | 1.096 | 0.998 | 0.979 |
| SC 2014 and 2015 | SC 2016 | 0.975 | 0.896 | 1.000 | 1.028 | 1.011 | 0.976 |
| **Rosthern** | | | | | | | |
| | Genomic prediction accuracy ($r_{GS}$)§ | | | | | | |
| SC 2015 | SC 2016 | 0.734 | 0.432 | 0.636 | 0.552 | 0.614 | 0.742 |
| Co-op + SC 2015 | SC 2016 | 0.680 | 0.367 | 0.613 | 0.529 | 0.595 | 0.722 |
| | Phenotypic prediction accuracy ($r_P$)¶ | | | | | | |
| SC 2015 | SC 2016 | 0.716 | 0.472 | 0.638 | 0.614 | 0.632 | 0.766 |
| | $r_{GS}/r_P$# | | | | | | |
| SC 2015 | SC 2016 | 1.025 | 0.915 | 0.997 | 0.899 | 0.972 | 0.969 |

†SC: selection candidates

‡HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight.

§Pearson's correlations between predicted values and trait phenotypes.

¶Pearson's correlations between phenotypes of the SC across years.

#$r_{GS}/r_P$ is the ratio of genomic prediction accuracy to phenotypic prediction accuracy.

Missing cells indicate that there was no data for the Co-op population in that year.

## 8.4 Conclusion

This study assessed the ability for GS to make predictions that address specific breeder needs. We compared different prediction scenarios using various cross-validation techniques and different statistical methods. Highly variable estimates of prediction accuracies were obtained across these methods and were strongly influenced by the cross-validation technique. The results showed that the genetic relationship between the TP and SC is an important factor affecting prediction accuracies, and it is important to develop a TP with strong genetic relationship to the SC. Genomic predictions across populations were close to zero even when few closely related lines were included in the TP, indicating that the LD phase between the two populations is more important than the genetic relationship from few highly-related lines. On the other hand, the commonly used fivefold cross-validation technique resulted in moderate to high genomic prediction accuracies. This indicates that it is important to use a cross-validation technique that resembles the actual application of GS in a breeding program. Comparison of different statistical methods based on within and across-population genomic predictions indicated that BayesB is superior to RR-BLUP, GS + de novo GWAS, G-BLUP, BL or RKHS-KA models when there were large effect QTL underlying traits, but in the absence of detectable large effect QTL there was no difference among the evaluated methods.

Genomic predictions based on data collected in one year to predict the performance of lines evaluated in a different year resulted in comparable accuracies to that based on the phenotypes. This indicates that GS can be successfully implemented to make across-year genomic predictions for subsequent generations of a cross in a breeding program. However, the potential of GS to make predictions in segregating populations from early generations needs further investigation. Combining data across locations or years in the TP resulted in higher prediction accuracy, suggesting that it is important to evaluate the training set in more than one environment to achieve higher prediction accuracy. However, combining data from two unrelated populations to increase the TP size did not improve prediction accuracy. Finally, the ratios of genomic to phenotypic prediction accuracies were close to one; this is an important finding because it suggests that GS can be just as effective as phenotypic selection. Therefore, GS can be used as an alternative to phenotypic selection and has the potential to transform wheat breeding.

# 9. General Discussion and Conclusions

## 9.1 Quantitative Trait Loci and Genome-Wide Association Mapping

One of the objectives of this study was to identify markers significantly associated with QTL that control important agronomic and end-use quality traits in wheat. QTL and genome-wide association mapping studies were conducted using two different populations. The QTL mapping population was composed of 304 heterogeneous RILs developed from three-way crosses (CDC Plentiful//Pasteur/CDC Utmost) and GWAS was performed using 231 elite breeding lines. The elite breeding lines were used as TP and the RIL population was developed to be used as SC for GS study (Chapter eight). These populations were used to identify QTL underlying important agronomic traits and to better understand the genetic architecture of traits for GS (Chapters six and seven). A total of 23 QTL that underlie six important agronomic traits were detected in the SC based on the average LS-means across five environments (Fig. 7-2). Six QTL were identified for heading date and test weight, four QTL for maturity, three QTL for thousand-kernel weight, and two QTL each for plant height and grain yield. Most of these QTL had only minor effects, with heading date and maturity having the only large effect QTL. Similarly, 34 significant marker-trait associations were identified for eight agronomic and end-use quality traits from GWAS in the TP (Table 6-1). Twelve markers were significantly associated with plant height, ten markers were associated with sedimentation volume, four markers were associated with heading date, three markers were associated with thousand-kernel weight, two markers were associated with maturity, and one marker was associated with each of grain yield, test weight and falling number (Table 6-1). Most of the significant markers identified in this population also had minor effects.

Further characterization of the QTL and GWAS results indicated that some loci may be the same in both populations and may correspond to known genes or QTL. One of the SNP marker (*wsnp_CAP12_c812_428290*) flanking the region on chromosome 2D (18.4 – 23.4 cM) that harboured a stable QTL for heading date, maturity and plant height in the RILs population was also significantly associated with heading date and maturity in the TP. In the TP, this marker explained 8 and 6% of the variation in heading date and maturity, respectively. In wheat, a major photoperiod insensitive gene (*Ppd-D1*) is located on the short arm of chromosome 2D (Worland, 1996). The *Ppd-D1* gene is involved in pleiotropic height reducing effects by accelerating ear emergence time and reducing plant life cycle (Worland et al., 1998; Worland et al., 2001). In this

study, the heading date, plant height and maturity QTL identified on chromosome 2D could be linked to the *Ppd-D1* gene. Marker-trait associations were also detected for heading date on chromosomes 5B and 5D, and maturity on chromosome 5B in the TP. These markers explained 5 to 8% of the phenotypic variance in this population. In wheat, the most important vernalization response genes *VRN-A1*, *VRN-B1*, and *VRN-D1* were mapped previously in collinear regions on the long arm of chromosomes 5A, 5B and 5D, respectively (Dubcovsky et al., 1998; Law et al., 1976). The genomic regions of significant markers associated with plant height also corresponded to genes or QTLs that were reported previously in wheat. For example, chromosome 4B is known to carry several reduced height genes including the major *Rht-B1b* gene that is known to affect plant height (Börner et al., 1996). Overall, several QTL underlying various important agronomic and end-use quality traits were detected in this study. Large effect QTL were identified for heading date and maturity in the SC but the genetic effects of most of the individual QTL identified for the other traits were relatively small, explaining < 10% of the phenotypic variance, which indicates the quantitative nature of the evaluated traits. Moreover, several environment-specific QTL were detected, indicating the presence of QTL-by-environment interactions. The identified QTL can be utilized for MAS following further validation in different genetic backgrounds.

### 9.2 Comparison of GS Approaches and Models

The main objective of this study was to identify the most appropriate statistical model and approach to implement GS in wheat breeding programs; to this end, we compared different GS approaches and models. In Chapter three, we used a TP of 231 varieties and advanced breeding lines to evaluate nine single-trait prediction models (RR-BLUP, G-BLUP, BRR, BL, BayesA, BayesB, BayesC$\pi$, RKHS, and RKHS-KA), three multiple-trait prediction models (multiple-trait BayesA, multiple-trait BayesA Matrix and Scalar) and a model that combines GWAS and GS (GS + de novo GWAS) using similar cross-validation folds. The average prediction accuracies based on Pearson's correlation between estimated GEBVs and trait phenotypes ranged from 0.55 to 0.77 across different model-trait combinations using fivefold cross-validation (Fig. 3-2). Generally, accuracies based on Spearman's rank correlation were smaller than accuracies based on Pearson's correlation, but all methods exhibited the same trend for both correlations. The results also suggested that the average accuracies of the evaluated single-trait prediction models were not significantly different for each trait (Fig. 3-2). Similarly, in Chapter eight, several statistical models

such as RR-BLUP, GS + de novo GWAS, G-BLUP, BayesB, BL, and RKHS-KA were compared using the 304 SC (RILs) based on fivefold cross-validation. Prediction accuracies within the SC ranged from 0.44 (in RKHS-KA for heading date) to 0.75 (in BayesB for heading date) for all model-trait combinations (Table 8-3). The results also indicated that prediction accuracy in BayesB was significantly higher than accuracies obtained in RR-BLUP, G-BLUP, BL and RKHS-KA models for heading date and maturity but there was no significant difference among these models for all the other traits (Table 8-3). QTL analyses in this population revealed that there were stable QTL with relatively large effects underlying heading date and maturity, but effects of the QTL identified for the other traits were not large enough to be captured by the model (Table 7-2).

The GS + de novo GWAS model is equivalent to RR-BLUP with up to three of the most significant SNPs identified from fold specific GWAS in the TP fitted as fixed effects (Spindel et al., 2016). This model was initially tested in rice and gains in prediction accuracy were reported for traits with large GWAS peaks (Spindel et al., 2016). In this study, no significant difference was observed between GS + de novo GWAS and the standard RR-BLUP in the TP (Fig. 3-6). This was because strong marker trait associations were not detected in the TP (Table 6-1). In the SC, GS + de novo GWAS outperformed RR-BLUP, G-BLUP, BL and RKHS-KS but it was inferior to BayesB for heading date and maturity (Table 8-3). However, for the other traits the accuracy of the GS + de novo GWAS model was not significantly different from the other models (Table 8-3). This suggests that fitting significant markers as fixed effects in GS is not advantageous if the effects of QTL underlying traits are small. This study also showed that BayesB is superior to GS + de novo GWAS when there are large-effect QTL underlying traits. The results from Chapters three and eight agree with previous studies that reported similar levels of accuracy among different GS models based on empirical data (Charmet et al., 2014; Daetwyler et al., 2013; Heslot et al., 2012). Simulation studies showed that when genetic variation is controlled by a few QTL with relatively large effects, variable selection methods such as BayesB have significantly higher accuracy than methods that shrink marker effects by assuming equal variance (Clark et al., 2011). Overall, the results from both Chapters three and eight suggest that variable selection methods, such as BayesB, should be considered when there are large effect QTL underlying traits, but when there are several QTL with minor effects, most GS models have similar performances.

This study also evaluated multiple-trait genomic prediction models in wheat. The results from Chapter three showed that there was no advantage of the evaluated multiple-trait models over

the single-trait models for all trait combinations (Table 3-4). Previous studies based on simulation and real data from animal breeding programs reported that multiple-trait models have advantages over single-trait models for the prediction of low-heritability traits genetically correlated with high-heritability traits (Guo et al., 2014a; Hayashi and Iwata, 2013; Jia and Jannink, 2012; Jiang et al., 2015). Jia and Jannink (2012) also indicated that multiple-trait models capture the genetic correlation between traits more efficiently under major QTL genetic architectures. Moreover, multiple-trait models resulted in higher prediction accuracy than single-trait models when phenotypic records are missing for some of the individuals and traits (Calus and Veerkamp, 2011; Guo et al., 2014a; Jia and Jannink, 2012). However, multiple-trait models were inferior to single-trait models in the absence of genetic correlation between traits, when trait heritability is high ($h^2 > 0.5$), and for traits with complete phenotypic data (Guo et al., 2014a; Hayashi and Iwata, 2013; Jia and Jannink, 2012). In this study, there was no benefit to the multiple-trait models over single-trait models, likely because the estimates of heritability were high for all traits compared to the heritability estimates used in most simulation studies (Table 3-2). Moreover, no major QTL were detected for all traits in the TP and this might also have affected the ability of the multiple-trait models to capture the genetic correlation among traits. Antedependence-based GS models, which consider SNP effects as being spatially correlated based on the relative physical location of SNP markers along the chromosome, were reported previously to have higher accuracy than their standard counterparts (Jiang et al., 2015; Yang and Tempelman, 2012). In this study, we did not find any benefit of the antedependence-based Bayesian multiple-trait prediction models.

### 9.3 Effects of Training Population Size, Marker Density, Heritability and Population Structure on GS accuracy

The effect of TP size was an important factor that affected prediction accuracy. Increasing the TP size from 50 to 200 increased the prediction accuracy for all traits on average by 28% (ranging from 20 to 35%) (Fig. 4-1). This suggests that accurate predictions can be obtained by using a large TP size to estimate marker effects. In GS, several studies showed that increasing the TP size increases the accuracy of GS by providing more data to estimate marker effects (Asoro et al., 2011; Meuwissen et al., 2001; Muir, 2007; Saatchi et al., 2010; VanRaden et al., 2009). Combs and Bernardo (2013) suggested that the product of TP size and heritability is an important factor that determines accuracy rather than heritability or TP size individually. All things considered, it

is especially important to increase the accuracy of prediction for traits with low heritability by increasing the TP size.

The effect of marker density on GS prediction accuracy was evaluated using five marker densities (770, 3K, 13K, 15K, and 18K SNPs). The results indicated that there was no difference in prediction accuracy among the evaluated marker densities for each trait (Fig. 4-2). This suggests that a reduced subset of evenly spaced markers can be sufficient for GS. Moser et al. (2010) also reported that a low-density assay of evenly spaced SNPs can provide sufficient prediction accuracies if the information content of the subset of SNPs is sufficient to estimate effects of distinct haplotypes. Similarly, Bassi et al. (2016) suggested that even marker distribution across the genome is crucial to capture important QTL. In this study, the similar prediction accuracies when marker number was increased could be because 770 evenly spaced SNPs were sufficient to accurately estimate the genomic relationships among the lines. The benefit of increasing marker density is to maximize the number of QTL in LD with at least one marker, which also maximizes the number of QTL whose effects will be captured by markers (Heffner et al., 2009). Marker density can also be a function of marker-QTL LD and the genetic architecture of the trait. It is therefore imperative to evaluate the optimum number of markers needed for accurate prediction of genetic values in different populations. Another explanation for the similar prediction accuracies when marker density was increased could be that there are insufficient degrees of freedom to benefit from the increase in marker density. Sample size is one of the most important factors limiting GS prediction accuracy (de los Campos et al., 2015). Increasing the number of markers without increasing the TP size may reduce accuracy because it increases collinearity among markers (Muir, 2007). In this study, increasing the TP size had a more important effect on accuracy than marker number.

In Chapters three and four, we evaluated the effects of trait heritability, marker number and density, TP size, and population structure on GS prediction accuracy. The results indicated that there was no direct relationship between trait heritability and prediction accuracy (Table 3-2; Fig. 3-2). Previous studies reported a strong relationship between the accuracy of genomic prediction and trait heritability, with the prediction being more accurate for traits with higher heritability (Combs and Bernardo, 2013; Heffner et al., 2011a; Moser et al., 2010; Saatchi et al., 2010). However, there are reported cases where prediction accuracy was higher for low heritability traits compared to traits with high heritability (Combs and Bernardo, 2013). In this study, heritability

had no affect on accuracy; this could be because heritability estimates were high for all traits in the population used for this study.

The effect of population structure on GS prediction accuracy has been reported previously. In Chapter four, plots of the first two PCs indicated that the three subpopulations were weakly differentiated (Fig. 4-4). This suggests that the effect of population structure is not strong in this population. This is because the population used in this study was composed of elite breeding materials that have been intercrossed frequently, resulting in various degrees of admixtures. The effect of population structure on genomic prediction accuracy was assessed by using an interaction model that decomposes marker effects into components that are constant across groups and components that are group specific (de los Campos et al., 2015). This approach was compared with an across-group model that assumes constant marker effects across subpopulations, thereby ignoring population stratification, and a stratified or within-group model that estimates marker effects separately in each subpopulation (de los Campos et al., 2015). The results showed that the prediction accuracy of an interaction model that accounted for population structure was similar to an across-group model that ignored population structure (Fig. 4-5). This indicates that population structure had no effect on prediction accuracy in this population.

### 9.4 Modelling Genotype-by-Environment Interaction in GS

Environment has a major impact on cultivar performance and breeding strategies; therefore, environmental effects should be considered when performing GS predictions. In Chapter five, two methods of modelling G × E in GS were examined based on 81 lines that were evaluated in seven environments. The first method was based on M × E model that was implemented in Lopez-Cruz et al. (2015). The performance of the M × E model was compared with a single-environment model that estimates marker effects in each environment separately and across-environment model that ignores G × E (Lopez-Cruz et al., 2015). The second method used a class of reaction norm models that incorporate the main and interaction effects of molecular markers, environments, and ECs using covariance functions as implemented in Jarquín et al. (2014a). Predictions were made for grain yield and protein content based on two cross-validation designs (CV1 and CV2) that mimic real situations faced by plant breeders (Burgueño et al., 2012). CV1 involved prediction of phenotypes for lines that have never been tested in any of the environments mimicking newly developed lines, while CV2 involved prediction of phenotypes for lines that were evaluated in

some but not in other environments, thereby mimicking incomplete field trials. Environmental variables related to temperature, humidity and precipitation were included in the reaction norm models.

When comparing the M × E to the single-environment and across-environment models, prediction accuracies varied depending on the cross-validation design used. The results indicated that for both traits the accuracy of the single-environment model was better than the accuracy of the M × E and across-environment models in CV1 (Tables 5-3 and 5-4). Previous studies also reported that the single-environment model performed either similar or better than the M × E model in CV1 (Crossa et al., 2015; Lopez-Cruz et al., 2015). This suggests that estimating marker effects separately in each environment is practical for predicting GEBVs of lines that were not tested in any of the environments. In CV2, the accuracy of the M × E model was 24.8 to 62.8% and 12.7 to 29.4% higher compared to the single-environment model for grain yield and protein content, respectively. This is because multi-environment analysis uses information for each line across-environments. However, prediction accuracies of the M × E model were comparable to the across-environment model in CV2, indicating that there was no benefit of modelling G × E for both traits. Previous studies reported that the M × E model was the best performing model in CV2 (Crossa et al., 2015; Lopez-Cruz et al., 2015). Some of the differences between the results obtained in this study and the results from previous studies could be due to differences in environments, TP size, or other population characteristics. Generally, prediction accuracy in CV1 was lower compared to the accuracy obtained in CV2. This is because CV2 uses information for lines across-environments, while this is not possible in CV1 because the lines have not been evaluated in any environment (Burgueño et al., 2012).

Six reaction norm models were also evaluated in this study for modelling G × E in GS. The first model (EG) included only the main effects of environments and markers without interaction terms. Similarly, the second model (EGW) included only the main effects of environments, markers and ECs without interactions. The other four models included interaction terms of G × E, G × W or both G × W and G × E in addition to the main effects. The results suggested that adding the main effects of the ECs to the main effects of markers and environments did not improve prediction accuracy for both grain yield and protein content. Similar results were reported by Pérez-Rodríguez et al. (2015) because ECs did not vary within the environment. Moreover, adding the interaction terms did not improve prediction accuracy in both cross-validation designs except

for grain protein, where in CV1 the most comprehensive model EGW-G×WG×E gave 2 to 11% higher prediction accuracy compared to models that included only the main effects (Tables 5-5 and 5-6). Previous studies reported improved prediction accuracy of models that included main and interaction terms compared to models based only on the main effects (Jarquín et al., 2014a; Jarquín et al., 2017; Pérez-Rodríguez et al., 2015; Sukumaran et al., 2017). In this study, lack of improvement in accuracy when modelling G × E in GS could be because the main effects of markers and environments explained a large proportion of the variance in the evaluated traits. However, the overall prediction accuracies obtained in this study were higher compared to the previously reported values (Crossa et al., 2015; Jarquín et al., 2014a; Jarquín et al., 2017; Lopez-Cruz et al., 2015; Sukumaran et al., 2017). Comparison of the two methods of modelling G × E in GS was inconclusive because the results were variable depending on the cross-validation design and the trait.

### 9.5 Effects of Genetic Relatedness, Training Population Composition and Cross-Validation Technique on GS accuracy

The purpose of GS is to improve the speed and efficiency of crop breeding. We therefore assessed the performance of GS in different scenarios that address specific breeder needs. Genomic predictions were made across populations, within population, and across years. Genomic prediction across populations were made by calculating GEBVs for SC using marker effects estimated from the TP. The genetic relationship between the TP and SC was evaluated by including parents of the SC in the TP, excluding parents from the TP, and by adding parents along with 50 and 100 randomly selected lines from SC in the TP. The results indicated that across-population genomic prediction accuracies were close to zero with or without the parents in the TP. Previous studies also reported accuracy close to zero when unrelated populations were used to make genomic predictions (Charmet et al., 2014; Crossa et al., 2014; Riedelsheimer et al., 2013; Windhausen et al., 2012). The low accuracy could be because of differences in allele frequencies and LD patterns between the TP and SC (Bassi et al., 2016; Charmet et al., 2014; de Roos et al., 2008; de Roos et al., 2009; Goddard, 2012; Meuwissen, 2009). A marker and QTL that are in LD in one population may not be in LD in another population, resulting in poor prediction accuracies (de Roos et al., 2008). This suggests that marker effects estimated in one population do not predict GEBVs accurately in a different, unrelated population. For example, significant marker-trait

associations were detected for plant height on chromosomes 2A, 4B, and 5B in the TP, while QTL for plant height were detected on chromosomes 2D and 6D in the SC (Tables 6-1 and 7-2). When predictions were made for plant height across these populations, markers that are in LD with the QTL on chromosomes 2A, 4B, and 5B are given more weight to estimate marker effects in the TP but these markers have no association to the QTL controlling plant height in the SC, resulting in negative prediction accuracies in BayesB (Table 8-1).

Including 50 lines from the SC in the TP resulted in 1.2 to 4.4-fold increase in accuracy compared to when only the parents were included (Table 8-1). A further increase of the number of SC in the TP to 100 resulted in 1.8 to 4.6-fold increase in accuracy for all traits compared to when only the parents were included (Table 8-1). This suggests that the genetic relationship between the TP and SC is important to improve prediction accuracy. Therefore, GS should not be used to make predictions across different populations unless these populations are closely related. In contrast, within-population genomic predictions based on fivefold cross-validation in the SC resulted in accuracies ranging from 0.44 (in RKHS-KA for heading date) to 0.75 (in BayesB for heading date) for all model-trait combinations (Table 8-3). This suggests that the k-fold cross-validation technique commonly used to assess prediction performance of GS models can be misleading if the actual application of GS in a breeding program is to make prediction across independent populations. The most attractive application of GS in plant breeding is to estimate GEBVs in one population based on marker effect estimates from an independent population. However, results from this study as well as previous studies showed that current GS models are not suitable to make across-population genomic predictions.

Another potential application of GS in crop breeding programs is to make prediction of GEBVs across years to advance generations of the same cross. In this case, the parental generation is used to estimate marker effects and GEBVs will be estimated for progenies. Both genomic and phenotypic prediction accuracies were variable across traits but were reasonable for making informed breeding decisions (Tables 8-4 and 8-5). In this study, genomic prediction accuracies ranged from 0.75 to 0.87 for heading date, 0.62 to 0.70 for plant height, 0.70 to 0.80 for maturity, 0.56 to 0.72 for grain yield, 0.69 to 0.81 for test weight, and 0.76 to 0.90 for thousand-kernel weight when predictions were made across years for the combined locations data set (Table 8-4). Phenotypic prediction accuracy was calculated as the correlation between observed phenotypes in the environments used for model training and cross-validation. Phenotypic prediction accuracies

ranged from 0.76 to 0.88 for heading date, 0.66 to 0.72 for plant height, 0.70 to 0.82 for maturity, 0.62 to 0.77 for grain yield, 0.74 to 0.81 for test weight, and 0.81 to 0.91 for thousand-kernel weight. Combining data across locations or years resulted in higher prediction accuracy, indicating the importance of evaluating the TP in more than one environment to achieve higher prediction accuracy. On the other hand, combining data from the Co-op and SC to make across-year predictions in the SC reduced prediction accuracies by 0.003 to 0.07 for all traits. This suggests that there is no advantage of combining different populations to increase the size of the TP. When comparing the phenotypic and genomic prediction accuracies to each other, their ratio ranged from 0.86 to 1.03 for all traits with a mean ratio of 0.96. This is a very important finding because it indicates that accuracies obtained from genomic and phenotypic predictions across years are highly comparable; therefore, GS can be applied to make across-year genomic predictions for subsequent generations of a cross in a wheat breeding program.

### 9.6 Future Research

The results of this study indicate that there is a strong potential for GS in wheat breeding; however, many of our results and analyses could be improved using larger population sizes. Based on a TP of 231 elite breeding lines and SC of 304 RILs, this study showed that marker effects estimated in one population do not predict GEBVs accurately in a different population. However, successful application of GS in crop breeding programs requires that accurate across-population genomic predictions need to be made. Based on simulated data, Meuwissen (2009) showed that breeding values of unrelated individuals can be predicted with accuracies of 0.88 – 0.93 if the population sizes are large enough. Similarly, due to the unbalanced nature of the data, the effect of modelling G × E in GS was evaluated based on a subset of 81 lines evaluated in seven environments. The effect of modelling G × E needs to be evaluated further using a larger TP size. Moreover, ECs related to temperature, humidity and precipitation were used to model G × E in this study. It is important that these and other environmental variables related to soil characteristics need to be investigated using a larger TP. It may be possible to better incorporate G × E and make more accurate across-population genomic predictions using GS in wheat if larger populations are used.

Important questions that need to be answered are when and how to implement GS in wheat breeding programs. These were beyond the objectives of this study and further research is required

to determine when and how to incorporate knowledge about GEBVs derived from GS models into breeding programs. Moreover, a detailed cost-benefit analysis is required to determine the benefit of GS over standard breeding approaches in wheat. Towards answering these questions, the results of this study showed that across-year genomic prediction accuracies are comparable to prediction accuracies obtained based on the phenotypes. Therefore, GS has a potential to accelerate wheat breeding using data from subsequent breeding stages; this may be the best way to implement GS in wheat breeding. However; in this study, genomic predictions were made for inbred lines (F6 generation and later) advanced without selection and it is important to evaluate across-year genomic predictions for early generation segregating populations in actual breeding programs.

## References


AACC International. Approved methods of analysis, 11th Ed. Method 46-30.01. Crude protein-combustion method. Approved November 8, 1995; reapproved November 3, 1999. Method 56-81.03. Determination of falling number. Approved November 2, 1972; reapproved November 3, 1999. AACC International, St. Paul, MN, U.S.A.

Abdurakhmonov, I.Y., and A. Abdukarimov. 2008. Application of association mapping to understanding the genetic diversity of plant germplasm resources. Int. J. Plant. Genomics 2008:574927.

Aguilar, I., I. Misztal, S. Tsuruta, G.R. Wiggans, and T.J. Lawlor. 2011. Multiple trait genomic evaluation of conception rate in Holsteins. J. Dairy Sci. 94:2621-2624.

Akdemir, D., J.I. Sanchez, and J.-L. Jannink. 2015. Optimization of genomic selection training populations with a genetic algorithm. Genet. Sel. Evol. 47:38.

Albrecht, T., V. Wimmer, H.-J. Auinger, M. Erbe, C. Knaak, M. Ouzunova, H. Simianer, and C.-C. Schön. 2011. Genome-based prediction of testcross values in maize. Theor. Appl. Genet. 123:339-350.

Annicchiarico, P., N. Nazzicari, X. Li, Y. Wei, L. Pecetti, and E.C. Brummer. 2015. Accuracy of genomic selection for alfalfa biomass yield in different reference populations. BMC Genomics 16:1020.

Arruda, M.P., P.J. Brown, A.E. Lipka, A.M. Krill, C. Thurber, and F.L. Kolb. 2015. Genomic selection for predicting head blight resistance in a wheat breeding program. The Plant Genome 8. doi:10.3835/plantgenome2015.01.0003.

Asif, M., R.-C. Yang, A. Navabi, M. Iqbal, A. Kamran, E.P. Lara, H. Randhawa, C. Pozniak, and D. Spaner. 2015. Mapping QTL, selection differentials, and the effect of *Rht-B1* under organic and conventionally managed systems in the Attila × CDC go spring wheat mapping population. Crop Sci. 55:1129-1142.

Asoro, F.G., M.A. Newell, W.D. Beavis, M.P. Scott, and J.-L. Jannink. 2011. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. The Plant Genome 4:132-144.

Avni, R., M. Nave, O. Barad, K. Baruch, S.O. Twardziok, H. Gundlach, I. Hale, M. Mascher, M. Spannagl, K. Wiebe, K.W. Jordan, G. Golan, J. Deek, B. Ben-Zvi, G. Ben-Zvi, A. Himmelbach, R.P. MacLachlan, A.G. Sharpe, A. Fritz, R. Ben-David, H. Budak, T. Fahima, A. Korol, J.D. Faris, A. Hernandez, M.A. Mikel, A.A. Levy, B. Steffenson, M. Maccaferri, R. Tuberosa, L. Cattivelli, P. Faccioli, A. Ceriotti, K. Kashkush, M. Pourkheirandish, T. Komatsuda, T. Eilam, H. Sela, A. Sharon, N. Ohad, D.A. Chamovitz, K.F.X. Mayer, N. Stein, G. Ronen, Z. Peleg, C.J. Pozniak, E.D. Akhunov, and A. Distelfeld. 2017. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science 357:93-97.

Axford, D.W.E., E.E. McDermott, and D.G. Redman. 1978. Small scale tests of bread-making quality. Mill. Feed Fert. 161:18-20.

Bagge, M., and T. Lübberstedt. 2008. Functional markers in wheat: Technical and economic aspects. Mol. Breed. 22:319-328.

Bassi, F.M., A.R. Bentley, G. Charmet, R. Ortiz, and J. Crossa. 2016. Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). Plant Science 242:23-36.

Bates, D., M. Maechler, B. Bolker, S. Walker, R.H.B. Christensen, H. Singmann, B. Dai and G. Grothendieck. 2016. lme4: Linear Mixed-Effects Models using 'Eigen' and S4. https://cran.r-project.org/package=lme4. R Package Version 1.1-7.

Battenfield, S.D., C. Guzmán, R.C. Gaynor, R.P. Singh, R.J. Peña, S. Dreisigacker, A.K. Fritz, and J.A. Poland. 2016. Genomic selection for processing and end-use quality traits in the CIMMYT spring bread wheat breeding program. The Plant Genome 9. doi:10.3835/plantgenome2016.01.0005.

Beaulieu, J., T. Doerksen, S. Clément, J. MacKay, and J. Bousquet. 2014. Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. Heredity 113:343-352.

Begum, H., J.E. Spindel, A. Lalusin, T. Borromeo, G. Gregorio, J. Hernandez, P. Virk, B. Collard, and S.R. McCouch. 2015. Genome-wide association mapping for yield and other agronomic traits in an elite breeding population of tropical rice (*Oryza sativa*). PloS One 10(3): e0119873.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Statist. Soc. B 57:289-300.

Bentley, A.R., M. Scutari, N. Gosman, S. Faure, F. Bedford, P. Howell, J. Cockram, G.A. Rose, T. Barber, J. Irigoyen, R. Horsnell, C. Pumfrey, E. Winnie, J. Schacht, K. Beauchêne, S. Praud, A. Greenland, D. Balding, and I.J. Mackay. 2014. Applying association mapping and genomic selection to the dissection of key traits in elite European wheat. Theor. Appl. Genet. 127:2619-2633.

Bernardo, R. 2008. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. Crop Sci. 48:1649-1664.

Bernardo, R. 2009. Genomewide selection for rapid introgression of exotic germplasm in maize. Crop Sci. 49:419-425.

Bernardo, R. 2014. Genomewide selection when major genes are known. Crop Sci. 54:68-75.

Bernardo, R., and J. Yu. 2007. Prospects for genomewide selection for quantitative traits in maize. Crop Sci. 47:1082-1090.

Beyene, Y., K. Semagn, S. Mugo, A. Tarekegne, R. Babu, B. Meisel, P. Sehabiague, D. Makumbi, C. Magorokosho, S. Oikeh, J. Gakunga, M. Vargas, M. Olsen, B.M. Prasanna, M. Banziger, and J. Crossa. 2015. Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. Crop Sci. 55:154-163.

Börner, A., M. Röder, and V. Korzun. 1997. Comparative molecular mapping of GA insensitive *Rht* loci on chromosomes 4B and 4D of common wheat (*Triticum aestivum* L.). Theor. Appl. Genet. 95:1133-1137.

Börner, A., J. Plaschke, V. Korzun, and A.J. Worland. 1996. The relationships between the dwarfing genes of wheat and rye. Euphytica 89:69-75.

Botstein, D., R.L. White, M. Skolnick, and R.W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32:314-331.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, E.S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635.

Breiman, A. and Graur, D. 1995. Wheat evolution. Israel Journal of Plant Sciences. 43:85-98.

Breseghello, F., and M.E. Sorrells. 2006a. Association analysis as a strategy for improvement of quantitative traits in plants. Crop Sci. 46:1323-1330.

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

Burgueño, J., G. de los Campos, K. Weigel, and J. Crossa. 2012. Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. Crop Sci. 52:707-719.

Burgueño, J., J. Crossa, J.M. Cotes, F.S. Vicente, and B. Das. 2011. Prediction assessment of linear mixed models for multienvironment trials. Crop Sci. 51:944-954.

Burstin, J., P. Salloignon, M. Chabert-Martinello, J.-B Magnin-Robert, M. Siol, F. Jacquin, A. Chauveau, C. Pont, G. Aubert, C. Delaitre, C. Truntzer, and G. Duc. 2015. Genetic diversity and trait genomic prediction in a pea diversity panel. BMC Genomics 16:105.

Bushuk, W., K.G. Briggs, and L.H. Shebeski. 1969. Protein quantity and quality as factors in the evaluation of bread wheats. Canadian Journal of Plant Science 49:113-122.

Calus, M. 2010. Genomic breeding value prediction: Methods and procedures. Animal 4:157-164.

Calus, M.P.L., and R.F. Veerkamp. 2007. Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. J. Anim. Breed. Genet. 124:362-368.

Calus, M.P.L., T.H.E. Meuwissen, A.P.W. de Roos, and R.F. Veerkamp. 2008. Accuracy of genomic selection using different methods to define haplotypes. Genetics 178:553-561.

Calus, M.P.L., and R.F. Veerkamp. 2011. Accuracy of multi-trait genomic selection using different methods. Genet. Sel. Evol. 43:26.

Camus-Kulandaivelu, L., J.-B. Veyrieras, B. Gouesnard, A. Charcosset, and D. Manicacci. 2007. Evaluating the reliability of outputs in case of relatedness between individuals. Crop Sci. 47:887-892.

Canadian Grain Commission, 2015. Canadian Wheat Classes. Government of Canada. https://www.grainscanada.gc.ca/wheat-ble/classes/classes-eng.htm (accessed 26 July 2016).

Cavanagh, C.R., S. Chao, S. Wang, B.E. Huang, S. Stephen, S. Kiani, K. Forrest, C. Saintenac, G.L. Brown-Guedira, A. Akhunova, D. See, G. Bai, M. Pumphrey, L. Tomar, D. Wong, S. Kong, M. Reynolds, M.L. da Silva, H. Bockelman, L. Talbert, J.A. Anderson, S. Dreisigacker, S. Baenziger, A. Carter, V. Korzun, P.L. Morrell, J. Dubcovsky, M.K. Morell, M.E. Sorrells, M.J. Hayden, and E. Akhunov. 2013. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc. Natl. Acad. Sci. U. S. A. 110:8057-8062.

Chao, S., P.J. Sharp, A.J. Worland, E.J. Warham, R.M.D. Koebner, and M.D. Gale. 1989. RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. Theor. Appl. Genet. 78:495-504.

Charmet, G., and E. Storlie. 2012. Implementation of genome-wide selection in wheat. Russian Journal of Genetics: Applied Research 2:298-303.

Charmet, G., E. Storlie, F.X. Oury, V. Laurent, D. Beghin, L. Chevarin, A. Lapierre, M.R. Perretant, B. Rolland, E. Heumez, L. Duchalais, E. Goudemand, J. Bordes, and O. Robert. 2014. Genome-wide prediction of three important traits in bread wheat. Mol. Breed. 34:1843-1852.

Chee, M., R. Yang, E. Hubbell, A. Berno, X.C. Huang, D. Stern, J. Winkler, D.J. Lockhart, M.S. Morris, and S.P.A. Fodor. 1996. Accessing genetic information with high-density DNA arrays. Science 274:610-614.

Chen, H., M. Iqbal, E. Perez-Lara, R.-C. Yang, C. Pozniak, and D. Spaner. 2015. Earliness *per se* quantitative trait loci and their interaction with *Vrn-B1* locus in a spring wheat population. Mol. Breed. 35:182.

CIMMYT. 2005. Laboratory Protocols: CIMMYT Applied Molecular Genetics Laboratory. Third Edition. Mexico, D.F.: CIMMYT.

Clark, S.A., J.M. Hickey, and J.H.J. van der Werf. 2011. Different models of genetic variation and their effect on genomic evaluation. Genetics Selection Evolution 43:18.

Clark, S.A., J.M. Hickey, H.D. Daetwyler, and J.H.J. van der Werf. 2012. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. Genetics Selection Evolution 44:4.

Clark, S.A. and J. van der Werf. 2013. Genomic best linear unbiased prediction (gBLUP) for the estimation of genomic breeding values. In: C. Gondro et al., editors, Genome-wide association

studies and genomic prediction. Humana Press. New York. p. 321-330. doi:10.1007/978-1-62703-447-0.

Cobb, J.N., G. DeClerck, A. Greenberg, R. Clark, and S. McCouch. 2013. Next-generation phenotyping: Requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. Theor. Appl. Genet. 126:867-887.

Collard, B.C.Y., M.Z.Z. Jahufer, J.B. Brouwer, and E.C.K. Pang. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142:169-196.

Collard, B.C.Y., and D.J. Mackill. 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 363:557-572.

Combs, E., and R. Bernardo. 2013. Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers. The Plant Genome 6. doi:10.3835/plantgenome2012.11.0030.

Crossa, J., P. Pérez, J. Hickey, J. Burgueño, L. Ornella, J. Cerón-Rojas, X. Zhang, S. Dreisigacker, R. Babu, Y. Li, D. Bonnett, and K. Mathews. 2014. Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity 112:48-60.

Crossa, J. 2012. From genotype × environment interaction to gene × environment interaction. Curr. Genomics 13:225-244.

Crossa, J., R.-C. Yang, and P.L. Cornelius. 2004. Studying crossover genotype × environment interaction using linear-bilinear models and mixed models. Journal of Agricultural, Biological, and Environmental Statistics 9:362-380.

Crossa, J., G. de los Campos, M. Maccaferri, R. Tuberosa, J. Burgueño, and P. Pérez-Rodríguez. 2015. Extending the marker × environment interaction model for genomic-enabled prediction and genome-wide association analysis in durum wheat. Crop Sci. 56:1-17.

Crossa, J., L. G. de los Campos, P. Pérez, D. Gianola, J. Burgueño, J.L. Araus, D. Makumbi, R.P. Singh, S. Dreisigacker, J. Yan, V. Arief, M. Banziger, and H.-J. Braun. 2010. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186:713-724.

Crossa, J., D. Jarquín, J. Franco, P. Pérez-Rodríguez, J. Burgueño, C. Saint-Pierre, P. Vikram, C. Sansaloni, C. Petroli, D. Akdemir, C. Sneller, M. Reynolds, M. Tattaris, T. Payne, C. Guzman, R.J. Peña, P. Wenzl, and S. Singh. 2016. Genomic prediction of gene bank wheat landraces. G3 (Bethesda). 6:1819-1834.

Cuevas, J., J. Crossa, O.A. Montesinos-López, J. Burgueño, P. Pérez-Rodríguez, and G. de los Campos. 2017. Bayesian genomic prediction with genotype × environment interaction kernel models. G3 (Bethesda). 7:41-53.

Cuthbert, J.L., D.J. Somers, A.L. Brûlé-Babel, P.D. Brown, and G.H. Crow. 2008. Molecular mapping of quantitative trait loci for yield and yield components in spring wheat (*Triticum aestivum* L.). Theor. Appl. Genet. 117:595-608.

Cuthbert, P.A., D.J. Somers, J. Thomas, S. Cloutier, and A. Brulé-Babel. 2006. Fine mapping *Fhb1*, a major gene controlling fusarium head blight resistance in bread wheat (*Triticum aestivum* L.). Theor. Appl. Genet. 112:1465-1472.

Daetwyler, H.D., U.K. Bansal, H.S. Bariana, M.J. Hayden, and B.J. Hayes. 2014. Genomic prediction for rust resistance in diverse wheat landraces. Theor. Appl. Genet. 127:1795-1803.

Daetwyler, H.D., R. Pong-Wong, B. Villanueva, and J.A. Woolliams. 2010. The impact of genetic architecture on genome-wide evaluation methods. Genetics 185:1021-1031.

Daetwyler, H.D., M.P.L. Calus, R. Pong-Wong, G. de los Campos, and J.M. Hickey. 2013. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. Genetics 193:347-365.

Dawson, J.C., J.B. Endelman, N. Heslot, J. Crossa, J. Poland, S. Dreisigacker, Y. Manès, M.E. Sorrells, and J.-L. Jannink. 2013. The use of unbalanced historical data for genomic selection in an international wheat breeding program. Field Crops Res. 154:12-22.

de los Campos, G., D. Gianola, and G.J.M. Rosa. 2009a. Reproducing kernel Hilbert spaces regression: a general framework for genetic evaluation. J. Anim. Sci. 87:1883-1887.

de los Campos, G., D. Gianola, G.J.M. Rosa, K.A. Weigel, and J. Crossa. 2010. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. Genet. Res., Camb. 92:295-308.

de los Campos, G., H. Naya, D. Gianola, J. Crossa, A. Legarra, E. Manfredi, K. Weigel, and J.M. Cotes. 2009b. Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genetics 182:375-385.

de los Campos, G., J.M. Hickey, R. Pong-Wong, H.D. Daetwyler, and M.P.L. Calus. 2013. Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics 193:327-345.

de los Campos, G., Y. Veturi, A.I. Vazquez, C. Lehermeier, and P. Pérez-Rodríguez. 2015. Incorporating genetic heterogeneity in whole-genome regressions using interactions. Journal of Agricultural, Biological, and Environmental Statistics 20:467-490.

de Oliveira, E.J., M.D.V. de Resende, V. da Silva Santos, C.F. Ferreira, G.A.F. Oliveira, M.S. da Silva, L.A. de Oliveira, and C.I. Aguilar-Vildoso. 2012. Genome-wide selection in cassava. Euphytica 187:263-276.

de Roos, A.P.W., B.J. Hayes, R.J. Spelman, and M.E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. Genetics 179:1503-1512.

de Roos, A.P.W., B.J. Hayes, and M.E. Goddard. 2009. Reliability of genomic predictions across multiple populations. Genetics 183:1545-1553.

Denis, M., and J.-M. Bouvet. 2013. Efficiency of genomic selection with models including dominance effect in the context of *Eucalyptus* breeding. Tree Genetics & Genomes 9:37-51.

DePauw, R.M., R.E. Knox, D.G. Humphreys, J.B. Thomas, S.L. Fox, P.D. Brown, A.K. Singh, C. Pozniak, H.S. Randhawa, D.B. Fowler, R.J. Graf, and P. Hucl. 2011. New breeding tools impact Canadian commercial farmer fields. Czech J. Genet. Plant Breed. 47: S28-S34.

Desta, Z.A., and R. Ortiz. 2014. Genomic selection: Genome-wide prediction in plant improvement. Trends Plant Sci. 19:592-601.

Doerge, R.W. 2002. Mapping and analysis of quantitative trait loci in experimental populations. Nat. Rev. Genet. 3:43-52.

Dubcovsky, J., D. Lijavetzky, L. Appendino, and G. Tranquilli. 1998. Comparative RFLP mapping of *Triticum monococcum* genes controlling vernalization requirement. Theor. Appl. Genet. 97:968-975.

Dvorak, J., P.E. McGuire, and B. Cassidy. 1988. Apparent sources of the A genomes of wheats inferred from polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. Genome. 30: 680-689.

Dvorák, J. and H.-B. Zhang. 1990. Variation in repeated nucleotide sequences sheds light on the phylogeny of the wheat B and G genomes. Proc. Natl. Acad. Sci. USA. 87: 9640-9644.

Earl, D.A. and B.M. vonHoldt. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the evanno method. Conservation Genetics Resources 4:359-361.

Ellis, M.H., W. Spielmeyer, K.R. Gale, G.J. Rebetzke, and R.A. Richards. 2002. " Perfect" markers for the *Rht-B1b* and *Rht-D1b* dwarfing genes in wheat. Theor. Appl. Genet. 105:1038-1042.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS One 6(5): e19379.

Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. The Plant Genome 4:250-255.

Endelman, J. 2013. Genomic prediction with rrBLUP 4. https://potatobreeding.cals.wisc.edu/wp-content/uploads/sites/21/2014/01/GS_tutorial.pdf (accessed 20 Feb. 2015).

Environment Canada. 2016. Historical data. Government of Canada. http://climate.weather.gc.ca/historical_data/search_historic_data_e.html (accessed 02 Feb. 2016).

Erbe, M., B.J. Hayes, L.K. Matukumalli, S. Goswami, P.J. Bowman, C.M. Reich, B.A. Mason, and M.E. Goddard. 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J. Dairy Sci. 95:4114-4129.

FAOSTAT, 2015. Crops production. Food and Agricultural Organization of the United Nations. http://faostat.fao.org (accessed 25 Dec. 2015).

Federer, W.T. 1961. Augmented designs with one-way elimination of heterogeneity. Biometrics 17:447-473.

Flint-Garcia, S.A., J.M. Thornsberry, and E.S. Buckler IV. 2003. Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. 54:357-374.

Francia, E., G. Tacconi, C. Crosatti, D. Barabaschi, D. Bulgarelli, E. Dall'Aglio, and G. Vale. 2005. Marker assisted selection in crop plants. Plant Cell, Tissue and Organ Culture 82:317-342.

Gianola, D., R.L. Fernando, and A. Stella. 2006. Genomic-assisted prediction of genetic value with semiparametric procedures. Genetics 173:1761-1776.

Gianola, D., H. Okut, K.A. Weigel, and G.J.M. Rosa. 2011. Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. BMC Genet. 12:87.

Gianola, D., G. de los Campos, W.G. Hill, E. Manfredi, and R. Fernando. 2009. Additive genetic variability and the bayesian alphabet. Genetics 183:347-363.

Gianola, D. 2013. Priors in whole-genome regression: the Bayesian alphabet returns. Genetics 194:573-596.

Gianola, D., and J.B.C.H.M. van Kaam. 2008. Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. Genetics 178:2289-2303.

Goddard, M.E. 2012. Uses of genomics in livestock agriculture. Animal Production Science 52:73-77.

Goddard, M.E., B.J. Hayes, and T.H.E. Meuwissen. 2011. Using the genomic relationship matrix to predict the accuracy of genomic selection. J. Anim. Breed. Genet. 128:409-421.

Goddard, M.E., and B.J. Hayes. 2007. Genomic selection. J. Anim. Breed. Genet. 124:323-330.

Gouy, M., Y. Rousselle, D. Bastianelli, P. Lecomte, L. Bonnal, D. Roques, J.-C. Efile, S. Rocher, J. Daugrois, L. Toubi, S. Nabeneza, C. Hervouet, H. Telismart, M. Denis, A. Thong-Chane, J.C. Glaszmann, J.-Y. Hoarau, S. Nibouche, and L. Costet. 2013. Experimental assessment of the accuracy of genomic selection in sugarcane. Theor. Appl. Genet. 126:2575-2586.

Grenier, C., T.-V. Cao, Y. Ospina, C. Quintero, M.H. Châtel, J. Tohme, B. Courtois, and N. Ahmadi. 2015. Accuracy of genomic selection in a rice synthetic population developed for recurrent selection breeding. PloS One 10(8): e0136594.

Guo, G., F. Zhao, Y. Wang, Y. Zhang, L. Du, and G. Su. 2014a. Comparison of single-trait and multiple-trait genomic prediction models. BMC Genet. 15:30.

Guo, Z., D.M. Tucker, C.J. Basten, H. Gandhi, E. Ersoz, B. Guo, Z. Xu, D. Wang, and G. Gay. 2014b. The impact of population structure on genomic prediction in stratified populations. Theor. Appl. Genet. 127:749-762.

Gupta, P.K., R.K. Varshney, P.C. Sharma, and B. Ramesh. 1999. Molecular markers and their applications in wheat breeding. Plant Breeding 118:369-390.

Gustafson, P., O. Raskina, X. Ma, and E. Nevo. 2009. Wheat evolution, domestication, and improvement. In: B.F. Carver, editor, Wheat Science and trade. Wiley-Blackwell. Iowa. p. 5-30.

Habier, D., R.L. Fernando, K. Kizilkaya, and D.J. Garrick. 2011. Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics 12:186.

Habier, D., J. Tetens, F.-R. Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genetics Selection Evolution 42:5.

Habier, D., R.L. Fernando, and D.J. Garrick. 2013. Genomic BLUP decoded: a look into the black box of genomic prediction. Genetics 194:597-607.

Habier, D., R.L. Fernando, and J.C.M. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389-2397.

Hartigan, J.A., and M.A. Wong. 1979. Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics) 28:100-108.

Hayashi, T., and H. Iwata. 2010. EM algorithm for bayesian estimation of genomic breeding values. BMC Genetics 11:3.

Hayashi, T., and H. Iwata. 2013. A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits. BMC Bioinformatics 14:34.

Hayes, B.J., J. Panozzo, C.K. Walker, A.L. Choy, S. Kant, D. Wong, J. Tibbits, H.D. Daetwyler, S. Rochfort, M.J. Hayden, and G.C. Spangenberg. 2017. Accelerating wheat breeding for end-use quality with multi-trait genomic predictions incorporating near infrared and nuclear magnetic resonance-derived phenotypes. Theor. Appl. Genet. (2017). doi:10.1007/s00122-017-2972-7.

Hayes, B.J., P.J. Bowman, A.C. Chamberlain, K. Verbyla, and M.E. Goddard. 2009a. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genetics Selection Evolution 41:51.

Hayes, B.J., P.M. Visscher, and M.E. Goddard. 2009b. Increased accuracy of artificial selection by using the realized relationship matrix. Genet. Res., Camb. 91:47-60.

Hayes, B.J., P.J. Bowman, A.J. Chamberlain, and M.E. Goddard. 2009c. Invited review: Genomic selection in dairy cattle: progress and challenges. J. Dairy Sci. 92:433-443.

He, J., X. Zhao, A. Laroche, Z. Lu, H. Liu, and Z. Li. 2014. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front. Plant Sci. 5:484.

He, S., A.W. Schulthess, V. Mirdita, Y. Zhao, V. Korzun, R. Bothe, E. Ebmeyer, J.C. Reif, and Y. Jiang. 2016. Genomic selection in a commercial winter wheat population. Theor. Appl. Genet. 129:641-651.

Heffner, E.L., J.-L. Jannink, and M.E. Sorrells. 2011a. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. The Plant Genome 4:65-75.

Heffner, E.L., M.E. Sorrells, and J.-L. Jannink. 2009. Genomic selection for crop improvement. Crop Sci. 49:1-12.

Heffner, E.L., A.J. Lorenz, J.-L. Jannink, and M.E. Sorrells. 2010. Plant breeding with genomic selection: gain per unit time and cost. Crop Sci. 50:1681-1690.

Heffner, E.L., J.-L. Jannink, H. Iwata, E. Souza, and M.E. Sorrells. 2011b. Genomic selection accuracy for grain quality traits in biparental wheat populations. Crop Sci. 51:2597-2606.

Helentjaris, T., G. King, M. Slocum, C. Siedenstrang, and S. Wegman. 1985. Restriction fragment polymorphisms as probes for plant diversity and their development as tools for applied plant breeding. Plant Mol. Biol. 5:109-118.

Heslot, N., H.-P. Yang, M.E. Sorrells, and J.-L. Jannink. 2012. Genomic selection in plant breeding: a comparison of models. Crop Sci. 52:146-160.

Heslot, N., D. Akdemir, M.E. Sorrells, and J.-L. Jannink. 2014. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. Theor. Appl. Genet. 127:463-480.

Holland, J.B. 2007. Genetic architecture of complex traits in plants. Curr. Opin. Plant Biol. 10:156-161.

Howey, R. and H.J. Cordell. 2011. MapThin. https://www.staff.ncl.ac.uk/richard.howey/mapthin/ (accessed 17 Mar. 2017).

Huang, M., A. Cabrera, A. Hoffstetter, C. Griffey, D.V. Sanford, J. Costa, A. McKendry, S. Chao, and C. Sneller. 2016. Genomic selection for wheat traits and trait stability. Theor. Appl. Genet. 129: 1697-1710.

Huang, X.Q., S. Cloutier, L. Lycar, N. Radovanovic, D.G. Humphreys, J.S. Noll, D.J. Somers, and P.D. Brown. 2006. Molecular detection of QTLs for agronomic and quality traits in a doubled haploid population derived from two Canadian wheats (*Triticum aestivum* L.). Theor. Appl. Genet. 113:753-766.

International Wheat Genome Sequencing Consortium. 2014. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345:1251788.

Isidro, J., J.-L. Jannink, D. Akdemir, J. Poland, N. Heslot, and M.E. Sorrells. 2015. Training set optimization under population structure in genomic selection. Theor. Appl. Genet. 128:145-158.

Iwata, H., and J.-L. Jannink. 2011. Accuracy of genomic selection prediction in barley breeding programs: a simulation study based on the real single nucleotide polymorphism data of barley breeding lines. Crop Sci. 51:1915-1927.

Jakobsson, M., and N.A. Rosenberg. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801-1806.

Jannink, J.-L., A.J. Lorenz, and H. Iwata. 2010. Genomic selection in plant breeding: from theory to practice. Brief. Funct. Genomics 9:166-177.

Jarquín, D., C.L. da Silva, R.C. Gaynor, J. Poland, A. Fritz, R. Howard, S. Battenfield, and J. Crossa. 2017. Increasing genomic-enabled prediction accuracy by modeling genotype × environment interactions in Kansas wheat. The Plant Genome 10. doi:10.3835/plantgenome2016.12.0130.

Jarquín, D., J. Crossa, X. Lacaze, P.D. Cheyron, J. Daucourt, J. Lorgeou, F. Piraux, L. Guerreiro, P. Pérez, M. Calus, J. Burgueño, and G. de los Campos. 2014a. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor. Appl. Genet. 127:595-607.

Jarquín, D., K. Kocak, L. Posadas, K. Hyma, J. Jedlicka, G. Graef, and A. Lorenz. 2014b. Genotyping by sequencing for genomic prediction in a soybean breeding population. BMC Genomics 15:740.

Jia, Y., and J.-L. Jannink. 2012. Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics 192:1513-1522.

Jiang, J., Q. Zhang, L. Ma, J. Li, Z. Wang, and J-F. Liu. 2015. Joint prediction of multiple quantitative traits using a Bayesian multivariate antedependence model. Heredity 115:29-36.

Jiang, Y., and J.C. Reif. 2015. Modeling epistasis in genomic selection. Genetics 201:759-768.

Kamran, A., M. Iqbal, A. Navabi, H. Randhawa, C. Pozniak, and D. Spaner. 2013. Earliness *per se* QTLs and their interaction with the photoperiod insensitive allele *Ppd-D1a* in the Cutler × AC Barrie spring wheat population. Theor. Appl. Genet. 126:1965-1976.

Kärkkäinen, H.P., and M.J. Sillanpää. 2012. Back to basics for Bayesian model building in genomic selection. Genetics 191:969-987.

Kibite, S., and L.E. Evans. 1984. Causes of negative correlations between grain yield and grain protein concentration in common wheat. Euphytica 33:801-810.

Knox, R.E., C.J. Pozniak, F.R. Clarke, J.M. Clarke, S. Houshmand, and A.K. Singh. 2009. Chromosomal location of the cadmium uptake gene (*Cdu1*) in durum wheat. Genome 52:741-747.

Koebner, R.M.D., and R.W. Summers. 2003. 21st century wheat breeding: plot selection or plate detection? Trends Biotechnol. 21:59-63.

Košner, J., and K. Pánková. 1998. The detection of allelic variants at the recessive *vrn* loci of winter wheat. Euphytica 101:9-16.

Kumar, S., M.C.A.M. Bink, R.K. Volz, V.G.M. Bus, and D. Chagné. 2012a. Towards genomic selection in apple (*Malus*×*domestica* Borkh.) breeding programmes: Prospects, challenges and strategies. Tree Genetics & Genomes 8:1-14.

Kumar, S., D. Chagné, M.C.A.M. Bink, R.K. Volz, C. Whitworth, and C. Carlisle. 2012b. Genomic selection for fruit quality traits in apple (*Malus*×*domestica* Borkh.). PloS One 7(5): e36674.

Kunert, A., A.A. Naz, O. Dedeck, K. Pillen, and J. Léon. 2007. AB-QTL analysis in winter wheat: I. synthetic hexaploid wheat (*T. turgidum* ssp. *dicoccoides* × *T. tauschii*) as a source of favourable alleles for milling and baking quality traits. Theor. Appl. Genet. 115:683-695.

Lado, B., P.G. Barrios, M. Quincke, P. Silva, and L. Gutiérrez. 2016. Modeling genotype × environment interaction for genomic selection with unbalanced data from a wheat breeding program. Crop Sci. 56:1-15.

Lander, E.S., and D. Botstein. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185-199.

Langridge, P., E.S. Lagudah, T.A. Holton, R. Appels, P.J. Sharp, and K.J. Chalmers. 2001. Trends in genetic and genome analyses in wheat: a review. Aust. J. Agric. Res. 52:1043-1077.

Law, C.N., A.J. Worland, and B. Giorgi. 1976. The genetic control of ear-emergence time by chromosomes 5A and 5D of wheat. Heredity 36:49-58.

Letunic, I., and P. Bork. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44: W242-W245.

Li, H., J. Wang, and Z. Bao. 2015. A novel genomic selection method combining GBLUP and LASSO. Genetica 143:299-304.

Lin, F., S.L. Xue, D.G. Tian, C.J. Li, Y. Cao, Z.Z. Zhang, C.Q. Zhang, and Z.Q. Ma. 2008. Mapping chromosomal regions affecting flowering time in a spring wheat RIL population. Euphytica 164:769-777.

Lipka, A.E., F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury, M.A. Gore, E.S. Buckler, and Z. Zhang. 2012. GAPIT: Genome association and prediction integrated tool. Bioinformatics 28:2397-2399.

Liu, Y., Z. He, R. Appels, and X. Xia. 2012. Functional markers in wheat: current status and future prospects. Theor. Appl. Genet. 125:1-10.

Long, N., D. Gianola, G.J.M. Rosa, and K.A. Weigel. 2011. Application of support vector regression to genome-assisted prediction of quantitative traits. Theor. Appl. Genet. 123:1065-1074.

Longin, C.F.H., X. Mi, and T. Würschum. 2015. Genomic selection in wheat: optimum allocation of test resources and comparison of breeding strategies for line and hybrid breeding. Theor. Appl. Genet. 128:1297-1306.

Lorenz, A.J., S. Chao, F.G. Asoro, E.L. Heffner, T. Hayashi, H. Iwata, K.P. Smith, M.E. Sorrells, and J.-L. Jannink. 2011. Genomic selection in plant breeding: knowledge and prospects. Vol. 110. In: D.L. Sparks, editor, Advances in agronomy. Elsevier. San Diego. p. 77-123. doi:10.1016/B978-0-12-385531-2.00002-5.

Lopez-Cruz, M., J. Crossa, D. Bonnett, S. Dreisigacker, J. Poland, J.-L. Jannink, R.P. Singh, E. Autrique, and G. de los Campos. 2015. Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. G3 (Bethesda). 5:569-582.

Lorenzana, R.E., and R. Bernardo. 2009. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor. Appl. Genet. 120:151-161.

Luan, T., J.A. Woolliams, S. Lien, M. Kent, M. Svendsen, and T.H.E. Meuwissen. 2009. The accuracy of genomic selection in Norwegian red cattle assessed by cross-validation. Genetics 183:1119-1126.

Ly, D., M. Hamblin, I. Rabbi, G. Melaku, M. Bakare, H.G. Gauch Jr., R. Okechukwu, A.G.O. Dixon, P. Kulakow, and J.-L. Jannink. 2013. Relatedness and genotype × environment interaction affect prediction accuracies in genomic selection: a study in cassava. Crop Sci. 53:1312-1325.

Mackay, T.F.C. 2001. The genetic architecture of quantitative traits. Annu. Rev. Genet. 35:303-339.

Mammadov, J., R. Aggarwal, R. Buyyarapu, and S. Kumpatla. 2012. SNP markers and their impact on plant breeding. Int. J. Plant. Genomics 2012:728398.

Marroni, F., S. Pinosio, G. Zaina, F. Fogolari, N. Felice, F. Cattonaro, and M. Morgante. 2011. Nucleotide diversity and linkage disequilibrium in *Populus nigra* cinnamyl alcohol dehydrogenase (*CAD4*) gene. Tree Genetics & Genomes 7:1011-1023.

McCallum, B.D., and R.M. DePauw. 2008. A review of wheat cultivars grown in the Canadian prairies. Can. J. Plant Sci. 88:649-677.

McCartney, C.A., D.J. Somers, D.G. Humphreys, O. Lukow, N. Ames, J. Noll, S. Cloutier, and B.D. McCallum. 2005. Mapping quantitative trait loci controlling agronomic traits in the spring wheat cross RL4452 × 'AC domain'. Genome 48:870-883.

McFadden, E.S. and E.R. Sears. 1946. The origin of *Triticum spelta* and its free-threshing hexaploid relatives. J. Hered. 37:81-89.

Mendes, M.P., and C.L. de Souza Jr. 2016. Genomewide prediction of tropical maize single-crosses. Euphytica 209:651-663.

Meng, L., H. Li, L. Zhang, and J. Wang. 2015. QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. The Crop Journal 3:269-283.

Meuwissen, T. 2007. Genomic selection: marker assisted selection on a genome wide scale. J. Anim. Breed. Genet. 124:321-322.

Meuwissen, T.H.E. 2009. Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. Genetics Selection Evolution 41:35.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819-1829.

Meuwissen, T., B. Hayes, and M. Goddard. 2016. Genomic selection: a paradigm shift in animal breeding. Animal Frontiers 6:6-14.

Michel, S., C. Ametz, H. Gungor, D. Epure, H. Grausgruber, F. Löschenberger, and H. Buerstmayr. 2016. Genomic selection across multiple breeding cycles in applied bread wheat breeding. Theor. Appl. Genet. 129:1179-1189.

Miedaner, T., Y. Zhao, M. Gowda, C.F.H. Longin, V. Korzun, E. Ebmeyer, E. Kazman, and J.C. Reif. 2013. Genetic architecture of resistance to septoria tritici blotch in European wheat. BMC Genomics 14:858.

Mohan, M., S. Nair, A. Bhagwat, T.G. Krishna, M. Yano, C.R. Bhatia, and T. Sasaki. 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. Mol. Breed. 3:87-103.

Morota, G., and D. Gianola. 2014. Kernel-based whole-genome prediction of complex traits: a review. Front. Genet. 5:363

Moser, G., M.S. Khatkar, B.J. Hayes, and H.W. Raadsma. 2010. Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. Genet. Sel. Evol. 42:37.

Muir, W.M. 2007. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J. Anim. Breed. Genet. 124:342-355.

Mutshinda, C.M., and M.J. Sillanpää. 2010. Extended Bayesian LASSO for multiple quantitative trait loci mapping and unobserved phenotype prediction. Genetics 186:1067-1075.

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

Nevo, E., A.B. Korol, A. Beiles, and T. Fahima. 2002. Domestication of wheats. In: E. Nevo et al., editors, Evolution of wild emmer and wheat improvement: population genetics, genetic

resources, and genome organization of wheat's progenitor, *Triticum dicoccoides*. Springer-Verlag Berlin Heidelberg. Berlin, Germany. p. 3-10.

Onogi, A., M. Watanabe, T. Mochizuki, T. Hayashi, H. Nakagawa, T. Hasegawa, and H. Iwata. 2016. Toward integration of genomic selection with crop modelling: the development of an integrated approach to predicting rice heading dates. Theor. Appl. Genet. 129:805-817.

Ornella, L., S. Singh, P. Perez, J. Burgueño, R. Singh, E. Tapia, S. Bhavani, S. Dreisigacker, H.-J. Braun, K. Mathews, and J. Crossa. 2012. Genomic prediction of genetic values for resistance to wheat rusts. The Plant Genome 5:136-148.

Park, T., and G. Casella. 2008. The Bayesian lasso. J. Am. Stat. Assoc. 103:681-686.

Paterson, A.H. 1996. Making genetic maps. In: A.H. Paterson, editor, Genome Mapping in Plants. R.G. Landes Company. Austin, Texas. p. 23-39.

Paterson, A.H., E.S. Lander, J.D. Hewitt, S. Peterson, S.E. Lincoln, and S.D. Tanksley. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721-726.

Payne, P.I. 1987. Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. Ann. Rev. Plant Physiol. 38:141-153.

Peng, J. H., Sun, D. and Nevo, E. 2011. Domestication evolution, genetics and genomics in wheat. Mol Breeding. 28:281-301.

Pérez, P., and G. de los Campos. 2014. Genome-wide regression and prediction with the BGLR statistical package. Genetics 198:483-495.

Pérez, P., G. de los Campos, J. Crossa, and D. Gianola. 2010. Genomic-enabled prediction based on molecular markers and pedigree using the Bayesian linear regression package in R. The Plant Genome 3:106-116.

Perez-Lara, E., K. Semagn, H. Chen, M. Iqbal, A. N'Diaye, A. Kamran, A. Navabi, C. Pozniak, and D. Spaner. 2016. QTLs associated with agronomic traits in the Cutler × AC Barrie spring wheat mapping population using single nucleotide polymorphic markers. PloS One 11(8): e0160623.

Pérez-Rodríguez, P., J. Crossa, K. Bondalapati, G.D. Meyer, F. Pita, and G. de los Campos. 2015. A pedigree-based reaction norm model for prediction of cotton yield in multienvironment trials. Crop Sci. 55:1143-1151.

Pérez-Rodríguez, P., D. Gianola, J.M. González-Camacho, J. Crossa, Y. Manès, and S. Dreisigacker. 2012. Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. G3 (Bethesda). 2:1595-1605.

Perrier X. and J.P. Jacquemoud-Collet. 2006. DARwin software http://darwin.cirad.fr/.

Petersen, G., O. Seberg, M. Yde, and K. Berthelsen. 2006. Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). Mol. Phylogenet. Evol. 39:70–82.

Poland, J., J. Endelman, J. Dawson, J. Rutkoski, S. Wu, Y. Manes, S. Dreisigacker, J. Crossa, H. Sánchez-Villeda, M. Sorrells, and J.-L. Jannink. 2012. Genomic selection in wheat breeding using genotyping-by-sequencing. The Plant Genome 5:103-113.

Price, A.L., N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38:904-909.

Pritchard, J.K., and N.A. Rosenberg. 1999. Use of unlinked genetic markers to detect population stratification in association studies. Am. J. Hum. Genet. 65:220-228.

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, D. Bender, J. Maller, P. Sklar, P.I.W. de Bakker, M.J. Daly, and P.C. Sham. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559-575.

Rafalski, J.A. 2002. Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant Science 162:329-333.

Ramasamy, R.K., S. Ramasamy, B.B. Bindroo, and V.G. Naik. 2014. STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. SpringerPlus 3:431.

Randhawa, H.S., M. Asif, C. Pozniak, J.M. Clarke, R.J. Graf, S.L. Fox, D.G. Humphreys, R.E. Knox, R.M. DePauw, A.K. Singh, R.D. Cuthbert, P. Hucl, and D. Spaner. 2013. Application of molecular markers to wheat breeding in Canada. Plant Breeding 132:458-471.

R Development Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org.

Resende, M.D.V., M.F.R. Resende, Jr., C.P. Sansaloni, C.D. Petroli, A.A. Missiaggia, A.M. Aguiar, J.M. Abad, E.K. Takahashi, A.M. Rosado, D.A. Faria, G.J. Pappas, Jr., A. Kilian, and D. Grattapaglia. 2012a. Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. New Phytol. 194:116-128.

Resende Jr., M.F.R., P. Muñoz, M.D.V. Resende, D.J. Garrick, R.L. Fernando, J.M. Davis, E.J. Jokela, T.A. Martin, G.F. Peter, and M. Kirst. 2012b. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). Genetics 190:1503-1510.

Riedelsheimer, C., F. Technow, and A.E. Melchinger. 2012. Comparison of whole-genome prediction models for traits with contrasting genetic architecture in a diversity panel of maize inbred lines. BMC Genomics 13:452.

Riedelsheimer, C., J.B. Endelman, M. Stange, M.E. Sorrells, J.-L. Jannink, and A.E. Melchinger. 2013. Genomic predictability of interconnected biparental maize populations. Genetics 194:493-503.

Riley, R. 1975. Origins of wheat. In: A. Spicer, editor, Bread: social, nutritional and agricultural aspects of wheaten bread. Applied Science Publishers LTD. London, UK. p. 27–45.

Rincent, R., D. Laloë, S. Nicolas, T. Altmann, D. Brunel, P. Revilla, V.M. Rodríguez, J. Moreno-Gonzalez, A. Melchinger, E. Bauer, C-C. Schoen, N. Meyer, C. Giauffret, C. Bauland, P. Jamin, J. Laborde, H. Monod, P. Flament, A. Charcosset, and L. Moreau. 2012. Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). Genetics 192:715-728.

Rincent, R., A. Charcosset, and L. Moreau. 2017. Predicting genomic selection efficiency to optimize calibration set and to assess prediction accuracy in highly structured populations. Theor. Appl. Genet. 130:2231-2247.

Rudd, J.C. 2009. Success in wheat improvement. In: B.F. Carver, editor, Wheat Science and trade. Wiley-Blackwell. Iowa. p. 387-395.

Rutkoski, J., R.P. Singh, J. Huerta-Espino, S. Bhavani, J. Poland, J.L. Jannink, and M.E. Sorrells. 2015. Efficient use of historical data for genomic selection: a case study of stem rust resistance in wheat. The Plant Genome 8. doi:10.3835/plantgenome2014.09.0046.

Rutkoski, J., J. Benson, Y. Jia, G. Brown-Guedira, J.-L. Jannink, and M. Sorrells. 2012. Evaluation of genomic prediction methods for fusarium head blight resistance in wheat. The Plant Genome 5:51-61.

Rutkoski, J.E., E.L. Heffner, and M.E. Sorrells. 2011. Genomic selection for durable stem rust resistance in wheat. Euphytica 179:161-173.

Rutkoski, J.E., J.A. Poland, R.P. Singh, J. Huerta-Espino, S. Bhavani, H. Barbier, M.N. Rouse, J. Jannink, and M.E. Sorrells. 2014. Genomic selection for quantitative adult plant stem rust resistance in wheat. The Plant Genome 7. doi:10.3835/plantgenome2014.02.0006.

Saatchi, M., S.R. Miraei-Ashtiani, A.N. Javaremi, M. Moradi-Shahrebabak, and H. Mehrabani-Yeghaneh. 2010. The impact of information quantity and strength of relationship between training set and validation set on accuracy of genomic estimated breeding values. Afr. J. Biotechnol. 9:438-442.

Saatchi, M., M.C. McClure, S.D. McKay, M.M. Rolf, J. Kim, J.E. Decker, T.M. Taxis, R.H. Chapple, H.R. Ramey, S.L. Northcutt, S. Bauck, B. Woodward, J.C.M. Dekkers, R.L. Fernando, R.D. Schnabel, D.J. Garrick, and J.F. Taylor. 2011. Accuracies of genomic

breeding values in American Angus beef cattle using K-means clustering for cross-validation. Genet. Sel. Evol. 43:40.

Saiki, R.K., D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, and H.A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487-491.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406-425.

Sallam, A.H., J.B. Endelman, J.-L. Jannink, and K.P. Smith. 2015. Assessing genomic selection prediction accuracy in a dynamic barley breeding population. Plant Genome 8. doi:10.3835/plantgenome2014.05.0020.

SAS Institute. 2015. SAS user's guide. Version 9.4. SAS Inst. Cary, NC.

Scarth, R., and C.N. Law. 1984. The control of the day-length response in wheat by the group 2 chromosomes. Zeitschrift Für Pflanzenzüchtung 92:140-150.

Shendure, J., and H. Ji. 2008. Next-generation DNA sequencing. Nat. Biotechnol. 26:1135-1145.

Simons, K., J.A. Anderson, M. Mergoum, J.D. Faris, D.L. Klindworth, S.S. Xu, C. Sneller, J.-B. Ohm, G.A. Hareland, M.C. Edwards, and S. Chao. 2012. Genetic mapping analysis of bread-making quality traits in spring wheat. Crop Sci. 52:2182-2197.

Snape, J.W., K. Butterworth, E. Whitechurch and A.J. Worland. 2001. Waiting for fine times: genetics of flowering time in wheat. Euphytica. 119:185-190.

Solberg, T.R., A.K. Sonesson, J.A. Woolliams, and T.H.E. Meuwissen. 2008. Genomic selection using different marker types and densities. J. Anim. Sci. 86:2447-2454.

Somers, D.J., T. Banks, R. DePauw, S. Fox, J. Clarke, C. Pozniak, and C. McCartney. 2007. Genome-wide linkage disequilibrium analysis in bread wheat and durum wheat. Genome 50:557-567.

Sourdille, P., J.W. Snape, T. Cadalen, G. Charmet, N. Nakata, S. Bernard, and M. Bernard. 2000. Detection of QTLs for heading time and photoperiod response in wheat using a doubled-haploid population. Genome 43:487-494.

Sourdille, P., T. Cadalen, H. Guyomarc'h, J.W. Snape, M.R. Perretant, G. Charmet, C. Boeuf, S. Bernard, and M. Bernard. 2003. An update of the Courtot × Chinese spring intervarietal molecular marker linkage map for the QTL detection of agronomic traits in wheat. TAG Theor. Appl. Genet. 106:530-538.

Spindel, J.E., H. Begum, D. Akdemir, B. Collard, E. Redoña, J.-L. Jannink, and S. McCouch. 2016. Genome-wide prediction models that incorporate *de novo* GWAS are a powerful new tool for tropical rice improvement. Heredity 116:395-408.

Spindel, J., H. Begum, D. Akdemir, P. Virk, B. Collard, E. Redoña, G. Atlin, J.-L. Jannink, and S.R. McCouch. 2015. Genomic selection and association mapping in rice (*Oryza sativa*):

effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. PLoS Genet. 11(2): e1004982.

Statistics Canada, 2017. Estimated areas, yield, production and average farm price of principal field crops, in metric units, annually. http://www.statcan.gc.ca/daily-quotidien/170421/dq170421b-cansim-eng.htm (accessed 22 Apr. 2017).

Sukumaran, S., J. Crossa, D. Jarquín, and M. Reynolds. 2017. Pedigree-based prediction models with genotype × environment interaction in multienvironment trials of CIMMYT wheat. Crop Sci. 57:1865–1880.

Storlie, E., and G. Charmet. 2013. Genomic selection accuracy using historical data generated in a wheat breeding program. The Plant Genome 6. doi:10.3835/plantgenome2013.01.0001.

Technow, F., C.D. Messina, L.R. Totir, and M. Cooper. 2015. Integrating crop growth models with whole genome prediction through approximate Bayesian computation. PLoS One 10(6): e0130855.

Thavamanikumar, S., R. Dolferus, and B.R. Thumma. 2015. Comparison of genomic selection models to predict flowering time and spike grain number in two hexaploid wheat doubled haploid populations. G3 (Bethesda). 5:1991-1998.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B-Methodol. 58:267-288.

Toosi, A., R.L. Fernando, and J.C.M. Dekkers. 2010. Genomic selection in admixed and crossbred populations. J. Anim. Sci. 88:32-46.

Tsuruta, S., I. Misztal, I. Aguilar, and T.J. Lawlor. 2011. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. J. Dairy Sci. 94:4198-4204.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414-4423.

VanRaden, P.M., C.P.V. Tassell, G.R. Wiggans, T.S. Sonstegard, R.D. Schnabel, J.F. Taylor, and F.S. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16-24.

Varshney, R.K., A. Graner, and M.E. Sorrells. 2005. Genomics-assisted breeding for crop improvement. Trends Plant Sci. 10:621-630.

Voorrips, R.E. 2002. MapChart: software for the graphical presentation of linkage maps and QTLs. J. Hered. 93:77-78.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23:4407-4414.

Wang, C.L., X.D. Ding, J.Y. Wang, J.F. Liu, W.X. Fu, Z. Zhang, Z.J. Yin, and Q. Zhang. 2013. Bayesian methods for estimating GEBVs of threshold traits. Heredity 110:213-219.

Wang, D., I.S. El-Basyoni, P.S. Baenziger, J. Crossa, K.M. Eskridge, and I. Dweikat. 2012. Prediction of genetic values of quantitative traits with epistatic effects in plant breeding populations. Heredity 109:313-319.

Wang, D.G., J.-B. Fan, C.-J. Siao, A. Berno, P. Young, R. Sapolsky, G. Ghandour, N. Perkins, E. Winchester, J. Spencer, L. Kruglyak, L. Stein, L. Hsie, T. Topaloglou, E. Hubbell, E. Robinson, M. Mittmann, M.S. Morris, N. Shen, D. Kilburn, J. Rioux, C. Nusbaum, S. Rozen, T.J. Hudson, R. Lipshutz, M. Chee, and E.S. Lander. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280:1077-1082.

Wang S., C.J. Basten, and Z.-B. Zeng. 2012. Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC.

Wang, S., D. Wong, K. Forrest, A. Allen, S. Chao, B.E. Huang, M. Maccaferri, S. Salvi, S.G. Milner, L. Cattivelli., A.M. Mastrangelo, A. Whan, S. Stephen, G. Barker, R. Wieseke, J. Plieske, International Wheat Genome Sequencing Consortium, M. Lillemo, D. Mather, R. Appels, R. Dolferus, G. Brown-Guedira, A. Korol, A.R. Akhunova, C. Feuillet, J. Salse, M. Morgante, C. Pozniak, M.-C. Luo, J. Dvorak, M. Morell, J. Dubcovsky, M. Ganal, R. Tuberosa, C. Lawley, I. Mikoulitch, C. Cavanagh, K.J. Edwards, M. Hayden, and E. Akhunov. 2014a. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnol J. 12:787-796.

Wang, Y., M.F. Mette, T. Miedaner, M. Gottwald, P. Wilde, J.C. Reif, and Y. Zhao. 2014b. The accuracy of prediction of genomic selection in elite hybrid rye populations surpasses the accuracy of marker-assisted selection and is equally augmented by multiple field evaluation locations and test years. BMC Genomics 15:556.

Weber, J.L., and P.E. May. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. 44:388-396.

Wellmann, R., and J. Bennewitz. 2012. Bayesian models with dominance effects for genomic evaluation of quantitative traits. Genet. Res., Camb. 94:21-37.

Whittaker, J.C., R. Thompson, and M.C. Denham. 2000. Marker-assisted selection using ridge regression. Genet. Res., Camb. 75:249-252.

Wiebe, K., N.S. Harris, J.D. Faris, J.M. Clarke, R.E. Knox, G.J. Taylor, and C.J. Pozniak. 2010. Targeted mapping of *Cdu1*, a major locus regulating grain cadmium concentration in durum wheat (*Triticum turgidum* L. var *durum*). Theor. Appl. Genet. 121:1047-1058.

Wientjes, Y.C.J., R.F. Veerkamp, and M.P.L. Calus. 2013. The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. Genetics 193:621-631.

Williams, J.G.K., A.R. Kubelik, K.J. Livak, J.A. Rafalski, and S.V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18:6531-6535.

Windhausen, V.S., G.N. Atlin, J.M. Hickey, J. Crossa, J.-L. Jannink, M.E. Sorrells, B. Raman, J.E. Cairns, A. Tarekegne, K. Semagn, Y. Beyene, P. Grudloyma, F. Technow, C. Riedelsheimer, and A.E. Melchinger. 2012. Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. G3 (Bethesda). 2:1427-1436.

Winfield, M.O., A.M. Allen, A.J. Burridge, G.L.A. Barker, H.R. Benbow, P.A. Wilkinson, J. Coghill, C. Waterfall, A. Davassi, G. Scopes, A. Pirani, T. Webster, F. Brew, C. Bloor, J. King, C. West, S. Griffiths, I. King, A.R. Bentley, and K.J. Edwards. 2016. High-density SNP genotyping array for hexaploidy wheat and its secondary and tertiary gene pool. Plant Biotechnol. J. 14:1195-1206.

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

Wolfinger, R.D., W.T. Federer, and O. Cordero-Brana. 1997. Recovering information in augmented designs, using SAS PROC GLM and PROC MIXED. Agron. J. 89:856-859.

Wong, C.K., and R. Bernardo. 2008. Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. Theor. Appl. Genet. 116:815-824.

Worland, A.J., E.J. Sayers and V. Korzun. 2001. Allelic variation at the dwarfing gene *Rht8* locus and its significance in international breeding programmes. Euphytica 119:155-159.

Worland, A.J., V. Korzun, M.S. Röder, M.W. Ganal, and C.N. Law. 1998. Genetic analysis of the dwarfing gene *Rht8* in wheat. Part II. The distribution and adaptive significance of allelic variants at the *Rht8* locus of wheat as revealed by microsatellite screening. Theor. Appl. Genet. 96:1110-1120.

Worland, A.J. 1996. The influence of flowering time genes on environmental adaptability in European wheats. Euphytica 89:49-57.

Wray, N.R., J. Yang, B.J. Hayes, A.L. Price, M.E. Goddard, and P.M. Visscher. 2013. Pitfalls of predicting complex traits from SNPs. Nat. Rev. Genet. 14:507-515.

Würschum, T., S. Abel, and Y. Zhao. 2014. Potential of genomic selection in rapeseed (*Brassica napus* L.) breeding. Plant Breeding 133:45-51.

Würschum, T., J.C. Reif, T. Kraft, G. Janssen, and Y. Zhao. 2013. Genomic selection in sugar beet breeding populations. BMC Genet. 14:85.

Xu, Y., and J.H. Crouch. 2008. Marker-assisted selection in plant breeding: from publications to practice. Crop Sci. 48:391-407.

Xu, S., D. Zhu, and Q. Zhang. 2014. Predicting hybrid performance in rice using genomic best linear unbiased prediction. Proc. Natl. Acad. Sci. U.S.A. 111:12456-12461.

Yamamoto, E., H. Matsunaga, A. Onogi, H. Kajiya-Kanegae, M. Minamikawa, A. Suzuki, K. Shirasawa, H. Hirakawa, T. Nunome, H. Yamaguchi, K. Miyatake, A. Ohyama, H. Iwata, and H. Fukuoka. 2016. A simulation-based breeding design that uses whole-genome prediction in tomato. Sci. Rep. 6:19454.

Yang, W., and R.J. Tempelman. 2012. A Bayesian antedependence model for whole genome prediction. Genetics 190:1491-1501.

Yu, J., G. Pressoir, W.H. Briggs, I.V. Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, and E.S. Buckler. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38:203–208

Zhang, X., A. Sallam, L. Gao, T. Kantarski, J. Poland, L.R. DeHaan, D.L. Wyse, and J.A. Anderson. 2016. Establishment and optimization of genomic selection to accelerate the domestication and improvement of intermediate wheatgrass. The Plant Genome 9. doi:10.3835/plantgenome2015.07.0059.

Zhang, Z., U. Ober, M. Erbe, H. Zhang, N. Gao, J. He, J. Li, and H. Simianer. 2014. Improving the accuracy of whole genome prediction for complex traits using the results of genome wide association studies. PloS One 9(3): e93017.

Zhao, Y., J. Zeng, R. Fernando, and J.C. Reif. 2013. Genomic prediction of hybrid wheat performance. Crop Sci. 53:802-810.

Zhao, Y., M. Gowda, W. Liu, T. Würschum, H.P. Maurer, F.H. Longin, N. Ranc, and J.C. Reif. 2012. Accuracy of genomic selection in European maize elite breeding populations. Theor. Appl. Genet. 124:769-776.

Zhao, Y., M.F. Mette, M. Gowda, C.F.H. Longin, and J.C. Reif. 2014. Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. Heredity 112:638-645.

Zhong, S., J.C.M. Dekkers, R.L. Fernando, and J.-L. Jannink. 2009. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. Genetics 182:355-364.

Zhu, C., M. Gore, E.S. Buckler, and J. Yu. 2008. Status and prospects of association mapping in plants. The Plant Genome 1:5-20.

Zohary D. and M. Hopf. 2000. Cereals. In: D. Zohary and M. Hopf, editors, Domestication of plants in the old world. Oxford University Press Inc. New York. p. 16-91.

Zou, J., K. Semagn, M. Iqbal, H. Chen, M. Asif, A. N'Diaye, A. Navabi, E. Perez-Lara, C. Pozniak, R.-C. Yang, H. Randhawa, and D. Spaner. 2017. QTLs associated with agronomic traits in the Attila × CDC go spring wheat population evaluated under conventional management. PloS One 12(2): e0171528.

## Appendices

**Appendix A.** Origin and pedigree information of the lines used as training population and for genome-wide association study.

| Name | Trial | Origin | Pedigree |
|---|---|---|---|
| 5500HR | Varcomp | Canada | N91-2381/AC Minto |
| 5600HR | Varcomp | Canada | N91-2071/AC Minto |
| 5601HR | Varcomp | Canada | N893-2410/AC Majestic |
| 5602HR | Varcomp | Canada | AC Barrie/Norpro |
| 5603HR | Varcomp | Canada | McKenzie// 97NPI15-55/Lars |
| 5604HR CL | Both | Canada | AC Barrie//Butte86*4/FS4/3/CDC Teal/4/AC Domain*2/AC Cora |
| 5700PR | Varcomp | Canada | N91-3051/AC Foremost |
| 5701PR | Varcomp | Canada | N89-3003/N87-446//Oslo |
| 5702PR | Varcomp | Canada | HY 437///Russ//Sumai #3/Dalen |
| 8021-V2 | Co-op | Canada | Kenya 321/Peck |
| AC Abbey | Varcomp | Canada | BW 608/ 93464//BW 591 |
| AC Barrie | Both | Canada | Neepawa/Columbus//Pacific |
| AC Cadillac | Both | Canada | Pacific*3/BW 553 |
| AC Cora | Co-op | Canada | Katepwa/RL 4509 |
| AC Crystal | Both | Canada | HY 377/L8474-D1 |
| AC Domain | Both | Canada | BW 83/ND585 |
| AC Eatonia | Varcomp | Canada | Leader/Lancer |
| AC Elsa | Varcomp | Canada | Pacific/Laura |
| AC Foremost | Varcomp | Canada | HY320*5/BW553//HY320*6/7424-BW5B4 |
| AC Intrepid | Both | Canada | Laura/RL 4596//CDC Teal |
| AC Karma | Co-op | Canada | HY320*5/BW553//HY 358///HY 358/ 7915-QX76B2 |
| AC Majestic | Co-op | Canada | Columbus*2//Saric-70/Neepawa///Columbus*5//Saric-70/Neepawa |
| AC Michael | Co-op | Canada | Park/Neepawa |
| AC Minto | Co-op | Canada | CLMS/BW63//Katepwa/BW552 |
| AC Splendor | Both | Canada | Laura/RL 4596//Roblin/BW 107 |

**Appendix A.** Continued

| Name | Trial | Origin | Pedigree |
|---|---|---|---|
| AC Taber | Varcomp | Canada | Tobari F 66/Romany//HY 320*2/BW 553 |
| AC Vista | Varcomp | Canada | HY344/7915-QX76B2//HY358*3/BT10 |
| Alikat | Co-op | Canada | Katepwa*3/Maringa |
| Alsen | Co-op | USA | ND-674/ND-2710//ND 688 |
| Alvena | Both | Canada | BW 711/AC Intrepid |
| Benito | Co-op | Canada | CT 257/RL 4008///RL 4255*4//MIT/CI7090 |
| Bluesky | Co-op | Canada | Potam S 70/Glenlea |
| Burnside | Varcomp | Canada | Glenlea*2// 90B07-W3B/2*RL4452 |
| BW270 | Co-op | Canada | BW 165/RL 4660 |
| BW275 | Co-op | Canada | BW 83/ND585///BW 34*6//Thatcher/Poso 48 |
| BW314A | Co-op | Canada | RL 4763*2/Howell |
| BW317 | Co-op | Canada | AC Cadillac/ 8405-JC3C//AC Elsa |
| BW334 | Co-op | Canada | 9007-FB1C/AC Elsa//AC Barrie |
| BW337 | Co-op | Canada | CDC Teal*2/ND-2710 |
| BW338 | Co-op | Canada | BW 83/ND585 |
| BW343 | Co-op | Canada | 94B42-V2A/Superb |
| BW360 | Co-op | Canada | McKenzie*3//BW174*2/Clark |
| BW369 | Co-op | Canada | BW 193/Grandin//BW 236/AC Domain |
| BW370 | Co-op | Canada | N96-2449/AC Splendor |
| BW371 | Co-op | Canada | BW 240/McKenzie |
| BW377 | Co-op | Canada | AC Barrie*2//CDC Teal*2/Seneca-90 |
| BW385 | Co-op | Canada | 00H01*X3/ 98B21-S4A04 |
| BW387 | Co-op | Canada | BW 193/Grandin//BW 236/AC Domain |
| BW389 | Co-op | Canada | N98-2670/McKenzie |
| BW391 | Co-op | Canada | N95-2249/AC Domain//BW 763 |
| BW395 | Co-op | Canada | 99B61-AM15A3/BW392 |

**Appendix A.** Continued

| Name | Trial | Origin | Pedigree |
|---|---|---|---|
| BW396 | Co-op | Canada | 99B61-AY30B5/BW392 |
| BW397 | Co-op | Canada | Alsen/P00.06-77 |
| BW403 | Co-op | Canada | unkown |
| BW410 | Co-op | Canada | McKenzie//BW257/94B92-Y3B |
| BW416 | Co-op | Canada | 97S2029-66/ 97S2177-41 |
| BW417 | Co-op | Canada | 98S2014-10/ 97S2177-41 |
| BW421 | Co-op | Canada | CDC Bounty/FHB9 |
| BW423 | Both | Canada | CDC Bounty/FHB9 |
| BW425 | Co-op | Canada | AC Domain/BW 257 |
| BW427 | Co-op | Canada | Superb/ 98B19*J191 |
| BW429 | Co-op | Canada | McKenzie/Alsen |
| BW430 | Co-op | Canada | Alsen/BW 313 |
| BW431 | Co-op | Canada | 00H01*F57/BW 344 |
| BW433 | Co-op | Canada | BW275W/N99-2587 |
| BW449 | Co-op | Canada | 00H01*F57/BW 344 |
| BW450 | Co-op | Canada | 00H01*D26/ 00H04*J3 |
| BW454 | Co-op | Canada | HC736/98B69-R28//2*Prodigy///HC374*3/ 98B69-L47 |
| BW455 | Co-op | Canada | 98B34-T4B/ 98B26-N1C01B |
| BW464 | Co-op | Canada | Kane/ 98B25-AS3C02 |
| BW469 | Co-op | Canada | BW 361/ 5602HR |
| BW479 | Co-op | Canada | 02S2004-2-12/Glenn |
| BW755 | Co-op | Canada | Grandin*3/Fidel |
| BW768 | Co-op | Canada | P8913-V2A5/2* 8405-JC3C |
| BW774 | Co-op | Canada | P8913-V2A5/2* 8405-JC3C |
| BW796 | Co-op | Canada | AC Cadillac/ 8405-JC3C//AC Elsa |
| BW809 | Co-op | Canada | AC Barrie/AC Elsa |

**Appendix A.** Continued

| Name | Trial | Origin | Pedigree |
|---|---|---|---|
| BW811 | Co-op | Canada | AC Elsa//W88499/BW148 |
| BW813 | Co-op | Canada | AC Elsa*2/BW 755 |
| BW814 | Co-op | Canada | AC Elsa*2/BW 755 |
| BW826 | Co-op | Canada | AC Barrie//Butte86*4/FS4///McKenzie |
| BW830 | Co-op | Canada | BW 674/AC Cadillac//AC Barrie |
| BW833 | Co-op | Canada | SD-3055/AC Barrie |
| BW834 | Co-op | Canada | BW 725/AC Intrepid//AC Barrie |
| BW839 | Co-op | Canada | 94B35-X3A/AC Barrie//Superb |
| BW843 | Co-op | Canada | AC Majestic/X95.4 |
| BW847 | Co-op | Canada | AC Barrie/W93079 |
| BW849 | Co-op | Canada | McKenzie//BW661/BW755 |
| BW852 | Co-op | Canada | W98085/AC Barrie |
| BW854 | Co-op | Canada | AC Barrie/BW 725 |
| BW859 | Both | Canada | N94-2189/N92-2308///AC Barrie//Butte86*4/FS4 |
| BW870 | Co-op | Canada | Alsen/AC Elsa//AC Barrie |
| BW871 | Co-op | Canada | Alsen/AC Elsa//AC Barrie |
| BW879 | Co-op | Canada | N92-2308/AC Majestic/5/AC Barrie//Butte86*4/FS4///BW-604/BW38/4/Columbus/Amidon |
| BW901 | Co-op | Canada | BW 807/Journey//Lillian |
| BW902 | Co-op | Canada | BW 317/2*Alsen |
| BW908 | Co-op | Canada | CDC Go/ND694 |
| BW911 | Co-op | Canada | BW 322/PT 607 |
| BW912 | Co-op | Canada | 99S2087-7L//ID580/Briggs |
| BW919 | Co-op | Canada | 01II01-20-5G/ 5602HR |
| BW922 | Co-op | Canada | BW 282/CDC Go |
| BW927 | Co-op | Canada | Infinity/BW 349//Alsen |
| BW928 | Co-op | Canada | Infinity/BW 349//Alsen |

**Appendix A.** Continued

| Name | Trial | Origin | Pedigree |
|---|---|---|---|
| BW929 | Co-op | Canada | Infinity/BW 349//Alsen |
| BW930 | Co-op | Canada | Prodigy/Lovitt//Alsen |
| BW935 | Co-op | Canada | Infinity/ 5602HR//Alsen |
| BW942 | Co-op | Canada | BW755/11A//2*ND694 |
| BW948 | Co-op | Canada | FHB148/BW 278//Snowbird |
| BW949 | Co-op | Canada | Helios/Somerset |
| BW956 | Co-op | Canada | Lillian/BW 349//BW 828 |
| BW959 | Co-op | Canada | Helios/ 5602HR |
| BW961 | Co-op | Canada | Alsen/Waskada |
| Carberry | Both | Canada | Alsen/Superb |
| CDC Abound | Both | Canada | Superb*2/BW 755 |
| CDC Alsask | Varcomp | Canada | AC Elsa/AC Cora |
| CDC Bounty | Varcomp | Canada | Katepwa/W82624 |
| CDC Go | Varcomp | Canada | Grandin/SD-3055 |
| CDC Imagine | Both | Canada | CDC Teal*4/FS2 |
| CDC Kernen | Varcomp | Canada | CDC Bounty/FHB4 |
| CDC Makwa | Co-op | Canada | S7432/MIT//Benito |
| CDC Merlin | Varcomp | Canada | RL4386//BW525/BW37 |
| CDC Osler | Varcomp | Canada | AC Cora/PT 534 |
| CDC Plentiful | Both | Canada | BW 282/CDC Go |
| CDC Rama | Varcomp | Canada | McNeal/Glenlea |
| CDC Stanley | Varcomp | Canada | W95132/AC Barrie |
| CDC Teal | Varcomp | Canada | BW-604/BW38 |
| CDC Thrive | Varcomp | Canada | CDC Bounty/W98501 |
| CDC Utmost | Both | Canada | AC Elsa//CDC Teal/Seneca |
| CDC Walrus | Varcomp | Canada | Glenlea*2/McNeal |

**Appendix A.** Continued

| Name | Trial | Origin | Pedigree |
|---|---|---|---|
| CDN Bison | Varcomp | Canada | ES 012/ES 009 |
| Columbus | Varcomp | Canada | CT 257/RL 4008//Neepawa*5/BW 34 |
| Conway | Co-op | Canada | CHR/7C//NEP/OPAL |
| Cutler | Co-op | Canada | CNO/CC//INIA F 66 |
| Glencross | Varcomp | Canada | 96E06*A8/ 94E13-D1B |
| Glenlea | Varcomp | Canada | Pembina*2/BAGE///SN64/TZPP//Nainari 60 |
| Glenn | Varcomp | USA | ND 2831/Steele, PI516196 |
| Goodeve VB | Varcomp | Canada | Goodeve/AC Intrepid |
| GP069 | Co-op | Canada | HY 459/ALSEN//Snowhite475 |
| Grandin | Co-op | Canada | LEN//Butte*2/ND507/8/North Dakota 499///Justin/RL4205//WIS261/7/Butte/6/Butte/5/Waldron/4/PBA//TH/TF/3/PBA//TH/AUS-6774 |
| Harvest | Varcomp | Canada | AC Domain*2/BN-142 |
| Helios | Varcomp | Canada | BW 674/AC Cadillac//AC Barrie |
| HW021 | Co-op | Canada | – |
| HW024 | Co-op | Canada | Snowbird///BW315//Snowbird*2/BW314 |
| HY682 | Varcomp | Canada | HY 639/ 99 EPWAMDG 61 |
| Infinity | Varcomp | Canada | ND-671/ 8405-JC3C//AC Elsa |
| Invader | Co-op | Canada | Sinton/STOA |
| Journey | Varcomp | Canada | CDC Teal//Grandin/PT819 |
| Kanata | Co-op | Canada | BW 83/ND585///BW 34*6//Thatcher/POSO 48 |
| Kane | Varcomp | Canada | AC Domain/McKenzie |
| Katepwa | Varcomp | Canada | Neepawa*6/RL2938///Neepawa*6//PI 59284/2*FCR |
| Kenyon | Co-op | Canada | Neepawa*5/Buck Manantial |
| Lancer | Co-op | Canada | Fortuna/4/K58/N//TH///FN/TH |
| Laser | Co-op | Canada | NBB134/ 70M009 |
| Laura | Varcomp | Canada | BW15/BW 517 |

**Appendix A.** Continued

| Name | Trial | Origin | Pedigree |
|---|---|---|---|
| Leader | Co-op | Canada | Fortuna/4/K58/N//TH///FN/TH |
| Lillian | Varcomp | Canada | BW 621*3/ 90B07-AU2B |
| Lovitt | Varcomp | Canada | 8405-JC3C*2/AC Cora |
| McKenzie | Both | Canada | Columbus/Amidon |
| Minnedosa | Varcomp | Canada | AC Vista*3/LR18 |
| Muchmore | Both | Canada | Alsen/Superb |
| Neepawa | Varcomp | Canada | CT 257/RL 4008 |
| NRG010 | Varcomp | Canada | ND-2710/HY 459//AC Vista |
| Pacific | Co-op | Canada | RL4302/RL4356//RL4359/RL4353 |
| Park | Varcomp | Canada | CT609/Thatcher |
| Pasqua | Co-op | Canada | BW63//BW63/CLMS |
| Peace | Co-op | Canada | BW 165/RL 4660 |
| Prodigy | Varcomp | Canada | SWP2242/STOA |
| PT206 | Co-op | Canada | AC Cadillac/ 8405-JC3C//AC Elsa |
| PT212 | Co-op | Canada | BW 711/AC Intrepid |
| PT224 | Co-op | Canada | AC Cadillac/Superb//AC Barrie |
| PT225 | Co-op | Canada | AC Cadillac/Superb//AC Barrie |
| PT228 | Co-op | Canada | 9229G-003B/AC Barrie//AC Elsa |
| PT242 | Co-op | Canada | Goodeve/ 96B23-AD2D//CDC Osler |
| PT246 | Co-op | Canada | Stettler/Glenn |
| PT421 | Co-op | Canada | AC Domain/CDC Teal |
| PT425 | Co-op | Canada | AC Splendor/AC Elsa |
| PT430 | Co-op | Canada | AC Intrepid*3//BW174*2/Clark |
| PT434 | Co-op | Canada | AC Domain*6/LR22A |
| PT435 | Co-op | Canada | AC Domain*6/LR22A |
| PT441 | Co-op | Canada | McKenzie*3//BW174*2/Clark |

**Appendix A.** Continued

| Name | Trial | Origin | Pedigree |
|---|---|---|---|
| PT443 | Co-op | Canada | BW 226/Harvest |
| PT446 | Co-op | Canada | BW 226/Harvest |
| PT447 | Co-op | Canada | McKenzie//BW257/94B92-Y3B |
| PT450 | Co-op | Canada | AC Splendor/Harvest |
| PT451 | Co-op | Canada | BW 226/BW 314 |
| PT452 | Co-op | Canada | AC Splendor/BW 314 |
| PT456 | Co-op | Canada | BW 226/BW 314 |
| PT458 | Co-op | Canada | Harvest/McKenzie//AC Intrepid |
| PT459 | Co-op | Canada | Harvest/McKenzie//AC Intrepid |
| PT460 | Co-op | Canada | Harvest/McKenzie//AC Intrepid |
| PT464 | Co-op | Canada | PT 425/Helios |
| PT465 | Co-op | Canada | AC Intrepid/Somerset |
| PT468 | Co-op | Canada | Helios/Somerset |
| PT551 | Co-op | Canada | CDC Teal//EE8/KYN18 |
| PT553 | Co-op | Canada | PT532//Columbus*2/CDC Makwa |
| PT554 | Co-op | Canada | PT532///Columbus//Roblin/5297 |
| PT558 | Co-op | Canada | AC Domain//ND640/PT532 |
| PT559 | Both | Canada | SD-3055/AC Domain |
| PT560 | Co-op | Canada | AC Domain//ND655/PT532 |
| PT565 | Co-op | Canada | AC Barrie/CDC Teal |
| PT570 | Co-op | Canada | McKenzie//BW661/BW755 |
| PT574 | Co-op | Canada | AC Intrepid//CDC Teal/97IMIEG2-18 |
| PT577 | Co-op | Canada | AC Intrepid/CDC Bounty |
| PT579 | Co-op | Canada | CDC Alsask/BW 280 |
| PT583 | Co-op | Canada | BW 282/P00.02-56 |
| PT610 | Co-op | Canada | BW 193/Grandin//BW 236/AC Domain |

**Appendix A.** Continued

| Name | Trial | Origin | Pedigree |
|---|---|---|---|
| PT612 | Co-op | Canada | BW 270//N93-2210/BW193 |
| PT613 | Co-op | Canada | Bergen/JAG SIB//BW193///AC Intrepid |
| PT616 | Co-op | Canada | N99-2234/ 97S2177-41 |
| PT619 | Co-op | Canada | McKenzie/BC97ROM-52//BRIGGS |
| PT624 | Co-op | Canada | Alsen/BW 350 |
| PT637 | Co-op | Canada | BW 337/AC Elsa |
| PT756 | Co-op | Canada | AC Domain/Saunders |
| Red Fife | Varcomp | Canada | LV-POL |
| Rescue | Co-op | USA | APEX/S615 |
| RL4137 | Co-op | Canada | CID 796 SID 1/Mentana//RL-2265/2*Redman///Thatcher*5/III-52-7 |
| RL4452 | Co-op | Canada | Glenlea*6/Kitt |
| Roblin | Both | Canada | BW15/BW38//BW40/RL4353 |
| Selkirk | Co-op | Canada | RL-2265/2*Redman//Regent/Canus |
| Snowbird | Both | Canada | BW 83/ND585///BW 34*6//Thatcher/Poso 48 |
| Snowstar | Both | Canada | RL 4869/McKenzie |
| Somerset | Varcomp | Canada | 90B01-AD4D/Pasqua |
| Stettler | Varcomp | Canada | Prodigy/Superb |
| Sunmist | Co-op | Australia | – |
| Superb | Both | Canada | Grandin*2/AC Domain |
| SY985 | Varcomp | Canada | N99-3098WL/N98-3080W |
| Thatcher | Varcomp | USA | Marquis/Iumillo//Marquis/Kanred |
| Unity | Both | Canada | McKenzie*3//BW174*2/Clark |
| Vesper | Varcomp | Canada | Augusta/Hard White Alpha//3*AC Barrie/6/BW 150*2//TP/TM/3/2*Superb/4/Grandin*2/Caldwell/5/Superb |
| Waskada | Both | Canada | BW 278/2*Superb |
| Wildcat | Co-op | Canada | NB113/Glenlea |

**Appendix B.** Steps used for SNP genotype calling of the training population (varieties and advanced breeding lines) using the polyploid version of Genome Studio.

**Step 1**

1. Set DBSCAN and OPTICS clustering algorithm parameters *Cluster Distance* to 0.1 and *Minimum Number of Points in Cluster* to 10.
2. Select all SNPs in SNP Table, right click over the selected SNPs and choose *Cluster Selected SNPs* in the menu that appears, then select *OPTICS* in the sub-menu.
3. Sort SNP Table by *C1 Freq* in ascending order.
4. Select SNPs with *C1 Freq* < 1 and set *Aux* to 5.

**Step 2**

1. Filter SNPs with *Aux* = 0
2. Set OPTICS clustering algorithm parameter *Cluster Distance* to 0.09
3. Select all SNPs in SNP Table, right click over the selected SNPs and choose *Cluster Selected SNPs*
4. Sort SNP Table by *C1 Freq* in ascending order.
5. Select SNPs with *C1 Freq* < 1 and set *Aux* to 10.

**Step 3**

1. Filter SNPs with *Aux* = 0
2. Set OPTICS clustering algorithm parameter *Cluster Distance* to 0.08
3. Select all SNPs in SNP Table, right click over the selected SNPs and choose *Cluster Selected SNPs*
4. Sort SNP Table by *C1 Freq* in ascending order.
5. Select SNPs with *C1 Freq* < 1 and set *Aux* to 15.

**Step 4**

1. Filter SNPs with *Aux* = 0
2. Set OPTICS clustering algorithm parameter *Cluster Distance* to 0.07
3. Select all SNPs in SNP Table, right click over the selected SNPs and choose *Cluster Selected SNPs*
4. Sort SNP Table by *C1 Freq* in ascending order.
5. Select SNPs with *C1 Freq* < 1 and set *Aux* to 20.

**Step 5**

1. Filter SNPs with *Aux* = 0
2. Set OPTICS clustering algorithm parameter *Cluster Distance* to 0.06
3. Select all SNPs in SNP Table, right click over the selected SNPs and choose *Cluster Selected SNPs*
4. Sort SNP Table by *C1 Freq* in ascending order.
5. Select SNPs with *C1 Freq* < 1 and set *Aux* to 25.

**Step 6**

1. Filter SNPs with *Aux* = 0
2. Set OPTICS clustering algorithm parameter *Cluster Distance* to 0.05
3. Select all SNPs in SNP Table, right click over the selected SNPs and choose *Cluster Selected SNPs*
4. Sort SNP Table by *C1 Freq* in ascending order.
5. Select SNPs with *C1 Freq* < 1 and set *Aux* to 30.

**Step 7**

1. Filter SNPs with *Aux* = 0
2. Set OPTICS clustering algorithm parameter *Cluster Distance* to 0.04
3. Select all SNPs in SNP Table, right click over the selected SNPs and choose *Cluster Selected SNPs*
4. Sort SNP Table by *C1 Freq* in ascending order.
5. Select SNPs with *C1 Freq* < 1 and set *Aux* to 35.

**Step 8**

1. Select SNPs with *Aux* = 0
2. Set OPTICS clustering algorithm parameter *Cluster Distance* to 0.03
3. Select all SNPs in SNP Table, right click over the selected SNPs and choose *Cluster Selected SNPs*
4. Sort SNP Table by *C1 Freq* in ascending order.
5. Select SNPs with *C1 Freq* < 1 and set *Aux* to 40.

**Step 9**

1. Select SNPs with *Aux* = 0
2. Set OPTICS clustering algorithm parameter *Cluster Distance* to 0.02

3. Select all SNPs in SNP Table, right click over the selected SNPs and choose *Cluster Selected SNPs*
4. Sort SNP Table by *C1 Freq* in ascending order.
5. Select SNPs with *C1 Freq* < 1 and set *Aux* to 45.

**Step 10**

1. Select SNPs with *Aux* = 0
2. Set OPTICS clustering algorithm parameter *Cluster Distance* to 0.01
3. Select all SNPs in SNP Table, right click over the selected SNPs and choose *Cluster Selected SNPs*
4. Sort SNP Table by *C1 Freq* in ascending order.
5. Select SNPs with *C1 Freq* < 1 and set *Aux* to 50.
6. Select SNPs with *Aux* = 0 and set comment as Monomorphic.

Finally, filter SNP Table by *Aux* > 0 and visually check clustering. Manually curate incorrectly clustered SNPs.

**Appendix C.** List of significant markers detected from fold specific GWAS in the training population and their $R^2$ values. These markers were fitted as fixed effects in the GS + de novo GWAS model for genomic predictions in the training population based on fivefold cross-validation

| Trait† | CV Fold‡ | SNP ID | Chromosome | Position (cM) | Marker $R^2$ |
|---|---|---|---|---|---|
| HD | 1 | wsnp_CAP12_c812_428290 | 2D | 19.03 | 0.11 |
| HD | 2 | wsnp_CAP12_c812_428290 | 2D | 19.03 | 0.10 |
| HD | 3 | wsnp_CAP12_c812_428290 | 2D | 19.03 | 0.08 |
| HD | 3 | BS00065128_51 | 5B | 110.56 | 0.07 |
| HD | 4 | wsnp_Ex_c13485_21225504 | 5B | 97.28 | 0.08 |
| HD | 4 | wsnp_CAP12_c812_428290 | 2D | 19.03 | 0.08 |
| HD | 5 | wsnp_CAP12_c812_428290 | 2D | 19.03 | 0.09 |
| HT | 1 | IAAV971 | 4B | 57.49 | 0.15 |
| HT | 1 | wsnp_Ra_c22026_31453420 | 4B | 72.53 | 0.09 |
| HT | 1 | Tdurum_contig33737_157 | 4B | 55.96 | 0.13 |
| HT | 2 | IAAV971 | 4B | 57.49 | 0.15 |
| HT | 2 | Excalibur_c56787_95 | 4B | 58.10 | 0.09 |
| HT | 2 | Tdurum_contig33737_157 | 4B | 55.96 | 0.13 |
| HT | 3 | IAAV971 | 4B | 57.49 | 0.14 |
| HT | 3 | wsnp_Ra_c22026_31453420 | 4B | 72.53 | 0.08 |
| HT | 3 | Tdurum_contig33737_157 | 4B | 55.96 | 0.11 |
| HT | 4 | IAAV971 | 4B | 57.49 | 0.11 |
| HT | 4 | wsnp_Ex_c47370_52604482 | 2A | 106.86 | 0.07 |
| HT | 4 | RAC875_c38018_278 | 2A | 110.13 | 0.09 |
| HT | 5 | IAAV971 | 4B | 57.49 | 0.15 |
| HT | 5 | Tdurum_contig64772_417 | 4B | 50.85 | 0.10 |
| HT | 5 | Tdurum_contig33737_157 | 4B | 55.96 | 0.14 |

**Appendix C.** Continued

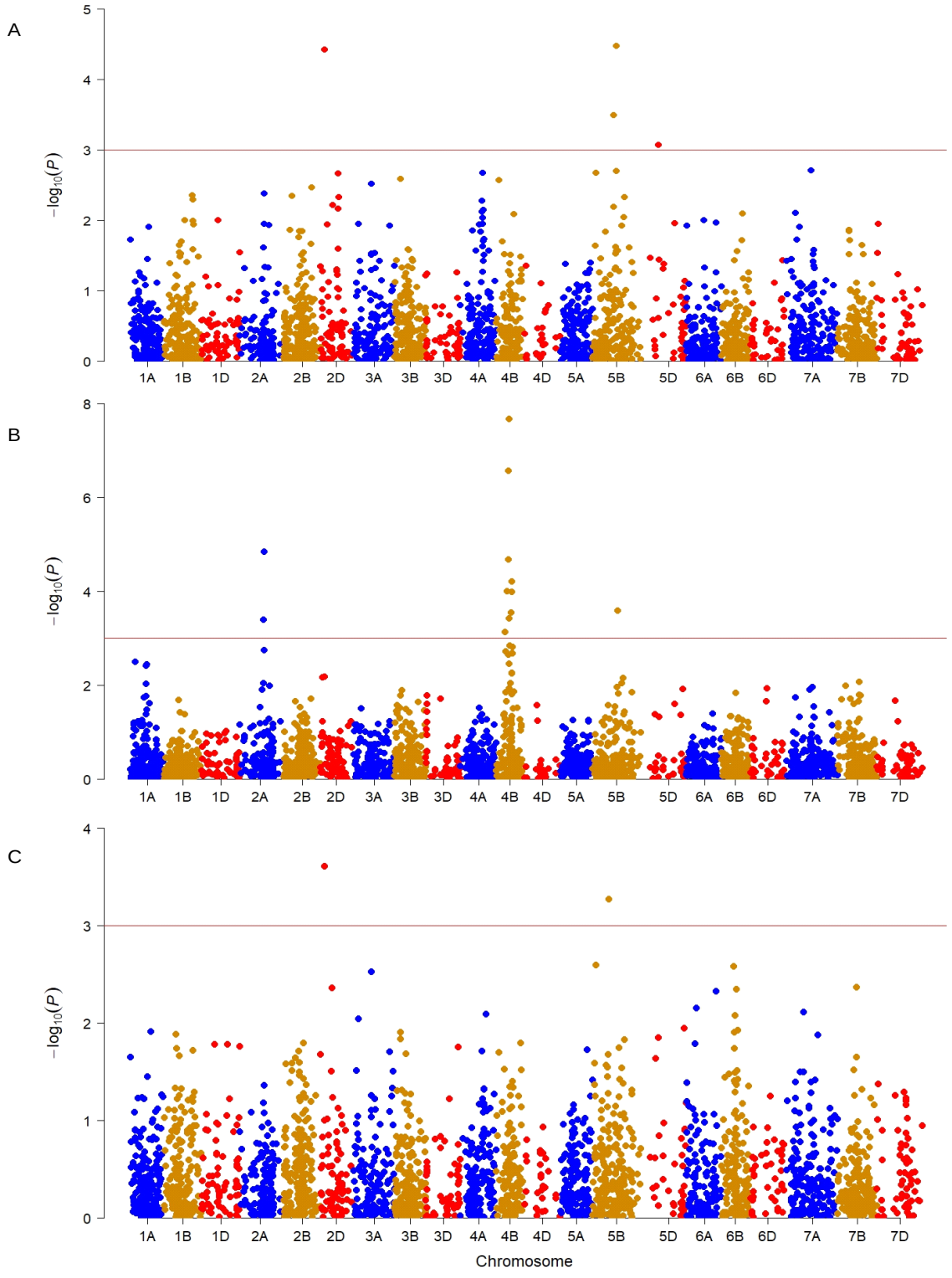
| Trait† | CV Fold‡ | SNP ID | Chromosome | Position (cM) | Marker $R^2$ |
|---|---|---|---|---|---|
| MAT | 1 | wsnp_CAP12_c812_428290 | 2D | 19.03 | 0.08 |
| MAT | 2 | wsnp_CAP12_c812_428290 | 2D | 19.03 | 0.06 |
| MAT | 3 | wsnp_CAP12_c812_428290 | 2D | 19.03 | 0.06 |
| MAT | 4 | RAC875_c62325_320 | 6A | 40.52 | 0.08 |
| MAT | 5 | wsnp_CAP12_c812_428290 | 2D | 19.03 | 0.08 |
| YLD | 1 | wsnp_Ku_c6065_10682531 | 7A | 125.26 | 0.06 |
| YLD | 2 | BS00067907_51 | 2A | 132.74 | 0.07 |
| YLD | 2 | Kukri_c6944_1636 | 2A | 140.94 | 0.06 |
| YLD | 2 | wsnp_JD_c640_960796 | 2B | 20.86 | 0.06 |
| YLD | 3 | IACX2250 | 6A | 81.17 | 0.07 |
| YLD | 4 | Tdurum_contig43552_666 | 5D | 193.91 | 0.08 |
| YLD | 5 | wsnp_Ku_c6065_10682531 | 7A | 125.26 | 0.06 |
| TWT | 1 | IACX8453 | 7A | 118.40 | 0.06 |
| TWT | 2 | Tdurum_contig8348_831 | 5A | 141.75 | 0.06 |
| TWT | 3 | IACX8453 | 7A | 118.40 | 0.08 |
| TWT | 4 | Ex_c25733_348 | 1B | 115.19 | 0.06 |
| TWT | 5 | BS00065296_51 | 5A | 15.86 | 0.06 |
| TKW | 1 | BobWhite_rep_c50057_164 | 1B | 145.80 | 0.07 |
| TKW | 2 | wsnp_Ex_c3372_6195001 | 1D | 75.04 | 0.05 |
| TKW | 3 | BS00068520_51 | 3A | 88.02 | 0.05 |
| TKW | 4 | RFL_Contig3869_808 | 7D | 148.62 | 0.06 |
| TKW | 5 | Excalibur_c77321_69 | 3A | 175.26 | 0.06 |

**Appendix C.** Continued

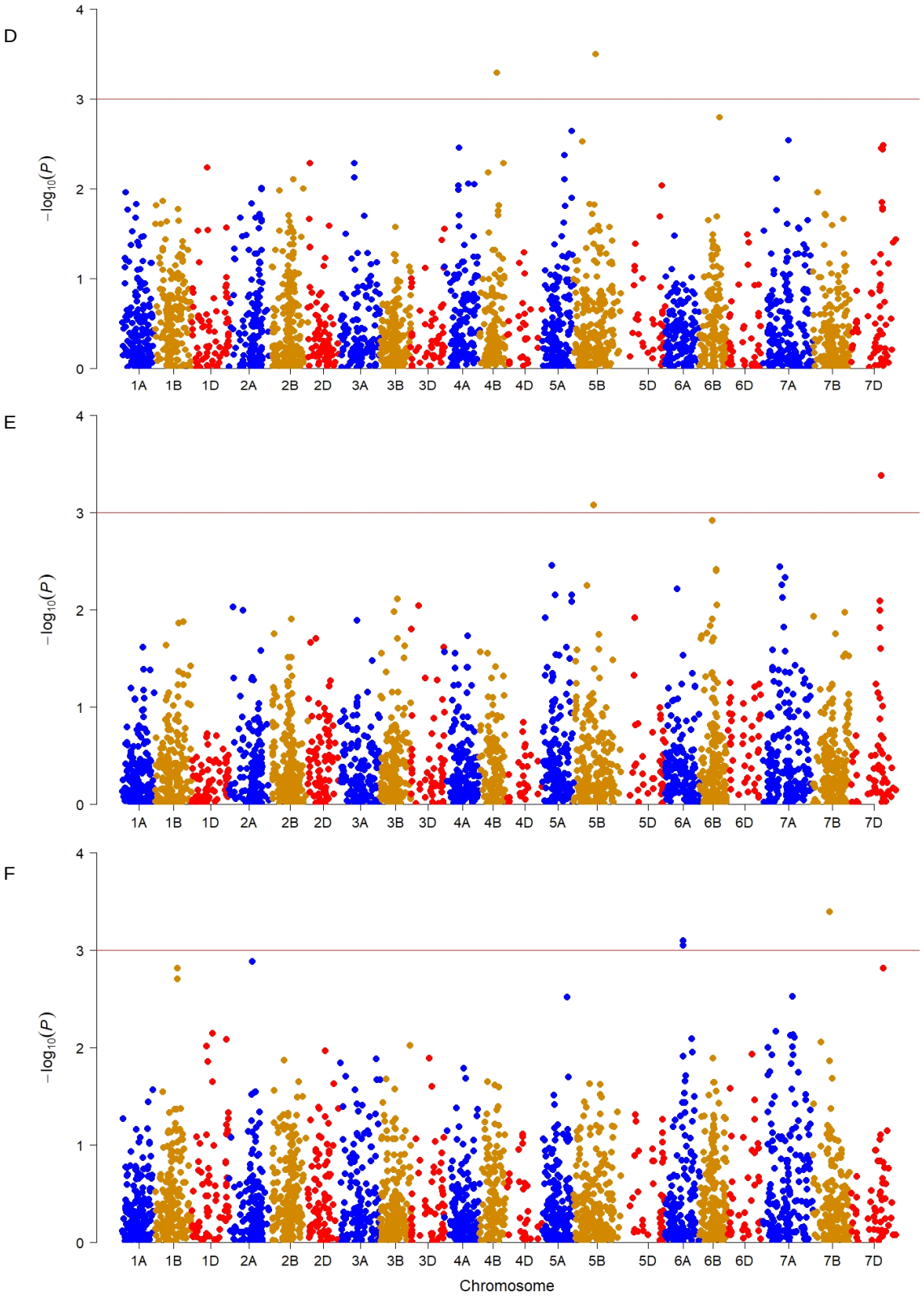
| Trait† | CV Fold‡ | SNP ID | Chromosome | Position (cM) | Marker $R^2$ |
|---|---|---|---|---|---|
| PRO | 1 | Excalibur_rep_c107573_54 | 7B | 66.62 | 0.07 |
| PRO | 2 | Kukri_c19178_2327 | 5B | 139.40 | 0.09 |
| PRO | 3 | Excalibur_c3165_730 | 5B | 114.95 | 0.07 |
| PRO | 4 | Kukri_c19178_2327 | 5B | 139.40 | 0.08 |
| PRO | 5 | BS00033782_51 | 7A | 36.05 | 0.12 |
| PRO | 5 | tplb0024a13_1332 | 7A | 34.90 | 0.10 |
| PRO | 5 | Tdurum_contig77759_52 | 7A | 60.47 | 0.11 |
| FN | 1 | BS00009458_51 | 2D | 95.95 | 0.06 |
| FN | 2 | IACX2250 | 6A | 81.17 | 0.09 |
| FN | 3 | Tdurum_contig44206_1503 | 7B | 159.67 | 0.08 |
| FN | 4 | BS00078844_51 | 3B | 85.03 | 0.08 |
| FN | 5 | BS00098432_51 | 5D | 200.54 | 0.12 |
| SDS | 1 | D_contig12192_450 | 1B | 122.76 | 0.08 |
| SDS | 1 | BS00035267_51 | 1B | 122.38 | 0.08 |
| SDS | 2 | D_contig12192_450 | 1B | 122.76 | 0.08 |
| SDS | 2 | BS00035267_51 | 1B | 122.38 | 0.07 |
| SDS | 2 | BobWhite_c26569_190 | 1A | 111.55 | 0.07 |
| SDS | 3 | D_contig12192_450 | 1B | 122.76 | 0.12 |
| SDS | 3 | BS00035267_51 | 1B | 122.38 | 0.10 |
| SDS | 3 | BobWhite_c14362_86 | 1B | 125.26 | 0.11 |
| SDS | 4 | D_contig12192_450 | 1B | 122.76 | 0.09 |
| SDS | 4 | Tdurum_contig42405_197 | 1A | 13.73 | 0.08 |
| SDS | 4 | BS00010849_51 | 3B | 27.75 | 0.09 |
| SDS | 5 | Tdurum_contig67350_771 | 3B | 86.97 | 0.10 |

†HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight, PRO: grain protein, FN: falling number, SDS: sedimentation volume.

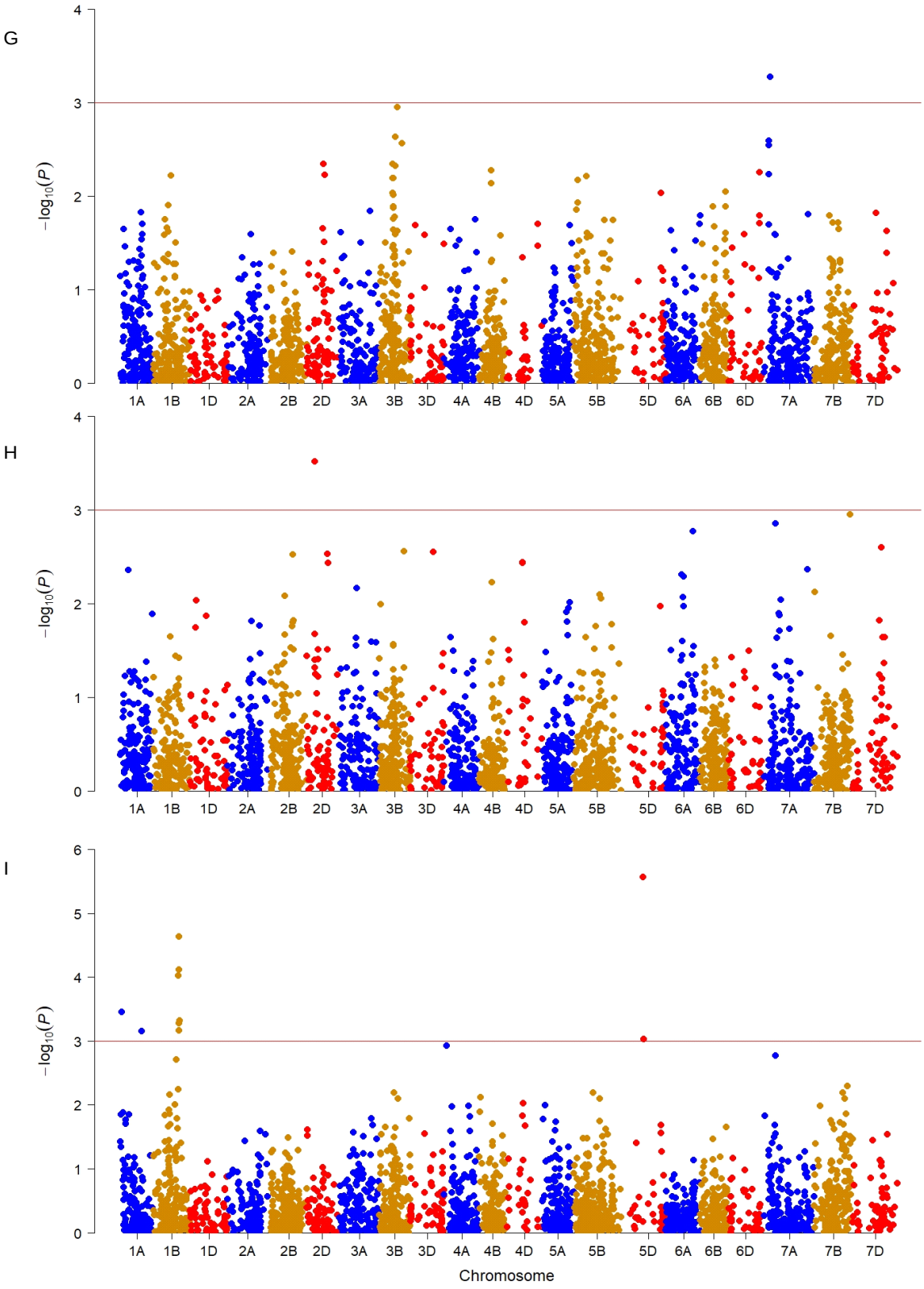
‡CV: cross-validation.




**Appendix D.** Manhattan plots of the genome-wide association study results. (A) heading date, (B) plant height, (C) maturity, (D) grain yield, (E) test weight, (F) thousand-kernel weight, (G) grain protein, (H) falling number, (I) sedimentation volume. The horizontal line indicates the threshold at $-\log_{10}$ (*P*-value) of 3.




**Appendix D**. Continued




**Appendix D**. Continued

**Appendix E**. Steps used for SNP genotype calling of the selection candidates (RILs) using the polyploid version of Genome Studio (Modified from Wang et al. 2014a)

**Step 1**

1. Set DBSCAN and OPTICS clustering algorithm parameters *Cluster Distance* to 0.07 and *Minimum Number of Points in Cluster* to 10.
2. Select all SNPs in SNP Table, right click over the selected SNPs and choose *Cluster Selected SNPs* in the menu that appears, then select *DBSCAN* in the sub-menu.
3. Sort SNP Table by # *Clusters* and *Call Freq* in ascending order.
4. Select SNPs with # *Clusters* equal to 2 and *Call Freq* > 0.9, and set *Aux* to 1.
5. Sort SNP Table by *Aux* in descending order then by *C1 Freq* and *C2 Freq* in ascending order.
6. Select SNPs with *C1 Freq* > 0.2 and *C1 Freq* < 0.8, and set *Comment* as "Polymorphic_Step1".
7. Select all SNPs and set *Aux* value to 0.
8. Select SNPs with # *Clusters* greater than 2, and set *Comment* as "Multiple Clusters".

**Step 2**

1. Select SNPs with # *Clusters* equal to 1 in SNP Table, right click over the selected SNPs, and choose *Cluster Selected SNPs* in the menu that appears, then select *2 Clusters* in the sub-menu.
2. Sort SNP Table by *Comment* in descending order, # *Clusters* and *Call Freq* in ascending order.
3. Select SNPs with # *Clusters* equal to 2 and *Call Freq* > 0.9, and set *Aux* value to 1.
4. Sort SNP Table by *Aux* in descending order, then by *C1 Freq* and *C2 Freq* in ascending order.
5. Select SNPs with *C1 Freq* > 0.2 and *C1 Freq* < 0.8, and set *Comment* as "Polymorphic_Step2".
6. Select all SNPs and set *Aux* value to 0.
7. Sort SNP Table by *Comment* in descending order, then by # *Clusters* and *Call Freq* in ascending order.
8. Select SNPs with # *Clusters* equal to 1, and set *Comment* as "Monomorphic".
9. Select SNPs with # *Clusters* equal to 2 and *Call Freq* < 0.2, and set *Comment* as "Monomorphic".

**Step 3**

1. Set DBSCAN clustering algorithm parameter *Cluster Distance* to 0.09. Increase the cluster distance allows the identification of clusters that were too broad to be detected in the first step.
2. Select SNPs in SNP Table that do not have an annotation in *Comment*, right click over the selected SNPs and choose *Cluster Selected SNPs* in the menu that appears, then select *DBScan*

in the sub-menu.

3. Sort SNP Table by *Comment* in descending order, then by *Cluster #* and *Call Freq* in ascending order.
4. Select SNPs with # *Clusters* equal to 2 and *Call Freq* > 0.9, and set *Aux* value to 1.
5. Sort SNP Table by *Aux* in descending order, then by *C1 Freq* and *C2 Freq* in ascending order.
6. Select SNPs with *C1 Freq* > 0.2 and *C1 Freq* < 0.8, and set *Comment* as "Polymorphic_Step3".
7. Select all SNPs and set *Aux* value to 0.
8. Select SNPs with # *Clusters* > 2, and set *Comment* as "Multiple Clusters".
9. Select SNPs with # *Clusters* equal to 1 and *Call Freq* > 0.99, and set *Comment* as "Monomorphic".

Finally, sort SNP Table by *Comment* and visually check clustering for SNPs marked as "Multiple Clusters". Manually curate incorrectly clustered SNPs.

**Appendix F.** Summary of the genetic map used for QTL analyses using 304 RILs (selection candidates).

| Chromosome | Length (cM) | Number of SNPs | SNP Density† | Minimum distance (cM)‡ | Maximum distance (cM)§ |
|---|---|---|---|---|---|
| 1A | 142.1 | 65 | 0.46 | 0.54 | 9.12 |
| 1B | 165.5 | 75 | 0.45 | 0.11 | 8.41 |
| 1D | 179.5 | 44 | 0.25 | 0.10 | 18.03 |
| 2A | 177.7 | 68 | 0.38 | 0.22 | 17.21 |
| 2B | 181.3 | 69 | 0.38 | 0.40 | 11.86 |
| 2D | 129.9 | 42 | 0.32 | 0.18 | 21.01 |
| 3A | 174.9 | 66 | 0.38 | 0.23 | 11.43 |
| 3B | 143.7 | 62 | 0.43 | 0.12 | 6.73 |
| 3D | 156.1 | 38 | 0.24 | 0.59 | 17.00 |
| 4A | 155.5 | 59 | 0.38 | 0.30 | 16.59 |
| 4B | 117.9 | 54 | 0.46 | 0.38 | 7.01 |
| 4D | 170.4 | 23 | 0.13 | 0.23 | 52.30 |
| 5A | 146.0 | 52 | 0.36 | 0.24 | 23.16 |
| 5B | 217.6 | 96 | 0.44 | 0.11 | 15.77 |
| 5D | 184.8 | 34 | 0.18 | 0.09 | 26.88 |
| 6A | 160.6 | 63 | 0.39 | 0.34 | 18.55 |
| 6B | 122.9 | 49 | 0.40 | 0.12 | 12.68 |
| 6D | 156.5 | 34 | 0.22 | 0.39 | 18.19 |
| 7A | 231.7 | 89 | 0.38 | 0.28 | 17.00 |
| 7B | 178.1 | 73 | 0.41 | 0.04 | 22.18 |
| 7D | 233.3 | 64 | 0.27 | 0.04 | 30.09 |
| **Whole genome** | **3526** | **1219** | **0.35** | **0.04** | **52.30** |

†Average number of SNPs per cM

‡Minimum genetic distance between adjacent SNPs

§Maximum genetic distance between adjacent SNPs

**Appendix G.** Summary of environment specific QTL identified for six agronomic traits based on 304 RILs evaluated across five environments. QTL analyses were conducted using LS-means of each environment.

| QTL | Trait† | Environment | Chromosome | Position (cM) | Confidence Interval | Left Marker | Right Marker | LOD | $R^2$ (%) | Add‡ |
|---|---|---|---|---|---|---|---|---|---|---|
| *QHd.usw-1B.1* | HD | Kernen-16 | 1B | 39.9 | 39.4 – 41.4 | Kukri_c14149_462 | BS00022304_51 | 5.2 | 2.4 | 0.5 |
| *QHd.usw-1B.2* | HD | Kernen-16 | 1B | 87.9 | 87.4 – 88.4 | Kukri_c14239_1995 | TA004264-0825 | 6.0 | 3.4 | -0.6 |
| *QHd.usw-7A* | HD | Rosthern-15 | 7A | 113.7 | 112.2 – 114.2 | BS00066015_51 | RAC875_c101928_381 | 13.3 | 6.6 | -0.9 |
| *QHd.usw-7D.2* | HD | Kernen-16 | 7D | 140 | 139.5 – 140.5 | Kukri_c34287_166 | IAAV3265 | 7.4 | 3.9 | -0.6 |
| *QTwt.usw-1D* | TWT | Kernen-16 | 1D | 100 | 97.5 – 103.5 | BS00110144_51 | wsnp_Ex_c35886_43950574 | 5.9 | 3.6 | 0.3 |
| *QTwt.usw-2B.1* | TWT | Rosthern-16 | 2B | 84.6 | 83.1 – 86.1 | Tdurum_contig53156_111 | Tdurum_contig28227_304 | 5.7 | 4.1 | -0.3 |
| *QTwt.usw-2D* | TWT | Kernen-14 | 2D | 94.9 | 93.4 – 95.4 | RAC875_c319_1776 | RAC875_c15518_236 | 5.8 | 4.9 | -0.3 |
| *QTwt.usw-3A* | TWT | Kernen-16 | 3A | 16.1 | 15.1 – 17.6 | RAC875_c51781_238 | wsnp_Ra_c9185_15386027 | 8.0 | 6.1 | -0.4 |
| *QTwt.usw-3D.1* | TWT | Kernen-16 | 3D | 73 | 72.5 – 75.5 | wsnp_Ex_c7260_12463738 | Kukri_c42075_156 | 5.3 | 4.0 | -0.3 |
| *QTwt.usw-3D.2* | TWT | Kernen-14 | 3D | 106 | 105.5 – 107.5 | wsnp_Ex_c8802_14726148 | Tdurum_contig67613_465 | 6.1 | 6.5 | -0.3 |
| *QTwt.usw-3D.2* | TWT | Kernen-15 | 3D | 107 | 106.5 – 107.5 | Tdurum_contig67613_465 | wsnp_Ra_rep_c71290_69343893 | 6.9 | 9.2 | -0.4 |
| *QTwt.usw-7D* | TWT | Kernen-14 | 7D | 104 | 103.5 – 104.5 | RAC875_c1834_694 | Ex_c19087_352 | 9.1 | 7.8 | 0.4 |
| *QTwt.usw-7D* | TWT | Kernen-15 | 7D | 96 | 95.5 – 97.5 | GENE-5000_1221 | D_contig63719_554 | 5.1 | 5.6 | 0.4 |
| *QTwt.usw-7D* | TWT | Kernen-16 | 7D | 96 | 95.5 – 97.5 | GENE-5000_1221 | D_contig63719_554 | 5.6 | 3.4 | 0.3 |
| *QMat.usw-3A* | MAT | Rosthern-15 | 3A | 86.1 | 83.6 – 86.6 | BS00073009_51 | BobWhite_c11225_941 | 5.2 | 3.7 | -0.5 |
| *QMat.usw-5A* | MAT | Kernen-16 | 5A | 66.3 | 65.8 – 66.8 | wsnp_BE495277A_Ta_2_5 | wsnp_Ku_c51039_56457361 | 6.2 | 6.3 | 0.7 |
| *QMat.usw-5B* | MAT | Rosthern-15 | 5B | 89 | 88.5 – 89.5 | wsnp_Ku_c3102_5810751 | RAC875_c36779_148 | 7.3 | 4.5 | 0.6 |
| *QMat.usw-7A* | MAT | Kernen-16 | 7A | 112.7 | 112.2 – 114.2 | wsnp_Ex_c2017_3787478 | BS00066015_51 | 17.4 | 8.7 | -1.0 |
| *QMat.usw-7A* | MAT | Rosthern-16 | 7A | 96.7 | 96.2 – 97.2 | Tdurum_contig11827_738 | BS00102773_51 | 5.9 | 9.3 | -0.6 |
| *QMat.usw-7B* | MAT | Kernen-14 | 7B | 33 | 27.5 – 35.5 | wsnp_Ex_c11658_18773086 | IACX198 | 7.1 | 5.8 | 1.0 |

**Appendix G.** Continued

| QTL | Trait† | Environment | Chromosome | Position (cM) | Confidence Interval | Left Marker | Right Marker | LOD | $R^2$ (%) | Add‡ |
|---|---|---|---|---|---|---|---|---|---|---|
| *QTkw.usw-1A* | TKW | Kernen-14 | 1A | 71.7 | 71.2 – 72.2 | Ex_c80400_458 | Kukri_c23350_433 | 5.6 | 3.9 | 0.7 |
| *QTkw.usw-1B* | TKW | Kernen-16 | 1B | 120.9 | 120.4 – 121.4 | CAP11_c599_115 | RFL_Contig16_132 | 7.7 | 7.7 | 0.8 |
| *QTkw.usw-2D* | TKW | Kernen-16 | 2D | 12.9 | 12.4 – 14.4 | D_contig39560_387 | BS00022276_51 | 5.0 | 3.5 | 0.5 |
| *QTkw.usw-5A* | TKW | Kernen-15 | 5A | 54.3 | 52.8 – 55.8 | IACX2581 | Ex_c19057_965 | 5.6 | 7.9 | -0.8 |
| *QTkw.usw-5B.1* | TKW | Kernen-14 | 5B | 175 | 174.5 – 175.5 | wsnp_Ex_c3874_7036132 | Excalibur_c23452_310 | 5.3 | 7.3 | 1.0 |
| *QTkw.usw-5B.2* | TKW | Rosthern-15 | 5B | 48 | 47.5 – 48.5 | BS00065390_51 | BobWhite_c45340_368 | 7.4 | 3.5 | -0.8 |
| *QTkw.usw-5D* | TKW | Kernen-16 | 5D | 190.9 | 189.4 – 192.4 | Excalibur_c91745_337 | BS00011794_51 | 5.0 | 3.6 | -0.6 |
| *QTkw.usw-5D* | TKW | Rosthern-16 | 5D | 190.9 | 189.4 – 192.4 | Excalibur_c91745_337 | BS00011794_51 | 5.2 | 3.3 | -0.8 |
| *QTkw.usw-7A* | TKW | Kernen-14 | 7A | 159.7 | 158.2 – 161.2 | RAC875_c24411_889 | RAC875_c47457_496 | 6.9 | 5.7 | -0.9 |
| *QTkw.usw-7A* | TKW | Rosthern-16 | 7A | 159.7 | 158.2 – 161.2 | RAC875_c24411_889 | RAC875_c47457_496 | 7.0 | 5.3 | -0.9 |
| *QHt.usw-1D* | HT | Kernen-16 | 1D | 28 | 26.5 – 28.5 | wsnp_Ex_c1358_2602235 | Tdurum_contig50555_632 | 4.9 | 4.5 | -1.9 |
| *QHt.usw-3B* | HT | Kernen-15 | 3B | 14 | 13.5 – 16.5 | RFL_Contig4531_1195 | Ra_c8459_632 | 4.9 | 4.7 | 2.1 |
| *QYld.usw-1A* | YLD | Rosthern-15 | 1A | 73.7 | 73.2 – 74.2 | Kukri_c23350_433 | Ex_c28144_1843 | 5.7 | 2.1 | 172.2 |
| *QYld.usw-1D* | YLD | Rosthern-15 | 1D | 84 | 83.5 – 84.5 | D_contig14507_369 | BS00066446_51 | 10.5 | 4.7 | -254.8 |
| *QYld.usw-2B* | YLD | Kernen-14 | 2B | 26.6 | 26.1 – 27.1 | wsnp_Ex_c25445_34710489 | GENE-1018_99 | 5.7 | 9.8 | -216.3 |
| *QYld.usw-5A* | YLD | Rosthern-16 | 5A | 74.3 | 72.8 – 75.8 | Excalibur_c76628_182 | BS00021669_51 | 5.0 | 7.6 | -213.1 |

†HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight

‡Add: Additive effect.

**Appendix H.** List of significant markers detected from fold specific single marker regression in the selection candidates. These markers were fitted as fixed effects in the GS + de novo GWAS model for genomic predictions in the selection candidates based on fivefold cross-validation.

| Trait† | CV Fold‡ | SNP ID | Chromosome | Position (cM) |
|---|---|---|---|---|
| HD | 1 | wsnp_CAP12_c812_428290 | 2D | 19.03 |
| HD | 1 | RAC875_c1834_694 | 7D | 103.64 |
| HD | 1 | D_contig63719_554 | 7D | 97.40 |
| HD | 2 | wsnp_CAP12_c812_428290 | 2D | 19.03 |
| HD | 2 | D_contig63719_554 | 7D | 97.40 |
| HD | 2 | RAC875_c1834_694 | 7D | 103.64 |
| HD | 3 | wsnp_CAP12_c812_428290 | 2D | 19.03 |
| HD | 3 | wsnp_Ex_c2054_3852564 | 7D | 93.65 |
| HD | 3 | D_contig63719_554 | 7D | 97.40 |
| HD | 4 | Kukri_c20975_765 | 7D | 114.05 |
| HD | 4 | D_GDRF1KQ02JPR1A_106 | 7D | 115.29 |
| HD | 4 | RAC875_c1834_694 | 7D | 103.64 |
| HD | 5 | D_contig63719_554 | 7D | 97.40 |
| HD | 5 | wsnp_CAP12_c812_428290 | 2D | 19.03 |
| HD | 5 | RAC875_c1834_694 | 7D | 103.64 |
| HT | 1 | RAC875_c28667_516 | 5B | 100.31 |
| HT | 1 | wsnp_CAP12_c812_428290 | 2D | 19.03 |
| HT | 1 | Kukri_rep_c68330_380 | 5B | 95.52 |
| HT | 2 | wsnp_CAP12_c812_428290 | 2D | 19.03 |
| HT | 2 | BS00047901_51 | 2D | 9.23 |
| HT | 2 | Kukri_c34967_226 | 6D | 83.44 |
| HT | 3 | wsnp_CAP12_c812_428290 | 2D | 19.03 |
| HT | 3 | Kukri_c34967_226 | 6D | 83.44 |
| HT | 3 | RAC875_c28667_516 | 5B | 100.31 |
| HT | 4 | Ra_c8459_632 | 3B | 14.10 |
| HT | 4 | Kukri_c34967_226 | 6D | 83.44 |
| HT | 4 | RAC875_c28667_516 | 5B | 100.31 |
| HT | 5 | wsnp_CAP12_c812_428290 | 2D | 19.03 |
| HT | 5 | BS00047901_51 | 2D | 9.23 |
| HT | 5 | RAC875_c28667_516 | 5B | 100.31 |

**Appendix H.** Continued

| Trait† | CV Fold‡ | SNP ID | Chromosome | Position (cM) |
|---|---|---|---|---|
| MAT | 1 | wsnp_CAP12_c812_428290 | 2D | 19.03 |
| MAT | 1 | D_contig63719_554 | 7D | 97.40 |
| MAT | 1 | BS00066015_51 | 7A | 113.30 |
| MAT | 2 | wsnp_CAP12_c812_428290 | 2D | 19.03 |
| MAT | 2 | wsnp_BE497845D_Ta_1_1 | 7D | 147.40 |
| MAT | 2 | D_contig63719_554 | 7D | 97.40 |
| MAT | 3 | wsnp_CAP12_c812_428290 | 2D | 19.03 |
| MAT | 3 | D_contig63719_554 | 7D | 97.40 |
| MAT | 3 | BS00066015_51 | 7A | 113.30 |
| MAT | 4 | Ex_c19087_352 | 7D | 105.30 |
| MAT | 4 | BS00066015_51 | 7A | 113.30 |
| MAT | 4 | wsnp_CAP12_c812_428290 | 2D | 19.03 |
| MAT | 5 | D_contig63719_554 | 7D | 97.40 |
| MAT | 5 | BS00066015_51 | 7A | 113.30 |
| MAT | 5 | RAC875_c1834_694 | 7D | 103.64 |
| YLD | 1 | wsnp_Ex_c25445_34710489 | 2B | 26.48 |
| YLD | 1 | Excalibur_c1787_1199 | 2A | 0.00 |
| YLD | 1 | Excalibur_rep_c101263_892 | 2B | 32.16 |
| YLD | 2 | Excalibur_c1787_1199 | 2A | 0.00 |
| YLD | 2 | wsnp_Ex_c25445_34710489 | 2B | 26.48 |
| YLD | 2 | Excalibur_rep_c101263_892 | 2B | 32.16 |
| YLD | 3 | wsnp_Ex_c25445_34710489 | 2B | 26.48 |
| YLD | 3 | Excalibur_c1787_1199 | 2A | 0.00 |
| YLD | 3 | Excalibur_rep_c101263_892 | 2B | 32.16 |
| YLD | 4 | Excalibur_c1787_1199 | 2A | 0.00 |
| YLD | 4 | wsnp_Ex_c25445_34710489 | 2B | 26.48 |
| YLD | 4 | Excalibur_rep_c101263_892 | 2B | 32.16 |
| YLD | 5 | wsnp_Ex_c25445_34710489 | 2B | 26.48 |
| YLD | 5 | Excalibur_c1787_1199 | 2A | 0.00 |
| YLD | 5 | Excalibur_rep_c101263_892 | 2B | 32.16 |

**Appendix H.** Continued

| Trait† | CV Fold‡ | SNP ID | Chromosome | Position (cM) |
|---|---|---|---|---|
| TWT | 1 | wsnp_Ex_c14026_21924297 | 4B | 59.51 |
| TWT | 1 | Tdurum_contig33737_157 | 4B | 55.96 |
| TWT | 1 | RAC875_c27536_611 | 4B | 54.64 |
| TWT | 2 | wsnp_Ex_c14026_21924297 | 4B | 59.51 |
| TWT | 2 | RAC875_c27536_611 | 4B | 54.64 |
| TWT | 2 | Tdurum_contig33737_157 | 4B | 55.96 |
| TWT | 3 | wsnp_Ex_c14026_21924297 | 4B | 59.51 |
| TWT | 3 | Tdurum_contig33737_157 | 4B | 55.96 |
| TWT | 3 | Excalibur_c52517_464 | 4B | 61.84 |
| TWT | 4 | tplb0037m09_1556 | 7B | 69.57 |
| TWT | 4 | wsnp_Ex_c14026_21924297 | 4B | 59.51 |
| TWT | 4 | RAC875_c22539_484 | 6B | 73.42 |
| TWT | 5 | wsnp_Ex_c14026_21924297 | 4B | 59.51 |
| TWT | 5 | Tdurum_contig33737_157 | 4B | 55.96 |
| TWT | 5 | CAP8_rep_c4633_93 | 4D | 80.68 |
| TKW | 1 | Ku_c1125_814 | 4A | 100.38 |
| TKW | 1 | RAC875_c95150_286 | 4A | 102.43 |
| TKW | 1 | wsnp_Ex_c1556_2972715 | 6A | 82.38 |
| TKW | 2 | wsnp_Ex_c1556_2972715 | 6A | 82.38 |
| TKW | 2 | Ku_c1125_814 | 4A | 100.38 |
| TKW | 2 | RAC875_c95150_286 | 4A | 102.43 |
| TKW | 3 | Ku_c1125_814 | 4A | 100.38 |
| TKW | 3 | RAC875_c95150_286 | 4A | 102.43 |
| TKW | 3 | RAC875_c59673_500 | 4A | 103.76 |
| TKW | 4 | Ku_c1125_814 | 4A | 100.38 |
| TKW | 4 | RAC875_c59673_500 | 4A | 103.76 |
| TKW | 4 | RAC875_c95150_286 | 4A | 102.43 |
| TKW | 5 | Ku_c1125_814 | 4A | 100.38 |
| TKW | 5 | RAC875_c95150_286 | 4A | 102.43 |
| TKW | 5 | RAC875_c59673_500 | 4A | 103.76 |

†HD: heading date, HT: plant height, MAT: maturity, YLD: grain yield, TWT: test weight, TKW: thousand-kernel weight.

‡CV: cross-validation