Genetic Architecture of Water-Soluble Carbohydrates in Wheat

Ben Ovenden BSc, BAppSci(Hons)

A thesis submitted to Charles Sturt University in fulfilment of the requirements for the degree of Doctor of Philosophy

School of Agricultural and Wine Sciences
Faculty of Science
Charles Sturt University

Wagga Wagga, Australia
August, 2015
Certificate of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma at Charles Sturt University or any other educational institution, except where due acknowledgment is made in the thesis. Any contribution made to the research by colleagues with whom I have worked at Charles Sturt University or elsewhere during my candidature is fully acknowledged. I agree that this thesis be accessible for the purpose of study and research in accordance with the normal conditions established by the Executive Director, Division of Library Services or nominee, for the care, loan and reproduction of theses.

Ben Ovenden

Wagga Wagga, 7 August 2015.
Acknowledgements

I would like to express my deepest appreciation for the direction, advice and time commitment throughout the six years it has taken me to complete this thesis provided from my supervisors: Professor Len Wade (CSU), Dr Greg Rebetzke (CSIRO) and Dr Andrew Milgate (NSW DPI). In hectic schedules they found time to guide me and somehow keep me on track. I have greatly benefited from Andrew for allowing me to work on his Irrigated Winter Cereals project experiments that made this study possible.

This thesis simply could never have been completed without the input and guidance on many aspects from Professor Jim Holland (North Carolina State University). I am particularly grateful for his generous hospitality in Raleigh and patient explanation and re-explanation of many quantitative genetics concepts.

I would like to acknowledge the receipt of an Australian Postgraduate Award scholarship from CSU, and a Grains Research Scholarship top-up from the GRDC to support me during my studies. Experiments for this thesis were part of the Irrigated Winter Cereals project (ICF00007), funded by GRDC, the Irrigated Cropping Forum, and NSW DPI. I would like to express my gratitude to the American Australian Association for the 2012 ConocoPhillips Education Fellowship to support my studies at North Carolina State University.

Many of the statistical methods I have used in this thesis were a completely foreign concept to me when I started my project. I would like to thank Chris Lisle and Sharon Nielsen (both CSU), for teaching me the analysis methods I now use almost every day. Learning to use R and ASReml has been particularly useful, and I look forward to learning more in years to come. It's scary how fun data analysis can be!
This thesis involved the analysis of thousands of samples and field measurements to obtain the dataset used. The list of people that provided technical help with these measurements is too long to mention them all by name, but I would particularly like to acknowledge Kerry Schirmer (CSU) for working through a mountain of samples with me over many months of sampling, counting, grinding, re-randomising, packing and NIRS preparation. I would also like to thank Aaron Hutchison (NSW DPI) for technical assistance in field measurements and compiling the weather and agronomy data for the Irrigated Winter Cereals experiments.

Last but not least, I would like to show my appreciation to the people that encouraged me for the duration of my studies: Katrina and my Mum and Dad for everything, Dr John Harper for his camaraderie, and Dr Peter Snell for his generosity and patience through constant support and doubtless work on my behalf in the rice breeding program at Yanco.
Abstract

Water-soluble carbohydrates (WSC) accumulate in wheat when assimilate supply is greater than assimilate demand from sinks. This characteristic has been identified as a potentially useful drought avoidance trait, as plants that accumulate carbohydrate reserves can remobilise that assimilate to contribute to grain filling under unfavourable growing conditions such as terminal drought. This study investigated a population of breeding germplasm constrained for maturity to explore the genetic control of WSC accumulation and the potential for genetic improvement of this trait. Experiments were conducted over two locations and two years with rainfed and irrigated treatments at each location to provide the multi-environment trial dataset.

The results of this study show that WSC accumulation involves complex interactions with other traits and the environment. For both WSC concentration (WSCC) and total WSC per m² (WSCA) strong environmental interactions were observed, with experiments clearly grouping into well-watered and water deficit environment clusters. Within these clusters, genetic correlations between experiments were high. Heritability for WSCC was higher than for WSCA, due to the difficulty in obtaining an accurate biomass measure, consequentially WSCC was used in the subsequent marker analysis. There was no correlation between grain yield and WSCC or WSCA, although significant correlations were observed in both well-watered and water deficit experiment clusters for reduced tillering, fewer grains per m² and greater grain weight.

Genome-wide association studies for WSCC were performed in each experiment, by employing two different marker sets and two analysis pipelines, and using the identification of known flowering time loci via the same methodology as a
positive control. Two quantitative trait loci (QTL) for WSCC were identified at multiple experiments and with both analysis pipelines. One of these markers has been previously associated with grain yield under water-limited conditions, and the other was located close to the Glu-1D locus controlling expression of glutenin grain storage proteins. Few loci for WSCC collocated with loci associated with flowering time. Notably, the association studies did not identify any major loci for WSCC that explained a large proportion of phenotypic variance.

The possible application of genomic selection was assessed by developing factor analytic models. These were then used to generate genomic estimated breeding values (GEBVs) from the whole-genome marker profiles for both the well-watered and water deficit environment clusters of experiments. The GEBVs for each of these target populations of environments showed 65-70% relative accuracy compared to phenotypic predictions. This translates to an approximately 30-40% greater genetic gain for genomic selection per unit time when the duration of the breeding cycle is reduced by half at the same time. Interestingly, the additive GEBVs were correlated with the non-additive genetic variance component, indicating possible phenotypic epistasis or genetic correlation between underlying component traits. However, the nature of whole genome marker profiles needs to be taken into consideration when applying them to statistical genetics models, as statistical measures of additivity are not necessarily additive genetic effects in the biological sense.

The empirical progression of methods used in this study highlight the nature of WSC genetic architecture as complex, polygenic, but ultimately a candidate for selection in breeding programs. This study demonstrates how some progress can be made towards genetic improvement of a complex trait even when the nature of inheritance is not completely understood.
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>Akaike Information Criterion</td>
</tr>
<tr>
<td>CIMMYT</td>
<td>International Maize and Wheat Improvement Center</td>
</tr>
<tr>
<td>CR</td>
<td>Correlated response to selection</td>
</tr>
<tr>
<td>DArT</td>
<td>Diversity arrays technology</td>
</tr>
<tr>
<td>DMC</td>
<td>Dry matter content</td>
</tr>
<tr>
<td>E-BLUP</td>
<td>Empirical Best Linear Unbiased Predictor</td>
</tr>
<tr>
<td>EGG</td>
<td>Estimated genetic gain, or response to selection</td>
</tr>
<tr>
<td>FA</td>
<td>Factor analysis</td>
</tr>
<tr>
<td>FDR</td>
<td>False discovery rate</td>
</tr>
<tr>
<td>G × E</td>
<td>Genotype × environment interaction</td>
</tr>
<tr>
<td>GAPIT</td>
<td>Genome Association and Prediction Integrated Tool</td>
</tr>
<tr>
<td>G-BLUP</td>
<td>Genomic best linear unbiased prediction</td>
</tr>
<tr>
<td>GEBV</td>
<td>Genomic estimated breeding value</td>
</tr>
<tr>
<td>GWAS</td>
<td>Whole genome association study</td>
</tr>
<tr>
<td>HRZ</td>
<td>High rainfall zone</td>
</tr>
<tr>
<td>IBD</td>
<td>Inbreeding by descent</td>
</tr>
<tr>
<td>LD</td>
<td>Linkage disequilibrium (gametic phase disequilibrium)</td>
</tr>
<tr>
<td>MAF</td>
<td>Minor allele frequency</td>
</tr>
<tr>
<td>MAS</td>
<td>Marker-assisted selection</td>
</tr>
<tr>
<td>MET</td>
<td>Multi-environment trial</td>
</tr>
<tr>
<td>NIRS</td>
<td>Near-infrared spectroscopy</td>
</tr>
<tr>
<td>QTL</td>
<td>Target population of environments</td>
</tr>
<tr>
<td>RSE</td>
<td>Relative selection efficiency</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>SSR</td>
<td>Single-sequence repeat</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>WGAIM</td>
<td>Whole genome average interval mapping</td>
</tr>
<tr>
<td>WSC</td>
<td>Water-soluble carbohydrate</td>
</tr>
<tr>
<td>WSCA</td>
<td>Total water-soluble carbohydrate per unit area</td>
</tr>
<tr>
<td>WSCC</td>
<td>Water-soluble carbohydrate concentration</td>
</tr>
<tr>
<td>WSCT</td>
<td>Water-soluble carbohydrate per tiller</td>
</tr>
<tr>
<td>WUE</td>
<td>Water-use efficiency</td>
</tr>
<tr>
<td>ZAD</td>
<td>Zadoks development score, in this thesis specifically relative maturity around flowering time</td>
</tr>
</tbody>
</table>
# Table of Contents

1. General Introduction ............................................................................................................ 1

2. Literature Review ................................................................................................................ 4

   2.1 Introduction ................................................................................................................... 4

   2.2 Drought is a major limitation to wheat production in south-eastern Australia .. 4

   2.3 Current progress in wheat breeding for grain yield (water productivity) in drought environments .............................................................. 5

   2.4 Physiological mechanisms for water-soluble carbohydrate accumulation in wheat ............................................................................................................................... 7

   2.5 Association of water-soluble carbohydrate accumulation with other traits ....... 9

   2.6 Genetic improvement of water-soluble carbohydrate accumulation capacity .. 11

   2.7 Previous studies for the genetic control of water-soluble carbohydrate accumulation ............................................................................................................................ 13

   2.8 Genetic analysis of water-soluble carbohydrates ...................................................... 16

   2.9 The relationship between relative maturity and water-soluble carbohydrate concentration ......................................................................................................................... 22

   2.10 The genetics of flowering time in wheat ................................................................. 24

   2.11 Progress in genetic mapping and analysis methodology .................................... 25

   2.12 The role of linkage disequilibrium in molecular quantitative genetics ........... 26

   2.13 Principles of genome-wide association studies .................................................... 28

   2.14 Factors that affect the outcome of genome-wide association studies used for plant breeding .................................................................................................................. 29

   2.15 Identifying causal associations using genome-wide association studies ....... 31

   2.16 ‘Missing’ heritability and the relationship between heritability in quantitative genetics and genome-wide association studies ......................................................... 33

   2.17 Genomic selection principles and background .................................................. 35

   2.18 Comparing genomic selection to marker-assisted selection ......................... 39

   2.19 Models and methods used to generate genomic predictions .................................. 40

   2.20 Accuracy of genomic selection and genetic gain ............................................. 41

   2.21 Genomic selection and genotype × environment interactions .......................... 46
2.22 Discussion of the literature reviewed ......................................................... 48

3. Overall Aims and Objectives of this Thesis .................................................. 53

4. Water-soluble carbohydrate genetic × environment interactions and selection potential ........................................................................................................... 54

4.1 Introduction .................................................................................................. 54

4.2 Aims and objectives .................................................................................. 55

4.3 Materials and methods ............................................................................. 55

4.3.1 Genotypes ............................................................................................... 55

4.3.2 Field experiment locations ...................................................................... 57

4.3.3 Experimental design ............................................................................... 59

4.3.4 Field experiment management and agronomy ....................................... 61

4.3.5 Field experiment weather, rainfall and irrigation ................................. 62

4.3.6 Measurements and observations ............................................................ 67

4.3.7 Statistical methods ................................................................................ 69

4.4 Results ....................................................................................................... 74

4.4.1 Trait means for agronomic and quality characteristics ....................... 74

4.4.2 Genotype × environment interactions ................................................... 84

4.4.3 Trait correlations .................................................................................... 85

4.4.4 Heritability, estimated genetic gain, and correlated response to selection 98

4.4.5 Variance partitioning in water-soluble carbohydrate concentration and total water-soluble carbohydrate basic models .................................. 103

4.4.6 Coefficient of determination between the well-watered and water deficit environment cluster predicted values ................................................ 105

4.5 Discussion ................................................................................................ 108

4.5.1 The target population of environments for water-soluble carbohydrate selection and genotype × environment interactions ............................. 108

4.5.2 Selection for water-soluble carbohydrate concentration and total watersoluble carbohydrates in well-watered and water deficit target populations of environments ............................................................................................................ 109
4.5.3 Using dry matter content as an indirect method of selection for water-soluble carbohydrate accumulation ........................................................................................................... 111

4.5.4 Correlated response of grain yield and maturity from selection for water-soluble carbohydrates ............................................................................................................. 112

4.5.5 Correlation of water-soluble carbohydrate concentration with other component traits .......................................................................................................................... 113

4.5.6 Potential to utilise water-soluble carbohydrate accumulation as a drought avoidance trait .................................................................................................................. 116

4.6 Conclusions ........................................................................................................................................................................................................................................... 116

5. Genome-Wide Association Studies for Water-Soluble Carbohydrate Concentration .............................................................................................................................. 118

5.1 Introduction ............................................................................................................................................................................................................................................. 118

5.2 Aims ...................................................................................................................................................................................................................................................... 119

5.3 Materials and methods ..................................................................................................................................................................................................................... 119

5.3.1 Genetic material and phenotyping for water-soluble carbohydrate concentration and relative maturity at flowering time .................................................................................................................. 119

5.3.2 Marker sets and statistical analysis of marker information .............................................................................................................................................................. 120

5.3.3 Linkage disequilibrium analysis ..................................................................................................................................................................................................................... 121

5.3.4 Genome-wide association study methods for GAPIT .............................................................................................................................................................................. 122

5.3.5 Genome-wide association study methods for WGAIM ............................................................................................................................................................................. 122

5.4 Results ............................................................................................................................................................................................................................................. 125

5.4.1 Heritability, population structure and linkage disequilibrium .............................................................................................................................................................................. 125

5.4.2 Genome-wide association study results for relative maturity at flowering time using GAPIT ............................................................................................................................................................................. 131

5.4.3 Genome-wide association study results for water-soluble carbohydrate concentration using GAPIT ............................................................................................................................................................................. 131

5.4.4 Genome-wide association study results for relative maturity at flowering time using WGAIM ............................................................................................................................................................................. 132

5.4.5 Genome-wide association study results for water-soluble carbohydrate concentration using WGAIM ............................................................................................................................................................................. 132

5.4.6 Markers for water-soluble carbohydrate concentration and relative maturity common between both GAPIT and WGAIM analyses ............................................................................................................................................................................. 133

5.5 Discussion ............................................................................................................................................................................................................................................. 161
5.5.1 Comparison of analysis at individual experiments........................................161
5.5.2 Comparison of WGAIM and GAPIT analysis pipelines..............................162
5.5.3 Comparison of DArT and SNP molecular marker sets..............................163
5.5.4 Loci associated with relative maturity at flowering time..........................165
5.5.5 Loci associated with both water-soluble carbohydrate concentration and relative maturity at flowering time .................................................................167
5.5.6 Loci associated with water-soluble carbohydrate concentration previously associated with other traits .................................................................168
5.5.7 Loci associated with water-soluble carbohydrate concentration across experiments and analysis pipelines .................................................................168
5.5.8 Genome-wide association study methodology limitations....................171
5.6 Conclusions ..............................................................................................173

6. Genomic Selection for Water-Soluble Carbohydrate Accumulation................176
6.1 Introduction ..........................................................................................176
6.2 Aims......................................................................................................177
6.3 Materials and methods ........................................................................177
6.3.1 Genetic material, phenotyping methods and the molecular marker information used in this study .................................................................177
6.3.2 Genomic selection mixed linear models..............................................177
6.3.3 Factor analytic model selection for genomic prediction.........................179
6.3.4 Cross validation of the G-BLUP model genomic estimated breeding value predictions .................................................................183
6.4 Results ..................................................................................................184
6.4.1 Genotypic and genotype × environment interaction variance explained by the model ..................................................................................................184
6.4.2 Genomic estimated breeding values for water-soluble carbohydrate concentration ..................................................................................................185
6.4.3 Genomic predictions for carbohydrate accumulation............................186
6.4.4 Cross-validation of genomic estimated breeding values.......................195
6.5 Discussion ............................................................................................199
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.5.1</td>
<td>Integration of genomic selection with genotype × environment effects</td>
<td>199</td>
</tr>
<tr>
<td>6.5.2</td>
<td>Practical application of genomic selection for water soluble carbohydrate improvement</td>
<td>202</td>
</tr>
<tr>
<td>6.5.3</td>
<td>Accuracy of genomic predictions</td>
<td>206</td>
</tr>
<tr>
<td>6.5.4</td>
<td>Correlations between the additive and non-additive genetic variance components in the model</td>
<td>208</td>
</tr>
<tr>
<td>6.6</td>
<td>Conclusions</td>
<td>212</td>
</tr>
<tr>
<td>7.</td>
<td>General Discussion</td>
<td>214</td>
</tr>
<tr>
<td>7.1</td>
<td>Genetic architecture of water-soluble carbohydrate accumulation</td>
<td>214</td>
</tr>
<tr>
<td>7.2</td>
<td>Selection for increased water-soluble carbohydrate accumulation</td>
<td>216</td>
</tr>
<tr>
<td>7.3</td>
<td>Genotype × environment interactions in quantitative trait analysis</td>
<td>218</td>
</tr>
<tr>
<td>7.4</td>
<td>Considerations on the methodology used in this study</td>
<td>220</td>
</tr>
<tr>
<td>7.5</td>
<td>Implications and applications of this thesis</td>
<td>222</td>
</tr>
<tr>
<td>8.</td>
<td>Conclusions</td>
<td>224</td>
</tr>
<tr>
<td></td>
<td>Appendix</td>
<td>228</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>252</td>
</tr>
</tbody>
</table>
Table 2.1: Loci for water soluble carbohydrates from previous studies (*UA is unlinked A genome marker).................................................................................................................... 19

Table 4.1: Predicted experiment mean confidence intervals (95%) for each trait. WSCC (mg/g), WSCA (g/m²), anthesis biomass (g/m²), dry matter content (DMC), tillers per m² at anthesis, height at maturity (cm), lodging score (scale 0-9 where 9 = completely lodged), relative maturity (ZAD) (Zadoks scale 0-99), thousand grain weight (g), screenings <2mm (%), grain protein content (%), and grain yield (t/ha.)..........................77

Table 4.2: Pearson’s product-moment correlation coefficient (r), coefficient of determination (r²), correlation p-value, and correlation 95% confidence intervals (CI) for trait E-GLUPs with WSCC at both well-watered (WW) and water deficit (WD) environment clusters, following Figure 4.16 and Figure 4.17. .............................................. 96

Table 4.3: Pearson’s product-moment correlation coefficient (r), coefficient of determination (r²), correlation p-value, and correlation 95% confidence intervals (CI) for trait E-GLUPs with WSCC at both well-watered (WW) and water deficit (WD) environment clusters, following Figure 4.18 and Figure 4.19. ........................................... 97

Table 4.4: Line mean heritability of WSC concentration, total WSC amount, anthesis biomass, dry matter content ratio, grain yield and maturity. ........................................ 100

Table 4.5: Estimated genetic gain (EGG, equivalent to response to selection) for WSCC at 10% and 25% (EGG 10 and EGG25) selection intensities. The correlated response to selection (CR) of dry matter content ratio on WSCC, and WSCC for selection on anthesis biomass, grain yield and maturity is given at 10% and 25% (CR10 and CR25) selection intensities, with the relative selection efficiency (RSE) for each correlated response. ........................................................................................................ 101

Table 4.6: Estimated genetic gain (EGG, equivalent to response to selection) for WSCA at 10% and 25% (EGG 10 and EGG25) selection intensities. The correlated response to selection (CR) of dry matter content ratio on WSCA, and WSCA for selection on grain yield and maturity is given at 10% and 25% (CR10 and CR25) selection intensities, with the relative selection efficiency (RSE) for each correlated response........................................................................ 102

Table 4.7: Variance components for WSC concentration, total WSC amount, maturity and grain yield from a simplified mixed linear model showing variance partitioning between genotype (G), and the environment components site location (L), irrigation treatment (T) and year (Y). .................................................................................................................... 104

Table 6.1 Factor analytic models fitted to the dataset for genomic prediction. Increasing order factor models were assessed using AIC and Chi-squared tests. Model E in this table (additive:FA2/non-additive:FA2) shows a significant improvement in fit from both additive:FA1/non-additive:FA1 and additive:FA1/ non-additive:FA2. Higher order FA models were not possible to fit with the computing resources available......180
Table 6.2: Genetic variances calculated from the average of the covariances from each environment cluster of experiments in the genotype and G × E covariance matrix from the G-BLUPs model following Zila et al. (2013).............................................................................................................. 187

Table 6.3: Cross-environment predictive ability of GEBVs from each experiment, and from each environment cluster of experiments. Predictive ability is calculated as the correlation between the additive GEBV from the cross validation model (without phenotype values) and true breeding value (total selection index from the full model with marker and phenotype values). The training population should be predictive of breeding values in the validation population, and the correlation between the two provides an estimation of how accurate each set of additive-only GEBVs predict the total additive and non-additive selection index.......................................................................................... 197

Table 6.4 Cross validation relative accuracy by experiment and environment cluster. The predictive ability of the GEBVs model at each experiment and environment cluster is divided by the mean line heritability to provide an accuracy measure relative to phenotypic selection........................................................................................................................ 198

Table A.1: Markers significant for WSCC at more than one site from WGAIM analysis for both the SNP and DArT marker sets. Chromosome and position are from the consensus maps for each marker set. Nearby genes are from CMAP GrainGene database (http://wheat.pw.usda.gov/cmap/) searches................................................................. 228

Table A.2: Markers significant for ZAD at more than one site from WGAIM analysis for both the SNP and DArT marker sets. Chromosome and position are from the consensus maps for each marker set. Nearby genes are from CMAP GrainGene database (http://wheat.pw.usda.gov/cmap/) searches................................................................. 230

Table A.3: Markers significant for WSCC at more than one site from the GAPIT analysis for the DArT marker set. No SNP markers were significant for a FDR adjusted p-value <0.10 for site by site GAPIT analyses. Chromosome and position are from the consensus map. Nearby genes are from CMAP GrainGene database (http://wheat.pw.usda.gov/cmap/) searches. GAPIT analyses returned significant markers for the two water deficit environments 2009 Coleambally and Yanco Rainfed, and not any of the well-watered environment experiments......................................................... 247

Table A.4: Markers significant for ZAD at more than one site from GAPIT analysis for the DArT and SNP marker datasets. Chromosome and position are from the consensus map. Nearby genes are from CMAP GrainGene database (http://wheat.pw.usda.gov/cmap/) searches................................................................. 248
List of Figures

Figure 2.1 Reproduced from Gebbing (2003), this figure shows the rise and fall of WSC (in terms of both fructans and sucrose amount per tiller) relative to anthesis date. ......23

Figure 2.2 Schematic of an inbreeding species breeding program with genomic selection adapted from Smith (2014). Standard phenotypic selection means that the breeding cycle (between crosses) is at least four years, as lines need to reach at least the F1 generation to be sufficiently genetically fixed (homozygous) for a stable phenotype to be obtained. Genomic selection uses the phenotypes measured in the line evaluation phases of the breeding program (called the training population) in a genomic prediction model to predict genotypes which are more likely to have the phenotype of interest in the early generations of a breeding program. Identified superior genotypes can be utilized in crossing, thus reducing the breeding cycle time from four years to two years. ........................................................................................................................................38

Figure 2.3 Validation approaches in genomic selection (reproduced from Sallam et al., 2015). Subset, or cross-validation uses the same dataset for both training and prediction. Interset validation refers to the method of testing training population predictions in a separate prediction population. Progeny validation refers to a prediction population composed of progeny from parents present in the training population. ..................................................................................................................................................45

Figure 4.1 Histogram of the Yanco irrigated experiment Zadoks’ development score of relative maturity (ZAD) for 10 September 2009. Breeding lines with scores between Z49 and Z56 (early head emergence to 50% anthesis, bars in blue) were identified for WSC measurement. This range is equivalent to approximately 3-5 days difference in anthesis date. ........................................................................................................................................58

Figure 4.2 Aerial view of experiment site at Coleambally NSW in 2009. The irrigated treatments can be seen in the centre bays and rainfed treatments on either side. Similar layouts were utilised for other experiments. .................................................................................................................................60

Figure 4.3 Daily maximum and minimum temperatures recorded by on-site weather stations throughout the growing season. The first date of sowing and the final date of harvest at each site are also indicated........................................................................................................................................65

Figure 4.4 Daily rainfall totals and soil moisture potentials recorded by on-site weather stations throughout the growing season, including the timing of irrigations (‘irr’) to the irrigated experiments, as well as the first date of sowing (‘sow’) and the final date of harvest (‘har’) at each experiment ........................................................................................................................................66

Figure 4.5 Histogram of Zadoks development score for the 365 genotypes identified for WSC measurement at each experiment ........................................................................................................................................78

Figure 4.6 Histogram of Zadoks development score for all of the genotypes at each experiment (990) including the lines grown but not measured for WSC ........................................................................................................................................79

Figure 4.7 Histogram of anthesis biomass for the genotypes identified for WSC measurement at each experiment ........................................................................................................................................80
Figure 4.8 Histogram of WSC concentration (WSCC) for the genotypes at each experiment. .......................................................... 81

Figure 4.9 Histogram of WSC amount (WSCA) for the genotypes at each experiment. 82

Figure 4.10 Histogram of grain yield for the genotypes identified for WSC measurement at each experiment. .......................................................... 83

Figure 4.11 Heatmap of relative maturity at flowering time (Zadoks development score) genetic correlations between experiments ............................................. 87

Figure 4.12 Heatmap of anthesis biomass per m² genetic correlations between experiments. The across-experiments FA model did not include 09YANA_RFD as this experiment had no significant genetic variance in the model. ............................................. 88

Figure 4.13 Heatmap of genetic correlation of WSC concentration between experiments. .......................................................................................... 89

Figure 4.14 Heatmap of genetic correlation of WSC amount per m² between experiments. The across-experiments FA model did not include 09YANA_RFD as this experiment had no significant genetic variance in the model. ............................................. 90

Figure 4.15 Heatmap of genetic correlation of grain yield between experiments. ....... 91

Figure 4.16 Plots of WSC concentration E-BLUPs with grain yield, grains per m², anthesis biomass, tillers per m² and maturity E-BLUPs. Genotypes are coloured by maturity group, where blue = winter types, purple = facultative, green = longer maturity spring types, red = shorter maturity spring types, and yellow = short season genotypes. .......................................................................................... 92

Figure 4.17 Plots of WSC concentration E-BLUPs with grain protein, lodging, mature plant height, screenings and thousand grain weight E-BLUPs. Lodging was not measured at either of the two water deficit environment experiments (09YANA_RFD and 09COLE_RFD). Genotypes are coloured by maturity group following Figure 4.1693

Figure 4.18 Plots of WSC amount E-BLUPs with grain yield, grains per m², anthesis biomass, tillers per m² and maturity E-BLUPs. Genotypes are coloured by maturity group following Figure 4.16. The water deficit environment cluster consists of only one experiment (09COLE_RFD) as shown in Figure 4.14.......................................................... 94

Figure 4.19 Plots of WSC amount E-BLUPs with grain protein, lodging, mature plant height, screenings and thousand grain weight E-BLUPs. Genotypes are coloured by maturity group following Figure 4.16. The water deficit environment cluster consists of only one experiment (09COLE_RFD) as shown in Figure 4.14, and lodging was not scored at that experiment.......................................................... 95

Figure 4.20 WSC Concentration E-BLUPs plotted by environmental group. Genotypes are coloured into maturity groups using maturity E-BLUPs to illustrate correlation between environment groups within groups of genotypes with comparable maturity .......................................................... 106
Figure 4.21 WSC amount E-BLUPs plotted by environmental group. Genotypes are coloured into maturity groups using relative maturity E-BLUPs. The water deficit environment consists of E-BLUPs from 09COLE_RFD which was the only water deficit experiment in the across-experiments FA model ................................................................. 107

Figure 5.1: Kinship relationship matrix heatmap and dendrogram output from GAPIT for the SNP data set........................................................................................................ 127

Figure 5.2: Kinship relationship matrix heatmap and dendrogram output from GAPIT for the DArT data set......................................................................................................... 128

Figure 5.3: Pairwise LD estimates ($r^2$) plotted against Euclidian pairwise marker distances for markers on the same consensus chromosome for the SNP marker set. Chromosomes are plotted for each genome, and a second degree LOESS curve fitted to each scatter plot. The dashed line indicates the critical value of $r^2$, above which marker pairs are assumed to be physically linked. For the SNP marker set the critical $r^2$ value ($r^2=0.0470$) intersects the LOESS curve at 24cM for the A genome and 21cM for both the B and D genomes ............................................................................................................ 129

Figure 5.4 Pairwise LD estimates ($r^2$) plotted against Euclidian pairwise marker distances for markers on the same consensus chromosome for the DArT marker set. Chromosomes are plotted for each genome, and a second degree LOESS curve fitted to each scatter plot. The dashed line indicates the critical value of $r^2$, above which marker pairs are assumed to be physically linked. The critical value of $r^2$ for the DArT marker set ($r^2=0.0456$) intersects the LOESS curve at approximately 18cM for the A genome, 16cM for the B genome and 25cM for the D genome. .............................................................................................................. 130

Figure 5.5 Chromosome maps showing DArT markers for consensus chromosomes 1A and 1B. Significant loci for WSCC and ZAD at each experiment are shown for both WGAIM (blue) and GAPIT (green) analyses................................................................. 134

Figure 5.6: Chromosome maps showing DArT markers for consensus chromosomes 1D, 2A and 2B. Significant loci for WSCC and ZAD at each experiment are shown for both WGAIM (blue) and GAPIT (green) analyses................................................................. 135

Figure 5.7 Chromosome maps showing DArT markers for consensus chromosomes 2D and 3A. Significant loci for WSCC and ZAD at each experiment are shown for both WGAIM (blue) and GAPIT (green) analyses................................................................. 136

Figure 5.8 Chromosome maps showing DArT markers for consensus chromosomes 3B, 3D and 4A. Significant loci for WSCC and ZAD at each experiment are shown for both WGAIM (blue) and GAPIT (green) analyses................................................................. 137

Figure 5.9 Chromosome maps showing DArT markers for consensus chromosomes 4B, 4D and 5A. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis .............................................................................................................. 138

Figure 5.10 Chromosome maps showing DArT markers for consensus chromosomes 5B and 5D. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis.............................................................................................................. 139
Figure 5.11 Chromosome maps showing DArT markers for consensus chromosomes 6A, 6B and 6D. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis. ........................................................................................................................................................................ 140

Figure 5.12 Chromosome maps showing DArT markers for consensus chromosomes 7A, 7B and 7D. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis. ........................................................................................................................................................................ 141

Figure 5.13 Chromosome maps showing SNP markers for consensus chromosome 1A. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis. ........................................................................................................................................................................ 142

Figure 5.14 Chromosome maps showing SNP markers for consensus chromosome 1B. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis. ........................................................................................................................................................................ 143

Figure 5.15 Chromosome maps showing SNP markers for consensus chromosome 1D. No markers on this chromosome were associated with either ZAD or WSCC at multiple experiments. ........................................................................................................................................................................ 144

Figure 5.16 Chromosome maps showing SNP markers for consensus chromosome 2A. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis. ........................................................................................................................................................................ 145

Figure 5.17 Chromosome maps showing SNP markers for consensus chromosome 2B. Significant loci for WSCC and ZAD at each experiment are shown for the WGAIM (blue) analysis. ........................................................................................................................................................................ 146

Figure 5.18 Chromosome maps showing SNP markers for consensus chromosome 2D. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis. ........................................................................................................................................................................ 147

Figure 5.19 Chromosome maps showing SNP markers for consensus chromosome 3A. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis. ........................................................................................................................................................................ 148

Figure 5.20 Chromosome maps showing SNP markers for consensus chromosome 3B. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis. ........................................................................................................................................................................ 149

Figure 5.21 Chromosome maps showing SNP markers for consensus chromosome 3D and 4A. No markers on this chromosome were associated with either ZAD or WSCC at multiple experiments. ........................................................................................................................................................................ 150

Figure 5.22 Chromosome maps showing SNP markers for consensus chromosome 4B. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis. ........................................................................................................................................................................ 151

Figure 5.23 Chromosome maps showing SNP markers for consensus chromosome 4D. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis. ........................................................................................................................................................................ 152
Figure 5.24 Chromosome maps showing SNP markers for consensus chromosome 5A. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis. ............................................................................................................................... 153

Figure 5.25 Chromosome maps showing SNP markers for consensus chromosome 5B. Significant loci for WSCC and ZAD at each experiment are shown for the WGAIM (blue) analysis. ............................................................................................................................... 154

Figure 5.26 Chromosome maps showing SNP markers for consensus chromosome 5D. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis. ............................................................................................................................... 155

Figure 5.27 Chromosome maps showing SNP markers for consensus chromosome 6A. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis. ............................................................................................................................... 156

Figure 5.28 Chromosome maps showing SNP markers for consensus chromosome 6B and 6D. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis ............................................................................................................................... 157

Figure 5.29 Chromosome maps showing SNP markers for consensus chromosome 7A. No markers on this chromosome were associated with either ZAD or WSCC at multiple experiments. ............................................................................................................................... 158

Figure 5.30 Chromosome maps showing SNP markers for consensus chromosome 7B. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis. ............................................................................................................................... 159

Figure 5.31 Chromosome maps showing SNP markers for consensus chromosome 7D. No markers on this chromosome were associated with either ZAD or WSCC at multiple experiments. ............................................................................................................................... 160

Figure 6.1 Biplot of REML estimates of rotated loadings and percentage variance from the G-BLUPs model for additive effects. This shows the cluster of well-watered environment experiments weighted towards loading one, and the two water-deficit environment experiments weighted towards loading two. ............................................................................ 181

Figure 6.2 Biplot of REML estimates of rotated loadings and percentage variance from the G-BLUPs model for non-additive effects. Similar to the additive effects biplot in Figure 6.1, this figure shows the cluster of well-watered environment experiments weighted towards loading one, and the two water-deficit environment experiments weighted towards loading two ............................................................................ 182

Figure 6.3 Additive GEBV genotypic and G × E correlations between experiments. The experiments cluster strongly into a well-watered environment grouping of all of the 2010 experiments, 09COLE_IRR and 09YANA_IRR, and a water deficit cluster consisting of the 09COLE_RFD and 09YANA_RFD experiments. ............................................................................ 188

Figure 6.4 Non-additive selection index genotypic and G × E correlations between experiments for the G-BLUP model. The clustering of well-watered and water deficit
experiments is maintained compared to the equivalent heatmap for the GEBVs, with some change of experiment clustering in the well-watered environment cluster.

Figure 6.5 Combined additive and non-additive genotypic and G × E correlations between experiments for the GBLUP model. This heatmap shows a similar clustering of experiments to both the additive and non-additive only heatmaps with a cluster of well-watered and water deficit environments.

Figure 6.6 Total selection index (consisting of combined additive GEBVs and non-additive selection index) from the G-BLUPs model regressed on the phenotype-only E-BLUPs from the model used in Chapter 4 (Equation 4.1) for each environment cluster. For the water deficit cluster, \( r = 0.979, r^2 = 0.958 \), significant at \( p<0.05 \). For the well-watered cluster, \( r = 0.993, r^2 = 0.987 \), significant at \( p<0.05 \).

Figure 6.7 Additive GEBVs regressed on the total (additive and non-additive) selection index for each environment cluster. For the water deficit cluster, \( r = 0.761, r^2 = 0.579 \), significant at \( p<0.05 \). For the well-watered cluster, \( r = 0.833, r^2 = 0.694 \), significant at \( p<0.05 \).

Figure 6.8 Non-additive selection index regressed on the total additive and non-additive total selection index for each environment cluster. For the water deficit cluster, \( r = 0.863, r^2 = 0.744 \), significant at \( p<0.05 \). For the well-watered cluster, \( r = 0.819, r^2 = 0.671 \), significant at \( p<0.05 \).

Figure 6.9 Additive GEBVs regressed on the non-additive selection index for each environment cluster. For the water deficit cluster, \( r = 0.329, r^2 = 0.108 \), significant at \( p<0.05 \). For the well-watered cluster, \( r = 0.365, r^2 = 0.133 \), significant at \( p<0.05 \).
1. General Introduction

Drought has been a dominant issue in Australian wheat production since the beginning of wheat cultivation in this country (Richards, 1991). Water deficit conditions for wheat crops can reduce both grain yield and quality. Work in wheat physiology has identified a number of drought escape traits that may contribute to performance under water deficit conditions (Araus et al., 2002). One drought avoidance trait of particular interest in terms of varietal improvement is the ability of some wheat genotypes to store water soluble carbohydrates (WSC) in their stems. As Schnyder (1993) describes, this characteristic is a consequence of assimilate production being higher than sink consumption, concomitant with the plant being able to temporarily store the excess. Under favourable conditions this is most likely to peak after anthesis, when the developing grain does not yet exert a significant demand on the photosynthetic resources of the plant (Rebetzke et al., 2009). Total concentrations of these carbohydrates can reach up to 40% of total stem weight (Schnyder, 1993). This storage pool of assimilates is utilised in the developing reproductive tissues of the plant, contributing around 10-20% under well-watered conditions and as much as 30-50% of grain yield under terminal drought conditions (Bidinger et al., 1977; Schnyder, 1993), as well as larger grain size (Rebetzke et al., 2009).

The duration and degree of WSC accumulation is influenced by a few key factors. In particular, reduced tiller numbers are associated with increased WSC (Rebetzke et al., 2008). Results imply that the reduced demand for assimilate for vegetative growth in reduced tillering wheats lead to greater levels of assimilate stored. High nitrogen (N) content was associated with lower WSC accumulation by van Herwaarden et al. (1998b) and Ruuska et al. (2006). Other traits known to influence
WSC accumulation include major dwarfing genes, morphology of wheat stems and WSC-related enzyme activity (Borrell et al., 1993; Zhang et al., 2008; Saint Pierre et al., 2010).

These confounding factors make genetic analysis difficult. Although WSC has been indirectly selected for in the UK and WA in general trends of varietal improvement (Rebetzke et al., 2009), few current studies have investigated the inheritance and genetic control of this trait. Ruuska et al. (2006) studied genetic variation for stem WSC concentration in a group of 22 Australian and international varieties. They concluded that variation in WSC concentration measured over multiple environments is strongly correlated with genetic variation, therefore breeding for increased stem WSC concentration should be possible. In a related study, Rebetzke et al. (2008) used three biparental populations to map quantitative trait loci (QTL) for WSC concentration and total WSC per unit of crop area. The number of QTL of small effect found indicated that WSC concentration (and total WSC) is controlled by multiple minor alleles, indicated as additive by the lack of evidence for epistasis or linkage disequilibrium. Rebetzke et al. (2008) suggest selection for total WSC, rather than WSC concentration which would favour the selection of earlier flowering, lower tillering genotypes. Genetic control of WSC was also exploited by Yang et al. (2007) by investigating a Chinese wheat population under well-watered and drought stress conditions. They found several QTL for WSC concentration at anthesis, with stronger associations in the drought stress environment.

The potential to increase the genetic capacity for WSC accumulation is an opportunity to improve the drought avoidance ability of Australian wheat varieties. The physiology of this trait is not well understood, although it is known to be complex, with relationships between WSC and a number of other important traits documented.
Some preliminary genetic studies have had promising results, however, a coordinated study of the physiological relationships and genetic components is required to determine the most effective way forward to select for WSC accumulation. Both conventional selection and QTL analysis need to be explored, and related to the physiology of WSC accumulation. Unraveling the genetic architecture of this complex trait will provide the next step towards producing high WSC accumulating varieties.

In this study the nature and determinants of WSC accumulation, as well as approaches to investigate complex traits such as WSC are reviewed in Chapter 2. The G × E interactions and potential for selection of this trait are explored in Chapter 4. Chapter 4 provides the foundation for Chapter 5, where genome wide association studies are employed to identify marker-trait associations for WSC concentration, and Chapter 6 where the application of genomic selection to improvement of WSC concentration is investigated. This thesis concludes by considering the implications of the complexity of genetic control of WSC accumulation, and opportunities for varietal improvement of this trait.
2. Literature Review

2.1 Introduction

This thesis describes an investigation of the genetic architecture of water-soluble carbohydrates (WSC) in wheat, specifically the accumulation of WSC over the lifecycle of the plant that is later remobilised to fill grain. In doing so it is important to understand the value of WSC accumulation as a breeding target, and the potential methods available to improve WSC accumulation and illuminate the nature of genetic control for this trait. Therefore, the review which follows is structured to cover the physiology of WSC accumulation and the usefulness of this trait for drought avoidance as well as the considerations involved in genome-wide association studies for complex traits like WSC accumulation. Finally this review covers the application of genomic selection to complex traits. From consideration of these concepts, a set of methodologies is formulated and used for the study of the genetic architecture of WSC in wheat.

2.2 Drought is a major limitation to wheat production in south-eastern Australia

Droughts are frequent in the Australian landscape, and water deficit during the growing season is a major limitation to wheat production nationally and also in south-eastern Australia (Richards, 1991). The reduction in grain yield and quality due to limitations in crop water availability reduces the profitability of farming systems, and potentially impacts on global food security (Ray et al., 2012). Most commonly, drought in south-eastern Australia occurs during grain filling at the end of the growing season, as rainfall tends to be evenly distributed or winter-dominant and vapour pressure deficit increases rapidly from late spring into summer (Howard, 2005; Chenu et al., 2013). However, terminal drought does not occur consistently, meaning the target
population of environments (TPE) for south-eastern Australia varies considerably, and field breeding programs will not encounter terminal drought selection pressure in every season (Chapman, 2008; Rebetzke et al., 2012). Further, the extent of drought severity tends to vary year-to-year (Chenu et al., 2013). The large-scale investigation of drought patterns of TPE in the Australian wheat belt by Chenu et al. (2013) identified four major environment types which were essentially categories of terminal drought stress, given early season drought is rarely a limit to production in the Australian wheat-belt. Across the wheat-belt, mild and moderate terminal drought environments were more likely to be encountered than stress-free conditions, with a similar pattern of drought occurrence to the national average modelled in south-eastern Australia. Similar observations based on historic climate data sets and modelling are also reported for north-eastern Australia by Chenu et al. (2011) and South Australia by Sadras and Rodriguez (2007).

2.3 Current progress in wheat breeding for grain yield (water productivity) in drought environments

The challenges of improving grain yield (and by extension water productivity) in water-limited environments is borne out by the historically low rates of genetic gain in water deficit environments compared to well-watered environments (for example, Araus et al., 2002; Richards et al., 2010). Drought incidence and severity, as well as the timing of drought during the growing season can differ markedly between sites and years in a breeding program multi-environment trial (MET). Together, these contribute to more complex genotype × environment interaction (hereafter G × E interaction) resulting in genotype rank changes for yield so that selection based on yield alone is less effective (Richards et al., 2002; Bennett et al., 2012; Rebetzke et al., 2012).
An alternative to selection for grain yield as a measure of performance is to identify useful traits that confer physiological adaptation to water deficit conditions. In theory these should translate to higher grain yield, and owing to their higher repeatability and reduced genetic complexity are more readily incorporated into new germplasm (reviewed by Ludlow & Muchow, 1990; Rebetzke et al., 2009; Richards et al., 2010). As outlined initially by Ludlow and Muchow (1990) and later by Passioura (1996) and Passioura (2007), there are many different interpretations of drought and the breeding strategies required to address performance under drought. Indeed, there is no appropriate universal approach for targeting drought tolerance, especially as Fleury et al. (2010) notes, selection for ‘drought resistance’ via drought survival traits would most probably translate to genotypes with uneconomically low yield potential. These drought resistance traits usually describe adaptations for maintenance of plant growth under water deficit which favour plant survival over productive growth, as discussed by Monneveux and Belhassen (1996). Instead, a promising avenue for breeding targets are drought escape and avoidance mechanisms, as cereal species and other annuals have a rich diversity of adaptations that assist them to avoid drought during their lifecycle rather than counter water deficit directly (Araus et al., 2002). Treating these adaptive traits as breeding targets is a promising way to improve wheat crop water use efficiency and overall productivity under water deficit (Tambussi et al., 2007; Rebetzke et al., 2009). An important aim for wheat genetic improvement in Australia is to assess and incorporate traits that confer performance under water limited environments into existing germplasm, and thus improve water-use efficiency (WUE) (Rebetzke et al., 2009). WUE can also be maximised through improved crop architecture, which is an adaptation of the crop design ideotype concept outlined by Donald (1968). The concept has already been deployed in the incorporation of major
genes for reduced height and photoperiod insensitivity to alter plant architecture and provide drought escape potential thus improving WUE (Passioura, 1996; Bennett et al., 2012). However many other traits are transitioning from physiological investigation to selection in breeding programs (Rebetzke et al., 2013). Water soluble carbohydrate (WSC) accumulation and remobilisation are highly promising traits with potential application to improve performance and water-productivity in wheat in south-eastern Australia (Richards et al., 2008; Rebetzke et al., 2009).

### 2.4 Physiological mechanisms for water-soluble carbohydrate accumulation in wheat

A number of plant families, including the Gramineae, have the ability to accumulate complex sugars as a carbohydrate reserve (Pollock & Cairns, 1991). Carbohydrate accumulation occurs when the crop synthesizes assimilate at a rate greater than needed by the various sinks (for example, developing florets, elongating shoots and roots). The excess carbohydrate is stored mainly in the lower parts of stems and culms with Gebbing (2003) observing that 88% of carbohydrate was stored in the enclosed heterotrophic part of the peduncle. Neither the roots nor leaves appear to be significant storage organs for carbohydrates (Gallagher et al., 1976). Most of the carbohydrate stored is in the form of fructans, with a minor component of sucrose and hexose (Schnyder, 1993; Wardlaw & Willenbrink, 1994), and are known as water soluble carbohydrates to distinguish them from cellulose carbohydrate cell walls. As Ruuska et al. (2006) reported, most of the variation in WSC between genotypes is due to variation in the quantity of fructans rather than other types of water soluble carbohydrates.

Both the accumulation and remobilisation of WSC is modified by environmental conditions that alter the balance between sources and sinks of
assimilate. In particular, the availability of source carbon (as sucrose) affects
accumulation (Xue et al., 2013). The WSC can be remobilised for use in growth or
respiration as shown by Kiniry (1993), who induced remobilisation with the use of
shading treatments to limit the supply of assimilate from photosynthesis. However the
main sink for remobilisation is the developing grain (Schnyder, 1993; van Herwaarden
et al., 1998a; Takahashi et al., 2001). Here, remobilised WSC contributes as much as 30-
50% of grain yield under terminal drought conditions, and around 10-20% under well-
watered conditions (Bidinger et al., 1977; Pheloung & Siddique, 1991; Schnyder, 1993;
Gebbing & Schnyder, 1999). Temporally, WSC accumulation increases from before
anthesis to a peak at 7-20 days after anthesis (Gebbing, 2003; Ehdaie et al., 2008; Zhu et
al., 2010) where the quantity of WSC can reach as much as 40% of stem weight
(Schnyder, 1993). After anthesis levels decline with remobilisation to other sinks.
Under water deficit conditions this peak can occur before anthesis (Goggin & Setter,
2004), and remobilisation is earlier and proportionally greater (Bidinger et al., 1977;

In terms of a physiological trait to improve the WUE of wheat crops under
terminal drought, it is the remobilisation of WSC and its contribution to grain yield
under water deficit that is of interest to plant breeders, although fructans have also
been shown to play a minor role as an osmoprotectant (Livingston et al., 2009). As
Yang et al. (2007) show, there is clearly no simple relationship reflecting the influence
of WSC on grain yield. However, a number of studies show that WSC is an important
source of assimilate for grain-filling under water limited conditions (Bidinger et al.,
1977; Pheloung & Siddique, 1991; Gebbing & Schnyder, 1999; Foulkes et al., 2007b).
Some workers such as Zhang et al. (2008) contend that remobilisation of WSC is
separate to and more important than WSC accumulation, and as Schnyder (1993)
observed, stored WSC does not have to be remobilised to filling grain. It could be used in supporting other sinks or consumed in respiration. Despite this observation, both van Herwaarden et al. (1998a) and Ruuska et al. (2006) indicate that wheat is generally efficient at remobilisation and Takahashi et al. (2001) reported that rates of decline in stem WSC corresponded to the four main stages of grain-filling.

2.5 Association of water-soluble carbohydrate accumulation with other traits

Water soluble carbohydrate traits including both WSC concentration (WSCC – WSC/total biomass) and WSC content (WSCA – WSC per unit area) have been associated with a number of other physiological characteristics which reveal a complex interaction of source-sink relationships, and particularly where WSC accumulation is a feature of reduced sink demand compared to source capacity to produce assimilate. Remobilisation is then the capacity to move and relocate this assimilate to needy sinks including grain. As with many drought escape traits, flowering time is a key trait associated with WSC accumulation owing partially to the nature of patterns of WSC accumulation across crop growth stages (Passioura, 1996; Rebetzke et al., 2008). As discussed above in section 2.4, the total amount of WSC (WSCA) peaks after anthesis in well-watered conditions (Ehdaie et al., 2008), or potentially earlier under water deficit conditions, and then declines in vegetative tissue with the stages of grain-filling (Takahashi et al., 2001). This temporal variation (reflecting the developmental stage of the crop) can make WSC a difficult trait to phenotype. Indeed, the concentration of WSC varies from tiller to tiller reflecting differences in shoot elongation and demand for carbohydrate amid competing shoots growing in the same canopy. A crop will be at different stages of accumulation or remobilisation depending on development (Blum, 1998), and depending on environmental conditions when WSC peaks, WSC can be
influenced by the conditions around the time of flowering (e.g. water availability and heat stress) (Talukder et al., 2013). Despite this, genotype ranking for WSC appear to be maintained both before and after anthesis (Dreccer et al., 2009).

The accumulation of WSC has been associated with characteristics that affect the capacity to store WSC in stems. Examples of such traits are those affecting stem length (Borrell et al., 1993) and therefore also plant height, which are especially revealed in studies contrasting genotypes varying in plant height (e.g. tall and semi-dwarf germplasm) such as Blum (1998), Ruuska et al. (2006) and Ehdaie et al. (2008). Stem thickness and weight, otherwise measured together as stem biomass volume, have been associated with WSC accumulation both with WSC amount per tiller and concentration (Saint Pierre et al., 2010; Dreccer et al., 2013). Similarly, WSC has been found to vary with stem structure (Xue et al., 2008), where germplasm with cell walls containing less cellulose and hemicellulose having higher WSCC. This association appears to be different to the reported correlation of increased WSCC with increased dry weight to fresh weight ratio of tillers reported by both Ehdaie et al. (2008) and Xue et al. (2009). Further, Xue et al. (2009) suggest that dry weight to fresh weight ratio of biomass may be a useful indirect measure of WSC.

The numbers of stems or tillers has also been shown to be strongly associated with WSCA and WSCC where smaller tiller numbers are generally correlated with higher WSCA and WSCC (Rebetzke et al., 2008; Rattey et al., 2009; Dreccer et al., 2013). Similarly Duggan et al. (2005) reported that near-isogenic tiller-inhibition (‘TIN’) containing lines (where lines with the TIN allele had half the number of tillers compared with the wild-type, non-TIN lines) had between 5-9% higher WSCC, a trend also reported by Mitchell et al. (2012). Moreover, the TIN lines had heavier and thicker stems which may provide some advantage in storage of WSC (Duggan et al., 2005).
However, Dreccer et al. (2013) also found that genotype rankings for WSCC were still maintained between varieties at similar tiller densities.

Among different traits, the most widely reported relationship is the association between WSC and reduced grain number per m² (Hendrix et al., 1986; Rebetzke et al., 2008; Dreccer et al., 2009), and larger grain weight (Ruuska et al., 2006; Yang et al., 2007; Rebetzke et al., 2008; Rattey et al., 2009; Saint Pierre et al., 2010; Bennett et al., 2012; Zhang et al., 2014b). Further, these relationships are maintained under both well-watered and water deficit conditions. Dreccer et al. (2009) highlighted the possibility that storage of WSC may reduce available carbohydrate during floret formation, therefore potentially WSC accumulation may compete with developing florets and other growth sinks. As a consequence, reduced fertility could contribute to fewer grain numbers per m². However studies comparing nitrogen uptake and WSC challenge this hypothesis, as WSC content has been shown to be strongly negatively correlated with plant nitrogen content (Batten et al., 1993; Rebetzke et al., 2008). Both van Herwaarden et al. (1998a) and Ruuska et al. (2006) show that in most situations, nitrogen is usually growth limiting, and increased nitrogen uptake will promote the diversion of WSC to extra growth arising from additional tillers. Therefore WSC accumulation mechanisms are clearly not a competitive sink. This is also seen in Gebbing and Schnyder (1999) where an increase in nitrogen fertiliser treatment was associated with a decrease in the contribution of WSC to grain yield.

2.6 Genetic improvement of water-soluble carbohydrate accumulation capacity

Direct selection for increased WSC in cereal breeding programs has been advocated for some time (Blum, 1998). However inadvertent selection by breeders has already occurred in some environments. For example, Rebetzke et al. (2009) review this
trend in breeding programs worldwide, observing a trend for increasing WSCC with year of variety release for Western Australian and the International Maize and Wheat Improvement Center (CIMMYT) wheat breeding programs. Rebetzke et al. (2009) noted in particular that there is no such trend in eastern Australian breeding programs. Common features of the Western Australian and CIMMYT breeding target environments is a profound and consistent terminal drought (Fischer, 1979; Dreccer et al., 2008), whereas in the south-eastern Australia TPE terminal drought occurs less frequently, but tends to be long, continuous and more severe water stress events when they do occur (Sadras & Rodriguez, 2007; Chenu et al., 2011; Chenu et al., 2013). However, the hypothesis that consistent terminal drought is required for selection is not supported by Shearman et al. (2005), who reported that UK wheat cultivars showed increased WSCC with progressive year of release, and Foulkes et al. (2007a), who observed that in the UK environment, WSCC showed a positive correlation with yield under both favourable well-watered conditions as well as terminal drought. This has also been reported by Zhang et al. (2014a), who showed an increase in frequency of ‘favourable alleles’ for WSC with year of release in China from an experiment with drought stress, heat stress and irrigated environments. However, given the summer-dominant rainfall patterns of the experiment location it is not clear if the drought stress occurred early or late in the growing season.

A limited number of studies have reported genotypic variation for WSC traits in wheat and subsequent potential for direct selection in commercial breeding programs. Ruuska et al. (2006) assessed genetic variation for stem WSC concentration in a diverse group of 22 Australian and international wheat varieties. They concluded that variation in WSCC measured over multiple environments was repeatable reflecting robust genotypic differences. Thereby breeding for increased stem WSC
concentration and content should be possible. Similarly, Rebetzke et al. (2008) used three biparental populations to map QTL for WSC concentration (WSCC) and total WSC per unit area (WSCA). The large number of QTL of small genetic effect reported in their study indicated that WSCC (and WSCA) was controlled by multiple additive effects with little or no evidence for epistasis or linkage disequilibrium. Rebetzke et al. (2008) also suggested selection for total WSC (WSCA) rather than WSC concentration (WSCC) would favour the selection of earlier flowering, lower tillering genotypes. Yang et al. (2007) used a Chinese wheat mapping population in their investigation of genetic control of WSC under well-watered and drought stress conditions. They found several QTL for WSCC at anthesis, with QTL identified in the drought stress environment explaining more genetic variance than the well-watered environment.

Further indirect evidence of strong genetic control for WSC traits is provided by studies which show that genotypes maintain their rankings for WSC traits across environments (for example, Foulkes et al., 2007a; Dreccer et al., 2009; 2013).

### 2.7 Previous studies for the genetic control of water-soluble carbohydrate accumulation

A number of studies have previously reported genetic control and architecture for WSCC and related characters, usually in the context of a wider investigation of physiological contributors to yield. In some cases, QTL were reported and these are summarised for different studies in Table 2.1. Snape et al. (2007) provided the first report of QTL for WSCC, using a doubled-haploid population (n = 144) derived from the European wheat varieties Spark and Rialto. This study utilised multiple QTL analysis methods assessed in multiple environments (usually mild terminal drought), with concurrence across analysis methods and presence in more than two environments taken as significance of QTL. The Spark/Rialto population segregated for
the 1BL/1RS translocation, and a QTL for WSCC was collocated at this locus with a QTL for increased biomass. Significantly this population varied for grain size, which was correlated with grain yield.

Yang et al. (2007) conducted a single study consisting of testing at one site in one year using a doubled-haploid population (n = 150) between two Chinese varieties. The split block design of the experiment contained a water deficit and a well-watered treatment, and QTL were identified for water deficit, well-watered, and across experiment phenotypic data, where WSCC was measured at flowering, grain filling and maturity. A large number of QTL were identified, which the authors categorise into additive QTL (a total of 11 for WSCC at flowering, grain filling and maturity), and 16 pairs of QTL with epistatic interactions. The QTL effects were small, ranging between 1 and 4% of the total phenotypic variance. Yang et al. (2007) noted that WSCC increased under the water deficit treatment of their experiment which was opposite to the effects noted by Ruuska et al. (2006).

Both McIntyre et al. (2010) and Pinto et al. (2010) utilised the CIMMYT Seri/Babax mapping population which is reportedly notable for a constrained flowering time range, and segregation for the 1BL/1RS translocation. McIntyre et al. (2010) identified five QTL for WSCC with each explaining between 6 and 7% of the total phenotypic variation, from phenotypes measured over three locations and four years in Australia. Population size was greater than for the two previous populations (n = 194). Of these, the QTL on 1D and 4A were also identified in the low yielding water deficit environment dataset, while only one QTL on 7A was identified in all combinations of environments (water deficit, well-watered and combined across-experiment phenotypes). The QTL identified were not collocated with QTL for flowering time, but the QTL on chromosomes 1D and 4A were collocated with QTL for
plant height, grain yield and grain weight under the ‘low yield’ (water deficit) conditions. Pinto et al. (2010) phenotyped for WSCC over three environments and two years in field experiments in Mexico (n=167) and identified QTL on chromosomes 1B, 3B and 4A, all collocating for QTL for grain yield, plant height, canopy temperature and NDVI.

Bennett et al. (2012) investigated a range of physiological traits in the RAC875/Kukri mapping population (n = 368) including WSCC in two water deficit experiments. A number of QTL on chromosomes 2D, 3A and 6A were detected for WSCC, some of which collocated with QTL for grain size (thousand grain weight) and plant height. The authors observed that the allele conferring greater plant height also increased WSC levels, although not necessarily a greater peduncle length, where WSC is mainly stored.

Zhang et al. (2008); (2009) and (2015), as well as Xue et al. (2013) have taken a different approach to the studies above and focused on the fructan exohydrolase (1-FEH) genes involved in the biosynthesis of WSC. These genes are located on chromosomes 6A, 6B and 6D. Zhang et al. (2008) studied a collection of bread and durum wheat lines, and some wheat diploid wild relatives, finding the expression of 1-FEH genes was correlated with tissue-specific WSCC. They noted that some of the QTL identified in their study collocated with QTL identified in Yang et al. (2007), although with correlated traits and not specifically WSCC QTL. Xue et al. (2013) also investigated a number of other genes involved in the synthesis of fructans and sucrose and found that the expression of a number of them were correlated with leaf sucrose levels, which in turn were correlated to WSCC. This approach assumes that the 1-FEH genes identified are the rate-limiting step in the biochemical steps in pathways that accumulate and remobilise WSC within the plant, without consideration of the
availability of assimilate, or the demand for assimilate in other sinks. This may not be
the case, although a number of studies have identified loci significant forWSCC on
chromosome 6A and 6B (Yang et al., 2007; McIntyre et al., 2011; Bennett et al., 2012).

Numerous QTL forWSCC have also been identified in other species, notably
rice, where two populations have been studied (Nagata et al., 2002; Takai et al., 2005),
and perennial ryegrass (Turner et al., 2006). No other traits were measured in the
ryegrass study. However, results of both rice studies show consistency with the
physiological interactions observed in wheat. Here, some alleles associated with
increased WSCC were also associated with earlier flowering time and reduced grain
number per m² (Nagata et al., 2002), and separately high WSCC was correlated with
low grain number per m² (Takai et al., 2005).

Most of these studies advocate the development of markers for marker-assisted
selection (MAS) from the QTL identified. While this is the end objective, much remains
to be done and understood before QTL for any quantitative trait from an isolated and
diverse mapping population can be applied in the elite genetic backgrounds in a
breeding program. In particular, the TPE for selection (Wade et al., 1996; Wade et al.,
1999; Cooper et al., 2014) need to be understood, as well as the genetic background, or
breeding germplasm pool the QTL are to be deployed in.

2.8 Genetic analysis of water-soluble carbohydrates

The most comprehensive QTL study to date on the inheritance of WSCC was
conducted by Rebetzke et al. (2008). The authors investigated carbohydrate
accumulation and other related traits in a follow-on study from Ruuska et al. (2006),
who found moderate to high heritability for WSCC in a diverse set of germplasm over
a number of environments and years. Rebetzke et al. (2008) measured WSCC and the
associated traits of WSCA and WSCT (WSC per tiller) in three biparental wheat populations, Sunco/Tasman, CD87/Katepwa and Cranbrook/Halberd assessed across multiple environments. In total, Rebetzke et al. (2008) identified 36 QTL for WSCC, 21 QTL for WSCA and 33 QTL for WSCT (summarised in Table 2.1). These QTL were typically of small genetic effect consistent with quantitative inheritance and genetic variance reflected the accumulation of additive alleles.

The populations used in the analysis varied for the major developmental genes for photoperiod sensitivity and reduced plant height, which can be a major cause of the observed phenotypic variability for grain yield and other traits in a mapping population (Rattey et al., 2009; Bennett et al., 2012; Edae et al., 2014). In the populations assessed in Rebetzke et al. (2008) these include the photoperiod sensitivity locus of Ppd-D1 in the Cranbrook/Halberd and CD87/Katepwa populations, and the semi-dwarfing loci of Rht-B1 and Rht-D1 varying in all three populations. Of these, QTL for WSCC, WSCA and WSCT collocated with alleles at the Ppd-D1 locus in all three populations, providing 33-38% of the genetic effect on phenotype, with the earlier flowering allele associated with increased WSCC. Dwarfing alleles at the Rht-B1 locus collocated with QTL for WSCC in the Sunco/Tasman population, and for WSCA and WSCT in the CD87/Katepwa population. In all cases, the reduced height allele was associated with greater carbohydrate accumulation.

For WSCT, QTL were detected in the Cranbrook/Halberd and Sunco/Tasman populations that collocated with the glutenin locus Glu-A1, with a further QTL in the Cranbrook/Halberd population that collocated with the Glu-B1 locus. The glutenin loci are involved in control and expression of polymeric proteins that link when flour is mixed with water to form dough. The rheological properties (resistance to stretching, and amount of stretching before breakage) of dough determine the end use of wheat
flour, which are largely determined by particular high and low molecular-weight glutenin alleles. Consequently, the marketing value of a variety is largely determined by the combination of glutenin alleles present, so these loci are significant selection targets in all Australian wheat breeding programs (Payne, 1987).

Many additional QTL identified in Rebetzke et al. (2008) were not collocated with these major loci, and some were identified across the three populations under study, suggesting that WSCC could be manipulated independently of major developmental and quality loci.
Table 2.1: Loci for water soluble carbohydrates from previous studies (*UA is unlinked A genome marker)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Population</th>
<th>Environment</th>
<th>WSC traits</th>
<th>QTL</th>
<th>Genetic variance</th>
<th>Collocating traits and major loci</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snape et al. (2007)</td>
<td>Spark/ Rialto</td>
<td>UK, moderate drought</td>
<td>WSCC per tiller</td>
<td>1B/1RS</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2B</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3D</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4B</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7A</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Yang et al. (2007)</td>
<td>Hanxuan 10/ Lumai 14</td>
<td>China, drought stress</td>
<td>WSCC per tiller at flowering</td>
<td>1A</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1D</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2D</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4A</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4B</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7B</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>WSCC per tiller at grain fill</td>
<td>4A</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>WSCC per tiller at maturity</td>
<td>6B</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6B</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>WSC accumulation</td>
<td>1A</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2A</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5A</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7B</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>WSC remobilisation</td>
<td>7A</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Location/Country</td>
<td>Environment</td>
<td>Accession</td>
<td>Yield, height, canopy temperature, NDVI, chlorophyll content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>---------------------------</td>
<td>------------------------------</td>
<td>-----------</td>
<td>-------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rebetzke et al. (2008)</td>
<td>Cranbrook/Halberd Australia, MET well-watered</td>
<td>WSCC</td>
<td>1A 2</td>
<td>Ppd-D1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2B 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2D 30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3A 30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3B 35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3D 35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sunco/Tasman</td>
<td>WSCA</td>
<td>2D 11</td>
<td>Rht-B1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4B 6-9</td>
<td>Rht-D1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4D</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McIntyre et al. (2011)</td>
<td>CD87/Katepwa</td>
<td>Australia, MET, separate analyses for drought stress and well watered environments</td>
<td>WSCC</td>
<td>Screenings, biomass</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Seri/Babax</td>
<td></td>
<td>1A 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1D 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4A 7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6B 7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(UA)* 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4A 7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4B 7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6B 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pinto et al. (2010)</td>
<td>Seri/Babax</td>
<td>Mexico, MET, separate analysis for terminal drought, heat stress and irrigated envt.</td>
<td>WSCC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1B 12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Variety</td>
<td>Region, Environment</td>
<td>Method</td>
<td>Year</td>
<td>Trait 1</td>
<td>Trait 2</td>
</tr>
<tr>
<td>---------------------</td>
<td>------------------</td>
<td>--------------------------------------------</td>
<td>--------</td>
<td>------</td>
<td>-----------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Pinto et al. (2010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yield, thousand grain weight, canopy temperature, NDVI</td>
<td>3B 11</td>
</tr>
<tr>
<td>Bennett et al.</td>
<td>RAC875/ Kukri</td>
<td>Australia, MET, moderate to severe drought</td>
<td>WSCC</td>
<td>2D</td>
<td>Grains per m², screenings, yield</td>
<td>4A 28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>WSCA</td>
<td>3A</td>
<td>Test weight, plant height</td>
<td>3A 19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6A</td>
<td>Tillers per m², thousand grain weight, grains per m², screenings</td>
<td>6A 6</td>
</tr>
<tr>
<td></td>
<td>WSC per tiller</td>
<td></td>
<td></td>
<td>3A</td>
<td>Test weight, plant height</td>
<td>3A 9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6A</td>
<td>Tillers per m², thousand grain weight, grains per m², screenings</td>
<td>6A 6</td>
</tr>
</tbody>
</table>
2.9  *The relationship between relative maturity and water-soluble carbohydrate concentration*

The accumulation and remobilisation of WSC is developmentally dependent, as the assimilate supply and demand changes over the lifecycle of the crop. This is detailed previously in section 2.4. This indicates there is potentially a strong link between relative maturity around flowering time (ZAD) and WSCC, although the nature of this relationship will depend on the timing of WSCC sampling. For example, populations with a range of flowering times sampled at the same time for WSCC risks that some genotypes will be in the carbohydrate accumulating phase and some in the remobilisation phase, and the concentrations measured may not be reflective of the actual maximum potential of the individual genotypes. Figure 2.1 reproduced from Gebbing (2003) illustrates the increase and decrease of WSC in terms of fructans and sucrose per tiller over the life cycle of a wheat crop. Studies such as Ruuska *et al.* (2006); Rebetzke *et al.* (2008) and Rebetzke *et al.* (2013) avoid much of this bias by sampling each plot in their trials as they reach 180°C past anthesis (Z65). This is not always logistically possible, and an alternative is to use a population constrained for flowering time variability, such as the approach taken by Rattey *et al.* (2009); Pinto *et al.* (2010) and McIntyre *et al.* (2011).
Figure 2.1 Reproduced from Gebbing (2003), this figure shows the rise and fall of WSC (in terms of both fructans and sucrose amount per tiller) relative to anthesis date.
2.10 The genetics of flowering time in wheat

Some key loci for flowering time have been well characterised, and are deployed in breeding programs around the world to ensure the crop life cycle best matches the growing environment for maximum grain yield (Evans, 1996). These loci are particularly important in environments such as the Australian wheat belt where drought escape is critical (Passioura, 1996; Richards, 1996). Wheat breeding in Australia has been an iterative process of continually adapting crop phenology so that the critical period of crop growth (when grain number is determined) occurs under the most favourable environmental conditions (Slafer et al., 2015). This favours the development of genotypes that flower early enough to make use of lower growing temperatures and vapour pressure deficit, and late enough to reduce the risk of frost damage (Nix, 1976; Perry & D’Antuono, 1989). This process of tailoring phenology commenced Farrer in the development and release of the cultivar Federation and continues today with the growth duration to flowering of spring wheats continuing to decrease (Perry & D’Antuono, 1989; Fischer, 2011; Sadras & Lawson, 2011). The success in this effort has been greatly facilitated by an understanding of the main photoperiod sensitivity and vernalisation genes (Pugsley, 1983). These loci have been the main means for plant breeders to alter crop phenology (Snape et al., 2001), and in Australian wheat, can account for around 45% of the genetic variation for flowering time (Eagles et al., 2010).

The major genes for vernalisation are located on chromosomes 5A (Vrn-A1), 5B (Vrn-B1) and 5D (Vrn-D1) with other minor effect vernalisation loci not directly utilised much by breeders (Yan et al., 2004). Vernalisation (meaning “to become spring-like”) occurs when the recessive alleles of Vrn loci are present, which confer a requirement for a low temperature period to avoid delays in flowering time. Dual-purpose winter
wheats typically carry these recessive alleles so that they can be planted earlier to
develop large vegetative biomass for grazing before flowering is initiated (Davidson et
al., 1990). Dominance at the 
loci removes the low temperature requirement and
genotypes become spring like (Law et al., 1976; Dubcovsky et al., 2007).

The main photoperiod response genes are on chromosomes 2A (Ppd-A1), 2B
(Ppd-B1) and 2D (Ppd-D1). The dominant alleles at these loci confer insensitivity to
changing day length and therefore promote earlier flowering, with different allele
combinations at the different loci producing a range of reproducible effects (Cane et al.,
2013). Photoperiod sensitive gene Ppd-A1 is not yet mapped and probably not as
commonly utilised as Ppd-B1 and Ppd-D1 by breeders (Slafer et al., 2015).

Aside from the Vrn and Ppd loci, flowering time is also influenced by earliness
per se genes with smaller effects and more complex interactions. The major earliness per
se genes are located on chromosome groups 2, 3 and 5 (Snape et al., 2001). Wheat
flowering time represents one of the best understood genetic traits in wheat, even if the
physiological mechanisms driving expression of these loci are not clear. Relative
maturity around flowering time has strong and repeatable genetic effects making it a
suitable trait to use as a positive control to evaluate different QTL discovery methods,
and an important trait to characterise alongside WSC given the interaction between
these traits discussed in section 2.9.

**2.11 Progress in genetic mapping and analysis methodology**

From the exploitation of allozyme information (Stuber et al., 1980) and the
beginning of linkage interval mapping (Lander & Botstein, 1989), marker-trait
associations have held much promise in the utilisation of crop diversity for genetic
improvement ( Tanksley & McCouch, 1997). Today, they have evolved to become a
valuable tool for plant breeders who can perform whole-genome association studies (GWAS) to dissect the underlying genetic architecture of plant phenotypes. Much of this progress has been dependent on advances in molecular marker technology (Ingvarsson & Street, 2011; Khan et al., 2014), and although progress is species-dependent, in general genetic polymorphism discovery has advanced rapidly based on advances in microarray technology. The rapid development in microarray features has been compared by Mockler and Ecker (2005) with the progress made in computer chip technology, which has followed Moore’s Law by increasing number of transistors per silicon chip exponentially over time.

Some of the first association studies were performed in hexaploid bread wheat (Bresegello & Sorrells, 2006; Crossa et al., 2007). However, wheat has proven to be a difficult species in which to develop genetic polymorphism information, due to the massively large polyploid genome and relatively low genetic diversity (Sukumaran & Yu, 2014). In particular, the homologous and paralogous nature of wheat loci make allocating markers to their correct genomes difficult (Cavanagh et al., 2013; Wang et al., 2014). Notwithstanding the challenges presented, some comprehensive marker resources have been developed, including the SNP assays of the previous two studies, and the Diversity Arrays Technology (DArT) marker system provided by Triticarte Pty Limited (Jaccoud et al., 2001; Wenzl et al., 2004; Akbari et al., 2006).

### 2.12 The role of linkage disequilibrium in molecular quantitative genetics

Genetic association studies, as reviewed by Flint-Garcia et al. (2003), are essentially the statistical association of linkage disequilibrium (LD) with variation in a trait. Linkage disequilibrium, more accurately called gametic phase disequilibrium (Lynch & Walsh, 1998), is the non-random association of alleles within a population
(Nordborg & Tavaré, 2002). As these authors explain, LD should not be confused with linkage, which is instead the physical association of loci on the same chromosome. While there is logically a relationship between LD and physical separation of loci on a chromosome, the correlation between the two is highly variable (Rafalski, 2010). The extent and nature of LD is determined by factors that influence the rate of LD decay and the inherent genetic variability of populations. These include mutation rates which give rise to new alleles and the rate of recombination which is largely governed by the pollination system of the species under consideration (inbreeding has fewer recombination opportunities than outbreeding).

Substantial progress has been made in association studies in maize, an outcrossing and genetically diverse crop species. Populations of elite US commercial maize lines have been shown to be more genetically diverse than the comparable interspecies variation between humans and chimpanzee (Rafalski & Morgante, 2004; Buckler et al., 2006), and LD decay is as rapid as 2,000 base pairs for diversity panels of maize inbred lines (Yu & Buckler, 2006). In contrast, wheat populations have been reported to have LD of around three times higher than comparable maize populations (Breseghello & Sorrells, 2006). Reasons for this include self-pollination, and the relatively recent (in evolutionary terms) hybridisation of goatgrass (Aegilops tauschii) with probably tetraploid durum wheat (Triticum durum) to include the D genome and become modern hexaploid bread wheat (Triticum aestivum) (Harlan, 1992). Wheat also contains a very large genome representing three separate genomes (A, B and D) (Morrell et al., 2012) so it is more complex and marker identification more difficult than comparable crop species (Sukumaran & Yu, 2014). Many workers report LD decay is around two or three times slower in the D genome compared to the A and B genomes,
with the D genome having a greater proportion of rare alleles (Akhunov et al., 2010; Wang et al., 2014).

### 2.13 Principles of genome-wide association studies

Association mapping is an open system design that exploits recombination events in the whole evolutionary history of the population (Myles et al., 2009). Association mapping is also called LD mapping, because it relies on the statistical association of phenotype between genotypes based on LD decay between markers. Essentially, association mapping compares the distribution of phenotypes amongst individuals carrying a particular locus and those that don’t, and then computing a probability of those differences arising randomly. Association mapping can be utilised across either a whole genome, or for a set of candidate loci. Apart from human genetics, candidate gene studies are rare, as they require detailed prior understanding of the underlying metabolic pathways in order to select the candidate loci (Rafalski, 2010). Moreover, as both Myles et al. (2009) and Ingvarsson and Street (2011) observe, identification of loci for candidate gene studies are essentially guesswork, and a genome-wide approach is the dominant method for plant breeding studies.

A GWAS was first applied to plant systems by Thornsberry et al. (2001) and this approach is now widely used in plant breeding. The use of GWAS differs from the more conventional biparental QTL mapping approach by utilising the different recombination events that occur in a population over its evolutionary history (Myles et al., 2009). It also allows the use of relevant plant breeding populations with a larger selection of the genetic diversity available for improvement of a trait and potentially multiple QTL (Mackay & Powell, 2007; Zhu et al., 2008; Yu et al., 2011). Together, GWAS may go some way to overcoming the limitations of biparental studies, where
identified QTL have rarely been deployed in plant breeding programs (Bernardo, 2008).

**2.14 Factors that affect the outcome of genome-wide association studies used for plant breeding**

The use of genetically diverse germplasm sets as populations has a number of implications that affect the results of a GWAS, so that it is difficult to apply a ‘black box’ approach to analysis. A careful understanding of these factors will drive interpretation of results (Myles et al., 2009; Rafalski, 2010). These factors are based off the same influences that affect the degree of LD in a given population, reviewed above in section 2.12. The previous section (2.13) outlines the main components of a GWAS, while this section discusses some of the factors with GWAS that can influence the outcome of a study. Clearly the nature of inheritance of the trait under study is important, the size of the genetic variance component and the degree of G × E interactions will determine the strength of genetic signal relative to total phenotype that can be associated with markers (Myles et al., 2009). Describing and understanding repeatable G × E is important, as Heslot et al. (2014) articulate, G × E can come from many different phenomena not just a few major stresses or geographic portions and may not be easily integrated as specific environmental covariates into a model.

The range in genetic diversity will determine the genetic variance component of the GWAS model, while allelic frequency across genotypes selected will play a role in detection. In breeding populations comprising advanced breeding lines diversity is often restricted by the effects of selection, so important loci may be fixed and rare loci difficult to detect within GWAS. On the other hand, a very diverse germplasm set is more likely to contain loci of large effect that mask the genetic contribution of other
and particularly smaller effect loci. In turn, population size can influence results as well (Wang *et al.*, 2012).

The genotypes selected to represent the population limit the genetic scope in an association study, but the markers used provide the degree of genetic resolution. Greater genome coverage of markers will take advantage of higher LD with causal variants, so the proportion of genetic variance explained by the markers should theoretically increase. It will also increase the stringency of multiple testing, as the false discovery rate (FDR) p-value adjustments commonly used are assumed to be independent. Obviously this is not the case, as biologically significant loci are often part of biochemical pathway networks and are likely to be linked (Alm & Arkin, 2003; Sorkheh *et al.*, 2008). The type of molecular marker employed has also shown to provide variation in GWAS results. For example, Remington *et al.* (2001) observed LD decay was different in SNP datasets than in single-sequence repeat (SSR) marker datasets in maize, and Matthies *et al.* (2012) observed differences between the GWAS results when using SSR markers and DArT markers in barley. Further, genotyping errors can significantly influence the false positive rate (Attia *et al.*, 2009).

The diversity of statistical approaches to GWAS indicates that plant breeders have a range of ‘black box’ analysis approaches available to them. These undoubtedly have a bearing on the nature and sometimes quality of GWAS results, although a discussion of the underlying algorithms of the various approaches is beyond the scope of this review. Briefly, most approaches use a two-step analysis, where first the variation among phenotypes is modelled, then those modelled trait values are used to perform the genetic-based association analysis. An alternative is to use an approach where both phenotype and marker-trait associations are modelled simultaneously so that experimental variance components for the phenotype are carried through to the
genetic model. Wang et al. (2012) reports that the two step modelling procedures are more effective than one step models, although this is in contrast to Verbyla et al. (2007), who discuss the advantages of including all of the phenotypic variance components within the marker-trait association model.

With these factors largely determining the outcome of an association study, plant breeders need to be cautious when using GWAS to gain insight into the inheritance of a trait. As espoused for human GWAS by Attia et al. (2009) and Oldmeadow et al. (2011), and for plant studies by Ingvarsson and Street (2011), multiple studies identifying associations are perhaps the most reliable way to elucidate the role of particular loci in the inheritance of a trait and validate GWAS associations.

2.15 Identifying causal associations using genome-wide association studies

A knowledge of population structure in the germplasm set under study conditions perhaps the most important question over any GWAS results – that of the degree of both Type II (false positive) and Type I (false negative) errors. Oftentimes the real biological significance of loci is poorly estimated by the statistical significance in the model used, and loci unrelated to the trait under study can show association due to unequal genetic relationships between individuals (Breseghello & Sorrells, 2006; Myles et al., 2009; Yang et al., 2010). Population structure is a consequence of non-random mating, arising from inbreeding or selection (Kennedy et al., 1992). This is commonplace in plant breeding populations, created as plant breeders and their selection environment trigger sweeps of selection fixing key loci (Smith & Haigh, 1974), and/or changing the allele frequency at many loci (Pritchard & Di Rienzo, 2010). Notably, population structure is also theoretically a confounder of the results of QTL studies from biparental populations, as some degree of selection pressure is often
inadvertently applied during population development although the question of uneven relatedness in such populations is rarely addressed (Sukumaran & Yu, 2014).

A number of strategies can be employed to reduce the probability of a false positive result. Firstly, a relationship (or ‘kinship’) matrix is included in the model, computed from marker or pedigree information. The commonly used inbreeding by descent (IBD) relationship matrix reflects the genetic structure of a population, where all lines are related to some degree (Habier et al., 2007b; Myles et al., 2009). The use of a relationship matrix is highlighted in the studies of Yu and Buckler (2006) and Larsson et al. (2013). Here, significant associations in the former study, in which the model used did not include a relationship matrix, disappear when a relationship matrix is used (in the latter study). However, it is important to note that the false negative rate will be greater with the inclusion of a relationship matrix, as significant and therefore important loci are more likely to be missed (Rafalski, 2010; Ingvarsson & Street, 2011). Other methods are also used in conjunction with the relationship matrix. Brachi et al. (2011) advocated ways of population restructuring that can be applied to break up population structure, depending on the G × E interactions of the trait under study. This can be taken further with the structured populations described by Holland (2007) such as the interconnected, shared parent structures of nested association mapping populations (Yu et al., 2008) and multi-parent advanced intercross populations (Cavanagh et al., 2008).

In most contemporary GWAS studies, the statistical probability obtained for significant loci is adjusted to account for false positives. One of the most common ways of doing so is to apply the Benjamini and Hochberg (1995) FDR procedure, where p-values are ranked, then each p-value is adjusted by its position down the list. Such adjustment also increases the false negative rate (Rafalski, 2010), and is widely
regarded as being overly stringent (Edae et al., 2014). Other alternatives to the
Benjamini and Hochberg (1995) procedure produce much the same results (Storey &
Tibshirani, 2003; Zila et al., 2013). An alternative to this approach is the adoption of a
consensus p-value, such as the p < 5x10⁻⁸ level used in human genetics (Attia et al.,
2009). However no consensus has emerged from plant breeding GWAS, and applying a
FDR p-value adjustment is at least indicative of p-values associated with biological
significance.

2.16 ‘Missing’ heritability and the relationship between
heritability in quantitative genetics and genome-wide
association studies

Discussion in the GWAS literature around the issue of ‘missing’ heritability is
illustrative of the role GWAS can have in plant breeding programs. It is widely
acknowledged that GWAS regardless of the species under study generally only
identify loci that explain a fraction of the calculable heritability for a trait (Holland,
2007; Manolio et al., 2009; Eichler et al., 2010). A frequent example here is illustrated
with the inheritance of human height. Human height has a heritability of around \( H^2 = 0.8 \), but studies prior to Yang et al. (2010) identified loci that accounted for around 10% of trait variation (Lango Allen et al., 2010). Yang et al. (2010) showed that some of the
discrepancy arises from statistical design and analysis considerations. The missing
heritability was in large part to do with ascertainment bias having filtered out many of
the low frequency markers, as GWAS rarely detects the rare alleles that contribute to
total genetic variation. The methodology of GWAS was originally designed to focus on
the higher end of the allele frequency and allele effect spectrum (Manolio et al., 2009),
which is in contrast with the objectives of plant breeders, who often seek to target rare
alleles and where possible rare combinations of alleles in plant improvement.
Population structure can also obscure causative loci, given these loci will be in LD with many other loci across the genome (Brachi et al., 2011), although as Korte and Farlow (2013) explain, non-causative loci can be better predictors of phenotype in some cases. Other sources of hidden heritability include the non-additive genetic variance (including dominance, epistatic and other genetic effects) or genotype × genotype epistatic interactions, and G × E interactions.

Care must be taken not to confuse heritability with the total genetic effects. As Eichler et al. (2010) explained, measures of heritability reflect the proportion of total variance. In turn, hidden environmental covariates that interact with genetic effects will move genetic variance to the denominator and reduce heritability estimates. Effectively this means that where environments are not well sampled, heritability is at a minimum and represents the smallest estimate of the genetic component of a phenotype. Thus Eichler et al. (2010) suggest that the focus of GWAS should be on genotype effect associations with phenotype rather than just genotype-phenotype associations, as that would allow characterisation of environmental effects on alleles. The QTL information provided by biparental studies is notably environment specific (Bernardo, 2008) and population specific (Holland, 2007) with these same caveats applying to GWAS. Just as alleles have a potentially different effect in different environments, they can also have different effects due to dominance and epistasis in different genetic backgrounds (Carlberg & Haley, 2004; Eichten & Borevitz, 2013). In theory, GWAS can detect epistatic effects, but in practice most models lack the power to do so compared to biparental studies (Rafalski, 2010). Additionally, methods that attempt to model epistasis in association studies are statistically difficult and not always biologically sensible (Holland, 2007). That is, they are largely statistical. Korte and Farlow (2013) outlined strategies for dealing with non-additive genetic effects,
noting that an allele in a different genetic background is statistically equivalent to a
different environment, and epistatic effects can subsequently be modelled like $G \times E$
interaction effects. So far, GWAS has been most useful for traits that are more
Mendelian rather than quantitative, and instances where such traits controlled by few
loci are difficult to phenotype, the loci identified are particularly useful to plant
breeders.

### 2.17 Genomic selection principles and background

Genomic selection refers to the simultaneous estimation of marker effects across
the genome, rather than the marker by marker approach of MAS (Meuwissen et al.,
2001; Heffner et al., 2009; Lorenz et al., 2011). Genomic selection uses the number of
markers shared between line pairs to model a genomic estimated breeding value
(GEBV), from the genomic best linear unbiased predictions, or G-BLUPs, rather than
patterns of markers across the genome (Habier et al., 2013). The nomenclature used
hereafter will define the additive genetic effects selection index as GEBVs, which
together with the non-additive selection index comprise the total selection index in a
G-BLUPs model. As reviewed by Hill (2012) and Nakaya and Isobe (2012), crop
genomic selection uses marker measures of realised relatedness from whole-genome
marker profiles to predict phenotypes of progeny, much like animal models (and less
commonly, crop models) that are designed to estimate breeding values utilising
pedigree information. The rationale of genomic selection is presented in Figure 2.2,
which illustrates the benefits of selection on phenotypic variance due to relatedness, as
selection is moved away from the basis of individual lines to the basis of allelic
combinations (Heffner et al., 2011b). A training population is composed of phenotyped
homozygous lines from the breeding program itself or from mapping populations
created for the purpose (Heffner et al., 2011a), with the corresponding marker realised-relationship matrix generated from whole genome marker profiles (Habier et al., 2007a). The statistical model is used to estimate the GEBV which is the phenotypic variance predicted from the relatedness between lines, as measured by shared markers from a whole-genome marker profile. Whole-genome profiles of lines without phenotypes are included in the training population model, and the relatedness of genotype-only lines with phenotyped lines is used to model GEBVs. As discussed in Heffner et al. (2010), GEBVs can be used to select lines and families of lines much earlier in the breeding process (illustrated also in Figure 2.2). In an inbreeding crop like wheat, germplasm needs to undergo several generations of inbreeding before lines are sufficiently fixed so they can be phenotypically evaluated as parental material for the next cycle of breeding, whereas for genomic selection, lines can be genotyped and selected as parents as early as the F2 and F3 stages. In this way, the rate of genetic gain per unit of time can be increased (sometimes significantly), even though prediction accuracies are somewhat lower than accuracies with phenotypic selection (Desta & Ortiz, 2014).

Clearly, the training population is critical to the accuracy and usefulness of the GEBVs, as genomic selection accuracy relies on the linkage disequilibrium (LD) between GEBV and trait being similar in both the training population and the prediction population. It should in theory be large enough to cover the scope of the genetic variation being tested in the breeding program (Hayes et al., 2009a; Daetwyler et al., 2010; Technow et al., 2013) but without strong population structure (Windhausen et al., 2012), otherwise the model will identify spurious genetic relationships thereby limiting its predictive ability. Empirical studies, such as that of Storlie and Charmet (2013), have shown a linear relationship between accuracy and shared alleles between
the training population and the prediction population. This is potentially a challenge for genomic selection in dynamic breeding programs, given selection pressure (intentional or otherwise) occurs throughout the breeding process, and lines that progress to the relatively advanced step (in genomic terms) of fixed-line evaluation are unlikely to represent the full spectrum of diversity present in the early stages of a breeding program. Heffner et al. (2011a) suggested a way of countering this could be the use of biparental crosses or other mapping populations as part of the training population model. Alternatively a higher proportion of genotypes with poorer prospects normally discarded in the breeding process could be retained for phenotyping and inclusion in the training population. While this may not appear an ideal use of breeding program resources, the resulting genetic gains from genomic selection may outweigh the cost (Lorenz et al., 2011).
Figure 2.2 Schematic of a inbreeding species breeding program with genomic selection adapted from Smith (2014). Standard phenotypic selection means that the breeding cycle (between crosses) is at least four years, as lines need to reach at least the F5 generation to be sufficiently genetically fixed (homozygous) for a stable phenotype to be obtained. Genomic selection uses the phenotypes measured in the line evaluation phases of the breeding program (called the training population) in a genomic prediction model to predict genotypes which are more likely to have the phenotype of interest in the early generations of a breeding program. Identified superior genotypes can be utilized in crossing, thus reducing the breeding cycle time from four years to two years.
2.18 Comparing genomic selection to marker-assisted selection

Marker-assisted selection differs from genomic selection fundamentally in that few loci are utilised in MAS, while whole-genome profiles are typically used for genomic selection although Hamblin et al. (2011) outlines methods for genomic selection utilising few SNPs of large effect. The overall difference is that genomic selection does not use markers individually, instead using measures of relatedness computed from whole-genome profiles. Some studies contend that the predicted usefulness of MAS for plant breeding is over-optimistic, as demonstrated by the lack of progress in the field (Jannink et al., 2010; Hill, 2012; Desta & Ortiz, 2014). However, a more consensus view is that MAS should be used where the genetic architecture of a trait is well understood and the genotypic variance is explained by a few major loci (Heffner et al., 2009; Ingvarsson & Street, 2011; Sukumaran & Yu, 2014). As a corollary, molecular marker technologies should not be automatically seen as the most appropriate methods for genetic improvement. As Holland (2004) reported, when heritability is high, and phenotyping is relatively straightforward and inexpensive, phenotypic selection may be preferable to marker-based approaches.

The challenge with MAS is that identifying significant loci for polygenic traits is difficult given that by definition they are controlled by many, sometimes hundreds of genes (Holland, 2007). In that case, genomic selection may be more appropriate, as marker effects across the whole genome are approximated in genomic selection models (Makowsky et al., 2011). Further, marker-QTL LD varies markedly across a genome and from population to population (Rafalski, 2010), so GWAS can sometimes miss alleles for highly heritable traits, because of fast decay of LD even with many markers. For example, Romay et al. (2013) reported results in GWAS for flowering time in maize,
where only one marker out of more than 600,000 tested was associated with the largest QTL identified for flowering time reported in previous studies.

2.19 Models and methods used to generate genomic predictions

There are a range of different statistical models developed for estimating marker effects across the whole genome. Numerous reviews of the different statistical models used in genomic selection studies have been published (for example, Jannink et al., 2010; Lorenz et al., 2011; Nakaya & Isobe, 2012; and Desta & Ortiz, 2014), and here only the main approaches are outlined. Perhaps the most common method, ridge-regression BLUP models, assume all of the markers in the model have equal variances, and is regarded as being more effective at capturing genetic relationships because of use of all markers. This approach is considered particularly suitable for traits where inheritance is controlled by many loci of small genetic effect. A second group of methodologies, Bayes A, Bayes B and Bayes Cπ, are quite different to ridge-regression as the genetic effects are not assumed to be equal, and assume that the genetic variance is dominated by few loci. Indeed, in the case of Bayes Cπ the proportion of markers used in the model is estimated from the data. Bayesian methods have been advocated for oligogenic traits where a large proportion of genetic variance is explained by a few QTL.

Mixed linear modelling approaches, including factor analytic (FA) models have also been used (Burgueño et al., 2012; Guo et al., 2013) for genomic prediction in multi-environment trial datasets. An example of their application was illustrated in Oakey et al. (2007), using a realised relationship matrix based on markers rather than pedigrees (Habier et al., 2007a).
It is generally advocated that the more markers used to compute the genomic predictions the better (Jonas & de Koning, 2013; Desta & Ortiz, 2014). However, Solberg et al. (2008) have shown in animal models and Makowsky et al. (2011) in human models that at some point, additional markers make little difference to the accuracy of predictions. The optimal number is likely to be population-specific and depend on the LD between the markers used and the trait of interest in both the training population and the prediction population.

### 2.20 Accuracy of genomic selection and genetic gain

A focus of recent genomic selection investigations has been in determining the accuracy (or effectiveness) of genomic predictions, and much recent work in crop breeding has followed developments in mammalian genetics, especially dairy cattle. There are typically three different validation approaches employed (illustrated in Figure 2.3). Progeny validation involves using the offspring of parents in the training population model as the prediction set. This is the most realistic test of accuracy as predictions are supposed to be based on the genetic selection value of a line, and progeny prediction takes into account changes in allele frequency due to recombination and selection that occurs over generations (Sallam et al., 2015). However, progeny validation has been rarely reported so far in an experimental setting due to the time and resources required to test predictions across generations. Some initial results of progeny validation have been reported from maize (Cooper et al., 2014) sugar beet (Hofheinz et al., 2012) and barley (Sallam et al., 2015).

Inter-set validation refers to the method of using separate training and prediction datasets, so that predictions modelled in one population are tested in a separate prediction population. This approach was explored by Hayes et al. (2009a)
using predictions for Holstein and Jersey populations of dairy cattle. Predictions for one breed were tested in the other, and compared to combined predictions from both breeds. The authors found that predictions between the two distinct breeds were poor. However, if the training set included both breeds the predictions improved markedly. Windhausen et al. (2012) tested inter-set predictions with CIMMYT maize populations and found very little (and even sometimes no) predictive ability where the training and prediction datasets were from genetically different populations. These results provide further weight to the requirement that training-populations need to be optimised for the trait architecture and population structure specific to the prediction population (Isidro et al., 2015).

One of the most widely used methods reported in the literature for assessing predictive accuracy is the cross-validation method, where the dataset used for training is also used to assess prediction (Luan et al., 2009; Heffner et al., 2010; Lorenz et al., 2011). While accuracy values from this method may be inflated, as in practice the predictions would not be utilised in the same generation and LD patterns would be somewhat different (Jonas & de Koning, 2013), experimental limitations dictate that this method is often the only one that can be applied to the dataset under study. Typically, the ‘fold’ method is used, where the genotypes with phenotypes in the dataset are split into a number of equal-sized groups. Each group in turn has the phenotype information removed from the model, leaving the phenotypes of genotypes in the other groups for obtaining genomic predictions. This process is repeated for each ‘fold’ or group, and the GEBVs are combined to form a full set of genotype only predictions to compare against a full set of genotype and phenotype predictions. Lin et al. (2014) reviewed the range of accuracies obtained using this method. Sallam et al. (2015) explains that cross validations gives the most optimistic measure of accuracy.
because they are based on the same genotype and environment covariance-based datasets. However at this early stage of genomic selection implementation it is a useful way of generating a measure of the usefulness of genomic predictions.

Based on the Falconer and Mackay (1996) response to selection ($R$) equation,

$$R = i \cdot r \cdot \sigma_A,$$

response to genomic selection can be estimated by:

$$\Delta G = \frac{i \cdot r \cdot \sigma_A}{\Delta t}$$

Equation 2.1

where $\Delta G$ is the annual selection gain, $\sigma_A$ is the additive genetic standard deviation of the trait, $r$ is the GEBV divided by the genomic prediction accuracy, to give selection accuracy, and $i$ is the selection intensity applied (Desta & Ortiz, 2014). An important addition is the time taken to achieve a breeding cycle ($\Delta t$, the time between crossing and the identification of superior progeny for the next round of crossing), as this is a key difference between genomic and phenotypic selection. Phenotypic selection for new parents for crossing (as opposed to early generation phenotypic selection for germplasm trait enrichment which is often performed on early generation material) cannot occur before fixed lines are generated. In the case of a wheat breeding program, this would commonly occur at approximately four years after crossing at the earliest and often later (Figure 2.2). However, genomic selection for new parents can be applied as soon as segregation begins in the F2- F3 generation, which in conjunction with rapid generation advance techniques could be as early as only one year (Heffner et al., 2010) although more typically occurs two years after crossing (Longin et al., 2015). The reduction in breeding cycle time is an indicator of how accurate genomic prediction needs to be to achieve genetic gains. Equation 2.1 shows that if breeding
cycle times can be reduced from four years to one year, relative accuracies need only be
greater than 0.25 to achieve greater genetic gain than phenotypic selection.

Another aspect considered by plant breeders when evaluating selection
efficiency is the complexity and expense required to obtain accurate phenotypes.
Genomic selection can significantly reduce the resources needed for difficult and
expensive phenotyping (Lin et al., 2014). Comprehensive phenotyping is still an
important part of the breeding process when genomic selection is applied, as it is
essential for the usefulness of the training population model. However, early
generation nursery screens can be replaced with genomic screens (Sallam et al., 2015),
and expensive field trials can be augmented by genetic material ‘enriched’ through
genomic selection (Heffner et al., 2009).

The combination of acceptable prediction accuracy and reduction in breeding
cycle time means that substantial genetic gain through the use of markers can be
achieved without having complete understanding of the genetic architecture of a
complex trait (Cooper et al., 2014), somewhat analogous to the ‘mapping-as-you-go’
methods of Podlich et al. (2004). Genomic prediction models need to be retrained over
time as after fixing a trait in the breeding population by selection, it no longer has any
predictive ability for the purpose of further selection.
Figure 2.3 Validation approaches in genomic selection (reproduced from Sallam et al., 2015). Subset, or cross-validation uses the same dataset for both training and prediction. Interset validation refers to the method of testing training population predictions in a separate prediction population. Progeny validation refers to a prediction population composed of progeny from parents present in the training population.
2.21 Genomic selection and genotype × environment interactions

In many of the studies on genomic selection thus far, G × E has been regarded as an impediment to genomic selection predictive ability, rather than an important consideration in the breeding process (Resende et al., 2012; Storlie & Charmet, 2013; Desta & Ortiz, 2014). Jonas and de Koning (2013) suggested that this inclination might in part be due to the origins of genomic selection methods in dairy cattle breeding, where G × E plays a smaller role in prediction accuracy as dairy herds are often kept in facilities that allow for some control of environmental influences. As Heslot et al. (2014) observed, genomic prediction studies so far have tended to focus on the computation of GEBVs presumed to be useful across all environments. Some studies do not address, or confound the effects of G × E on genomic prediction models (Asoro et al., 2011; Storlie & Charmet, 2013; Longin et al., 2015), and some deliberately attempt to minimise any environmental interactions, (for example, Heffner et al., 2010; and Sallam et al., 2015). Crossa et al. (2010) showed that marker effects and prediction accuracy estimated through genomic selection on a wheat breeding population differed between environments demonstrating that G × E can even be characterised through the process of genomic prediction, although in that study separate models were fitted for each trait-environment combination so it was not a true MET. Ly et al. (2013) discuss the impact of G × E in genomic prediction in a cassava breeding study and suggested that exploitation of G × E would be more effective than attempting to remove it from the prediction model, noting that G × E is not independent to the genetic variance components of the model. Their study showed that G × E from location effects was a strong confounding factor similar to the results of Resende et al. (2012) who found that
genomic selection models had poor predictive ability across sites in a eucalyptus dataset.

Together, these studies show that G × E interactions are a driver of prediction thereby suggesting that part of the differences observed empirically in the predictive ability of GEBVs (for example the table in Lin et al. (2014) of different crops), where trait heritability is similar could be the presence of G × E contributing to genetic architecture.

With reference to the Falconer and Mackay (1996) response to selection equation and Equation 2.1, an increase in G × E relative to genetic variance will reduce relative prediction accuracy (Asoro et al., 2011). However this does not account for differing environmental targets in a multi environment trial study, which need to be understood in order to maximise selection efficiency (Wade et al., 1996; Chenu et al., 2011; Cooper et al., 2014).

The question of G × E interactions was first considered by Burgueño et al. (2012) who implemented FA models to demonstrate an increase in accuracy. Similarly, others (including Heslot et al., 2013; and Heslot et al., 2014) adopted an alternative approach through the inclusion of environmental covariates into genomic selection models by seeking to identify the environmental factors responsible for G × E and determining genotype sensitivity to G × E. This is a complex undertaking, as the full scope of environmental variation is challenging to characterise into actual causes, and in the latter study the authors investigated 37 different environmental covariates (with only eight covariates significant in their model). While G × E can be modelled in this way, it is difficult to know which environmental covariate is significant beforehand, and the information required may not always be easy to acquire for a breeding program. Ly et al. (2013) investigated G × E with a different approach, focusing on
location rather than year-based interactions with the intention of identifying similar environments for selection as well as outlier environments. However, their study did not consider seasonal effects as this was deemed to be outside the ability to predict in the future. Probe genotypes are one way of characterising target populations of environments (Wade et al., 1999; Mathews et al., 2011), alternatively genetic correlations (variety rankings) between experiments can be used (Cullis et al., 2010).

2.22 Discussion of the literature reviewed

The above review of literature demonstrates the challenges involved in the genetic improvement of wheat for drought avoidance. However improving wheat performance under terminal drought conditions is important for the Australian wheat industry in particular in south-eastern Australia. It will not only improve the productivity of wheat varieties, but also the stability of yield performance in a variable environment. There are a number of physiological traits under broad consideration that may contribute to drought avoidance potential, and one of the most promising of these traits is the ability to store excess carbohydrate in stems as a reserve, so that assimilate is available to fill grain in the event that photosynthesis is limited by water deficit conditions at the end of the growing season.

Accumulation of WSC can be measured in multiple ways, most usually via the concentration of fructans in stems in the period after anthesis and before grain filling. This WSC is often measured in terms of concentration of WSC in stems, however an important measure derived from both WSCC and biomass at anthesis is the total WSC per m², or WSCA. The advantages of WSCA is this measure is directly relatable to grain yield (per unit area). Both WSCC and WSCA are widely used in wheat WSC studies. The expression of both WSCC and WSCA present as an integration of the
expression of other traits, notably those affecting the supply of assimilate and the sinks for assimilate, as WSC accumulation is essentially expressed as an imbalance between assimilate supply and demand. As an alternative sink, tillering will draw assimilate from WSC reserves or divert them from being accumulated as WSC. An increase in the availability of N is seen in many studies to result in a decrease in WSC levels, as N is often a limitation to crop growth. With extra N, plants have the means to invest in extra biomass and will use stored WSC in the process of doing so. Grain filling at the end of the crop lifecycle presents as a strong sink when plants are not typically producing further biomass into non-reproductive structures. As such, most reports concur that any WSC present is mobilised to the developing grains.

Lower grain number is shown to be related to higher WSC, although this may be to do with the relationship between grain number and tillering, and tillering and WSC, rather than a direct link between the two traits. Flowering time, or development stage when the crop is sampled for WSC will have a major effect. Genotypes growing at the same location in the same year but flowering at different times will essentially experience different environmental conditions after anthesis. This is generally the stage most favourable to WSC accumulation. Care needs to be taken when sampling for WSC with genotypes that vary for flowering time and maturity, so that these factors can be considered in any interpretation of WSC measurements.

The supply of assimilate, and therefore any excess to accumulate as WSC, depends on the capacity of the photosynthetic machinery of the plant. This has been shown to be affected by shading, or environmental conditions that limit transpiration and photosynthesis. Clearly environmental effects are a large driver of assimilate supply, and this will drive expression of the trait. Reports of WSC physiology have been conducted in a range of environments globally, broadly grouped into well-
watered and water-deficit environments. However there is a paucity of research around G × E effects on WSC accumulation, including the repeatability of WSC phenotypes across experiments and different environmental types. Research on G × E effects are important for selection and genetic improvement of this trait as reliability of the TPE appears to be key in making genetic gain for WSC accumulation. This is supported by evidence from the WA and CIMMYT breeding programs which have progressively increased WSC in varieties over time with selection for performance in environments with consistent terminal drought. The interaction of G × E is of additional interest in south-eastern Australia, where the TPE is inconsistent from season to season and there have been no trends by variety year of release for WSC accumulation.

Investigations have shown that substantial genetic variation exists for WSCC and WSCA, and that genotype rankings in physiology studies are well maintained, even persisting when sampling is conducted at development stages outside the main accumulation phase after anthesis. A number of QTL studies have extended knowledge of the genetic control of WSC, from experiments conducted under both well-watered and water deficit conditions. The QTL reportedly associated with WSC are located across the wheat genome, with few loci repeatedly identified. Because of the intrinsic link between the plant development cycle and the expression of WSC, major flowering time loci are consistently identified as explaining large proportions of the phenotypic variation for WSC traits, although loci independent of flowering time are also identified. The nature of QTL identified, commonly many QTL of small effects, provides evidence for the quantitative nature of WSC accumulation and remobilisation inheritance.
Recently developed methods for marker-trait association provide new techniques for plant breeders to understand the genetic architecture of complex traits. This is made possible by a combination of availability of whole genome profiles of markers and the development of statistical methods to utilise them. GWAS provides the means to use populations of breeding material from a breeding program rather than specifically designed mapping populations, allowing the exploitation of LD generated over many generations to identify marker-trait associations. Key to applying GWAS methods is an understanding of the strengths and limitations of the technique, and an appreciation of the difference between causal (biological) marker-trait associations, and statistical associations. The nature of GWAS means results are likely to be influenced by highly variable LD patterns and marker coverage. Previously, QTL interval mapping has been used to analyse genetic control of WSC, however GWAS in a diverse population of breeding material has not yet been tried.

Another technology developed and adapted from animal breeding methods for use in the genetic improvement of crops is genomic selection. Here whole genome profiles are incorporated into the phenotypic prediction model so that breeding values can be estimated from the full marker effects across the whole genome. Genomic prediction and selection has the potential to be applied to polygenic traits where genetic control is defined by many QTL of small effect, or for traits where genetic background is particularly important. The application of genomic selection to plant breeding is in its infancy, so little is known about the usefulness of this technique for the improvement of complex physiological traits in wheat.

Therefore in subsequent chapters of this thesis, the selection potential of WSCC and WSCA in experiments reflecting selection environments in south-eastern Australia are described along with the extent of G × E interactions for these traits. GWAS was
applied to WSC in an attempt to reveal some of the genetic architecture of the trait.

Then, genomic selection methods were explored to examine the potential of genomic selection for improvement of WSC.
3. Overall Aims and Objectives of this Thesis

The physiology of WSC has been investigated in a number of studies, although much remains to be done to understand the complex interactions between sources and sinks of assimilate. The evidence available suggests that WSC accumulation is expressed when there are no other strong competing sinks available, and that accumulated WSC is remobilised for grain filling. Little has been done to investigate the potential to select for high WSC accumulating genotypes in variety development. To facilitate the selection of WSC in breeding programs targeting TPE in the south-eastern Australian wheat belt, this study aims to investigate the potential for selection of WSC and the nature of G × E interactions in the TPE. In particular, this study will use a population of elite breeding material representative of germplasm pools in Australian wheat breeding programs to maximise the relevance of results to varietal improvement.

Plant breeders have more breeding technologies available to them than ever before, in particular the methods for applying molecular markers to plant breeding continue to evolve. This study will apply one of the latest iteration of marker association methods, GWAS, to identify marker-trait associations for WSC that could be utilised for marker-assisted selection. This study also investigated the application of genomic selection to WSC improvement as an alternative means to utilise markers for complex trait selection, to provide an indication of the usefulness of this technology to breeding programs. By a comparison of the methods employed, this thesis will provide insight for plant breeders into the best approaches to select for WSC accumulation in south-eastern Australia and beyond.
4. Water-soluble carbohydrate genetic × environment interactions and selection potential

4.1 Introduction

The potential to increase the genetic capacity for WSC accumulation is an opportunity to improve the drought avoidance ability of Australian wheat varieties, particularly for south-eastern Australia where terminal drought is a significant constraint to wheat production. Modelling by Asseng and van Herwaarden (2003) showed that increasing WSCC by 20% in the wheat varieties used in southern Australia wheat growing regions could increase grain yield by 12% in water limited seasons.

To date, there is little known about genetic variability in large populations, or parental material used in Australian wheat breeding programs. Characterising this genetic variability is a necessary step if wheat breeders are to select for WSC accumulation based on variation for WSC traits already present in breeding populations. An important aspect of characterising genetic variability is determining how performance changes between TPE, in particular, how expression of the trait changes between well-watered and terminal water deficit environments. WSC can be measured in different ways, both in terms of concentration of WSC in stem tissue (WSCC), and total amount of WSC (WSCA). Selection outcomes may differ depending which measure is used, and this has not been investigated previously in breeding material (Rebetzke et al., 2008). The reported interactions between WSC and other traits show that selection for increased WSC could alter plant architecture and phenology. Characterising the potential impacts of WSC on crop growth and quality traits will better inform wheat breeders as to the approach they may take to selection of this trait.
and the indirect consequences for other traits. The concept of indirect selection for WSC has been raised, and this will be investigated as well.

4.2 Aims and objectives

Following van Herwaarden and Richards (2002), Ruuska et al. (2006), Rebetzke et al. (2008) and then Rebetzke et al. (2009), this study aimed to provide a broad assessment of WSC among wheat breeding lines and released varieties, and across drought stress and well-watered environmental conditions. The purpose of this study was to assess the extent of genotypic variation in describing the potential for genetic improvement of WSC in south-eastern Australian wheat breeding programs. The potential increase in the genetic capacity for WSC accumulation and remobilisation may lead to improvement in the drought avoidance ability of Australian and particularly south-east Australian wheat varieties. The physiology of WSC is not well understood, although it is known to be complex, with relationships between WSC and a number of other important traits documented. This study will also investigate some of those relationships, to determine how they might impact breeding strategies for WSC. An exploration of G × E interactions encountered will help interpret the response of expression of WSC in different environments, which is a key part of determining an efficient means of selection for WSC traits. To this end, experiments will be conducted using sites that contrast in water availability (well-watered and water deficit) at the end of the growing season.

4.3 Materials and methods

4.3.1 Genotypes

A set of 365 genotypes used in this study were selected from a NSW Department of Primary Industries multi-site, multi-year irrigated winter cereals
evaluation trial described briefly in Mathews et al. (2011). The experiment included 1,314 genotypes overall, provided by eight Australian wheat breeding companies and a small number of genotypes (some synthetically derived) sourced from CIMMYT via the International Adaptation Trials project (Mathews et al., 2007). The genotypes included both elite breeding lines and contemporary commercial varieties representing a range of grain quality and maturity types. Also included were a small number of synthetically derived lines sourced from the Victorian Institute for Dryland Agriculture (Department of Environment and Primary Industries, Victoria). Experiments for the broader irrigated winter cereals evaluation were grown at eight locations across southern NSW, however only the experiments grown at Yanco and Coleambally were selected and sampled for this study. These two sites were used to plant experiments with contrasting irrigation and rainfed treatments to investigate the effects of water availability on \( G \times E \) interaction and trait expression, and also these sites were located in close proximity to laboratories and dehydrators needed for accurate determination of WSC.

As WSC accumulation varies according to plant development (Ehdaie et al., 2008), this study aimed to assess genotypes as close as practicable to a common anthesis date given individual sampling of genotypes in the experiments was not possible. In 2009, a total of 319 breeding lines and varieties out of the 990 grown in the experiment were selected based on common Zadoks’ development score in the Yanco irrigated site, which was taken on 10 September 2009 (Figure 4.1). Lines selected were between \( Z49 \) (early head emergence) and \( Z56 \) (60% heading) which corresponds to a range of approximately 3-5 days difference in anthesis date in southern NSW. Additionally all 46 commercially grown varieties in the experiment were included in this study regardless of Zadoks’ score. For the second year of this study in 2010, the
same breeding lines and named varieties were selected for WSC measurement except for 18 breeding lines that were excluded from the overall experiment in that year.

4.3.2 Field experiment locations

This study utilised field sites at Coleambally Community Experimental Demonstration Farm (elevation: 122m, location 34° 44’ 01.63”S 145° 57’ 30.16”E) and Yanco Agricultural Institute (elevation: 138m, location 34° 37’ 16.88” S, 146° 24’ 43.79” E). The Coleambally site soil type is a highly variable gilgai complex of red-brown loam (known locally as Willbriggie loam) and brown self-mulching clay (known locally as Yooroobla clay), while the Yanco site is predominately red-brown earth, locally known as Birganbigal clay loam (van Dijk, 1961; Hornbuckle & Christen, 1999).

Experiment sites were set out as land-formed irrigation bays for flood irrigation, with the exception of the Coleambally irrigated site in 2010 which utilised a raised bed irrigation layout.
Figure 4.1 Histogram of the Yanco irrigated experiment Zadoks’ development score of relative maturity (ZAD) for 10 September 2009. Breeding lines with scores between Z49 and Z56 (early head emergence to 50% anthesis, bars in blue) were identified for WSC measurement. This range is equivalent to approximately 3-5 days difference in anthesis date.
4.3.3 Experimental design

The field experiments were grown in irrigation bays at both the Yanco and Coleambally sites. A split-plot design was used, in which the irrigated and rainfed treatments were the main plots, and the 990 genotype entries were the sub-plots, among them the subset of genotypes for WSC phenotyping described above in section 4.3.1. There were two replicates of each treatment at each site. The placement of genotypes within field experiment layouts was optimised with the spatial design package DiGGer (Coombes, 2002). A buffer of the variety Chara, popular in southern NSW irrigated wheat areas, was sown around the experiment to completely fill the irrigation bays with cropped area. Figure 4.2 illustrates the general layout of experiments, where each replicate of a treatment has been grown in a separate irrigation bay.

For the laboratory phase measuring WSCC using near-infrared spectroscopy (NIRS), an experimental design structured by day of measurement and NIRS instrument carousel and well was developed and implemented. Samples from both field sites were pooled into one experimental design for each year, and the placement of genotypes within the laboratory experimental phase was also optimised with DiGGer (Coombes, 2002), with partial replication of 12% of experiment field plots sampled (i.e. a replication level of 1.12), following the methods in Cullis et al. (2006) and Smith et al. (2006).
Figure 4.2 Aerial view of experiment site at Coleambally NSW in 2009. The irrigated treatments can be seen in the centre bays and rainfed treatments on either side. Similar layouts were utilised for other experiments.
4.3.4 Field experiment management and agronomy

Agronomic practices for each field experiment followed best management practice for high-yielding irrigated wheat in southern NSW, derived from the recommendations of Lacy and Giblin (2006) for achieving high-yielding wheat crops under irrigation for the irrigated treatments at each site. Briefly, experiments were sown on a full profile of moisture, achieved by flood irrigating each site four to six weeks before sowing. This technique of pre-sowing irrigation is a common practice in southern NSW irrigated wheat systems and was included in both irrigated and rainfed treatments so that the focus on water deficit conditions would be in the later stages of crop growth. Sowing dates were targeted for the first two weeks of May, in line with recommendations of McRae et al. (2009), and sowing rates of 115kg/ha in the irrigated and 70kg/ha in the rainfed treatments, respectively. Nitrogen was applied at 17kg/ha for the irrigated and 10kg/ha for the rainfed treatments, and phosphate was applied at 73kg/ha for irrigated and 44kg/ha for the rainfed treatments, in the form of mono-ammonium phosphate fertiliser incorporated with the seed at sowing. Fertiliser was treated with 100g/ha flutriafol (400ml/ha Impact®) for control of fungal diseases in the early stages of growth.

Experiments were subject to a strict weed, pest and disease control regime to maximise yield potential. Fungal diseases including stripe rust (Puccinia striiformis f.sp. tritici), stem rust (Puccinia graminis f.sp. tritici), leaf rust (Puccinia triticina), crown rot (Fusarium pseudograminearum), Septoria tritici blotch and yellow spot (Pyrenophora tritici‐repentis) disease were controlled with regular 125g/ha applications of the foliar fungicide proiconazole (Bumper®). In 2009 sowings, aphid infestation was identified as a risk to crop health so experiments at both sites were treated prophylactically with 60g/ha pirimicarb (Pirimor® WG). Locusts were observed at establishment at the
Coleambally site in 2010, and were subsequently treated with 20g/ha alpha-cypermethrin (Astound®).

Nitrogen management for the irrigated treatments at each site targeted a 10t/ha crop yield potential (at 11.5% grain protein) adapted from the protocol of Lacy and Giblin (2006). Pre-sowing nitrogen was targeted to be in the vicinity of 120kg N/ha from the combination of deep soil nitrogen (following soil testing – data not shown) and fertiliser at seeding in order to provide adequate but not excessive nitrogen for vegetative growth. In general, sites met this objective although total starting nitrogen at Coleambally in 2009 was approximately 30-40kg N/ha more than targeted. To compensate for the high soil nitrogen levels at Coleambally in 2009, and for potentially high nitrogen mineralisation at Yanco in 2009 due to a paddock history of lucerne (Medicago sativa L.), the first nitrogen topdressing was limited to 50kg N/ha. First nitrogen topdressings for the irrigated treatments at both sites in 2010 were increased to 70kg N/ha. Second topdressings for the irrigated treatments at each site were applied at 110kg N/ha bringing total nitrogen for the irrigated treatments in the experiments to approximately 300kg N/ha consistent with estimated nitrogen demand by the crops. Rainfed treatments at each site were not supplied with additional fertiliser after sowing, in view of the limited yield potential without irrigation. In this way, expression of WSC accumulation at each treatment should differ relative to water availability rather than nutrient limitation, as the fertiliser requirement is adequate for the level of growth relative to water availability.

4.3.5 Field experiment weather, rainfall and irrigation

The 2009 and 2010 seasons contrasted in weather conditions across the separate growing seasons. Soil moisture was monitored using gypsum block AM400 soil
moisture data loggers (Hansen, Wenatchee, USA) from August through to physiological maturity. Data loggers were placed at 15, 30 and 60cm depths in 2009, and 30 and 60cm in 2010. The data loggers were located in the buffer surrounding the experiments (visible in Figure 4.2). Onsite weather stations (Davis Instruments, Hayward, USA) were used to record rainfall and temperatures from sowing through to harvest. Daily maximum and minimum temperatures are presented in Figure 4.3, while daily rainfall totals are shown in Figure 4.4.

In 2009, both sites experienced below average rainfall, and warmer air temperatures, particularly during the later stages of grain-filling (Figure 4.3 and Figure 4.4). Effective in-season (1 May to 30 November) rainfall for the Yanco site was 171mm, and 161mm for the Coleambally site. Irrigation scheduling was intended to maintain soil moisture availability above -100kPa, with irrigations commencing as soil water potential fell below -75kPa. High evapotranspiration required four irrigations at the Yanco irrigated site and three at the Coleambally irrigated site. Despite this, Figure 4.4 shows that the Yanco irrigated site experienced short periods of water deficit before each irrigation. Moisture availability in both rainfed treatments declined rapidly from the commencement of soil moisture measurement, and data loggers at all depths fell below -100kPa at the Yanco rainfed site during mid to late August and Coleambally rainfed site in mid-September.

Field experiments in 2010 commenced with a full moisture profile following heavy rainfall at both sites during February and March. Weather conditions throughout the year were cooler than average, with substantially above average rainfall during late spring (coinciding with the crop grain-filling period). The Yanco site received 379mm rainfall in the growing season, while the Coleambally site received 472mm. Moisture availability declined slowly through spring at both sites
with rainfall maintaining soil moisture above -100kPa until early October. Irrigated treatments at both sites received irrigations at this time and the Yanco rainfed site experienced a short period of water deficit before all sites received heavy rainfall in mid-October. As rainfall followed irrigation, waterlogging occurred at the irrigated treatments at both sites for a period of approximately two weeks (Figure 4.4) although the Coleambally irrigated treatment was grown on raised beds and was waterlogged for a shorter period of time than the rainfed treatment at the same site. Soil water availability remained high at the Coleambally site until harvest, and at Yanco declined below -100kPa by mid-November. Temperatures were correspondingly mild through spring, with low maxima and minima particularly in late spring (Figure 4.3), corresponding with the peak grain-filling period.
Figure 4.3 Daily maximum and minimum temperatures recorded by on-site weather stations throughout the growing season. The first date of sowing and the final date of harvest at each site are also indicated.
Figure 4.4 Daily rainfall totals and soil moisture potentials recorded by on-site weather stations throughout the growing season, including the timing of irrigations ('irr') to the irrigated experiments, as well as the first date of sowing ('sow') and the final date of harvest ('har') at each experiment.
4.3.6 Measurements and observations

Phenotypic measurements were conducted through the growing season at each site. Relative maturity was determined using the Zadoks decimal score for plant development (Zadoks et al., 1974). Three separate scores were made over the period across anthesis. For each field experiment, the score with mean development stage closest to the midpoint of the range Z50-Z69 (head emergence to completion of anthesis) was selected for analysis given scores at this stage are approximately linear (Dr Greg Rebetzke, pers. comm.) and this measure of relative maturity around flowering time is referred to hereafter as ZAD.

Genotypes selected for WSC analysis were sampled when the irrigated treatments at each site were approximately 180˚Cd post-anthesis following the sampling method of Rebetzke et al. (2008). A 50cm long section of row (0.09m²) was cut at ground level from a representative section of each plot and weighed. Samples were dried at 80˚C for three days in a forced-air dehydrator before weighing. The length of dried biomass bundles was measured as a surrogate for plant height, and the numbers of stems and spikes was recorded as an estimate of tiller number per m².

For WSC analysis approximately 5-10 stalks (including leaves, leave sheaths and heads, but not senesced plant material) were sampled from each bundle and ground through a 2mm screen using a pasture chopper. Ground biomass samples were homogenised and subsampled for scanning by NIRS with a Bruker Multi-purpose Analyser (Bruker Optik GmbH, Ettlingen, Germany) and OPUS software (version 5.1). The NIRS samples were dehydrated at 80˚C to remove moisture, and allowed to cool to room temperature in a desiccation cabinet before scanning. Scanned spectra were transformed using the first derivative and multiplicative scatter correction to minimise the influence of sample particle size on spectra interpretation. The spectral region
between 904-2354nm was selected for calibration construction to cover major absorption bands for carbohydrate groups and avoid the often distorted ends of the spectra profile.

Calibrations to obtain predicted WSCC values from spectra measurements were constructed using the “Quant 2 Method” component of the OPUS software. A random subset of 10% of samples from each laboratory phase experiment were identified, and WSCC for the calibration samples determined by the NSW Department of Primary Industries Feed Quality Service (Wagga Wagga) using the alkaline ferricyanide method (Piltz & Law, 2007). This method uses cold water with benzoic acid to extract WSC from the sample, as well as kill microorganisms that would otherwise metabolise the carbohydrates. Filtered extracts are then analysed by flow injection analyser by quantifying alkaline ferricyanide discolouration. This involves hydrolysisation of the extracted carbohydrate to invert sugar by hydrochloric acid and heat, then dialysis of the invert sugar into an alkaline stream of potassium ferricyanide which was then heated. The invert sugar reduces the yellow ferricyanide to colourless ferrocyanide, and the decrease in colour at 420nm is proportional to the concentration of WSC in the extract. WSC values for calibration set samples were used to develop a linear model for predicting WSC from all spectra. For the 2009 experiment, the coefficient of determination \((r^2)\) for the linear model was 92.36 and the root-mean-square error of cross-validation (RMSECV) was 15.4. For the 2010 experiment, the coefficient of determination for the linear model was 91.79 and the RMSECV = 16.

Prior to harvest, lodging (straw strength) was measured by a system of 0-10 visual scoring where 0 corresponds to erect plants and 10 to completely lodged plants. Rainfed treatments at both sites in 2009 were not scored for lodging as little variation was observed for this trait in these sowings. Harvest samples were used to measure
grain yield, and subsequent grain quality measurements of grain protein, screenings, thousand grain weight, and test weight following the methods described in Sissons et al. (2014).

A number of additional traits were estimated from the physical data collected. Dry matter content ratio (DMC) was estimated from dry and fresh weights following Xue et al. (2009). WSC mass per unit area (WSCA) was calculated fromWSCC and biomass as described in Rebetzke et al. (2008). Numbers of grains per m² was calculated from grain yield and thousand grain weight, while grains per spike was estimated from yield, thousand grain weight and tillers per m².

4.3.7 Statistical methods

A multiplicative mixed linear model (Gilmour et al., 1997) was used to analyse the multi-experiment data for each trait following the approach of Smith et al. (2001b) and Beeck et al. (2010). The linear mixed model for the response variable \( y = (y_1, ..., y_p) \), for \( p \) experiments \((j = 1, 2, 3, ... p)\) and \( m \) varieties \((i = 1, 2, 3, ... m)\) combining the variety main effects (genotype) and the G × E interaction effects is given by

\[
y = X\tau + Z_g g + Z_u u + \eta
\]

Equation 4.1

where \( y \) is the \( (n \times 1) \) data vector of the response variable across \( p \) experiments with \( N_j \) plots per experiment \( j \); \( \tau \) is a \( (t \times 1) \) vector of fixed effects with associated design matrix \( X \), including experiment (a combination of site, year, and irrigation treatment), other designed based effects, and any linear trends of experiment range and row where they were statistically significant. The term \( u \) is a random component with associated design matrix \( Z_u \) and contains experiment-specific terms used to capture design based extraneous variation, such as the blocking structure of the field experiments (replicate and irrigation bay), and the interaction of experiment range and row effects.
The WSC analysis included additional laboratory phase terms for day of measurement, NIRS carousel and carousel well according to the multiphase experimental design. Linear trends for NIRS carousel were fitted as fixed effects where statistically significant. The laboratory phase blocking structure of day of measurement, and the interaction of day of measurement with NIRS carousel were fitted as random effects. Some other experimental factors were fitted as random effects where statistically significant. For the analysis of WSCC and WSCA, the batch in which the anthesis biomass cuts were dried, and the batches in which these dried samples were ground (between grinder blade adjustments) were fitted to the model. For the analysis of tiller numbers per m² and anthesis bundle length (both manually measured) operator was fitted as a random effect.

The residual error is \( \eta = (\eta_1, \ldots, \eta_p) \) which at the \( j \)th experiment was assumed to have distribution \( \eta_j \sim N(0, \sigma_j^2 R_j) \) where \( \sigma_j^2 \) is the residual variance for the \( j \)th experiment and \( R_j \) is a matrix that for the field-only trait models contains a parameterization for a separable autoregressive ARI \( \otimes \) AR1 process to model potential correlation of the observations due to the neighbouring plots of the experiment, where \( \otimes \) is the Kronecker product. For the multiphase analysis of water soluble carbohydrate concentration and content, residual variance for each year was modelled.

For each trait analysed, the most parsimonious random model was determined using log likelihood ratio tests following the methodology of Stram and Lee (1994). Fixed effects were tested for significance using Wald F-statistics (Kenward & Roger, 1997). Variograms for each experiment were generated for spatial model selection following the approach of Gilmour et al. (1997) in Stefanova et al. (2009).
The term $g$ is a random component with associated design matrix $Z_g$ used to model the genotype and G × E interaction, where $g$ is expressed as a multiplicative $k$-factor analytic (FA) model following the methods of Smith et al. (2001b), and is given by

$$g = (A \otimes I_m)f + \delta$$

**Equation 4.2**

where $A$ is a matrix with $j$th column containing the $j$th factor loadings for the $p$ experiments, $f$ is a vector of genotype scores across the $p$ experiments and $\delta = (\delta_1, ..., \delta_p)$ is a residual genetic term where, at the $j$th experiment, $\delta_j \sim N(0, \sigma_{gj}^2 I_m)$, and $\sigma_{gj}^2$ is the residual genetic variance for the $j$th experiment. The term $I_m$ represents an $m \times m$ identity matrix. The FA model with $k$ factors may be thought of as a multiple regression of the genotype effects in each environment on $k$ environmental covariates (also called loadings) with separate slopes (scores) for each genotype.

The variance model for the combined genotype and G × E effects is given by

$$\text{var}(g) = (AA' + \psi) \otimes I_m$$

**Equation 4.3**

where $\psi$ is a diagonal matrix of the $p$ environment specific variances, thus accommodating the variance-covariance heterogeneity between experiments. The process of selecting the optimum FA model for each trait followed the procedure described in Beeck et al. (2010), which is based on both formal and informal diagnostics. Here the most parsimonious models were identified using the Akaike Information Criterion (AIC); (Akaike, 1974) and the proportion of genotype and G × E interaction variation explained by the regression implicit in the FA model (Cullis et al., 2010). For traits where the AIC continued to decrease beyond a four factor FA model,
the fourth order model was used, given Smith (1999) has shown that the number of factors \( k \) that can be fitted must be no greater than the largest integer satisfying

\[
k \leq \frac{1}{2} \left( 2p + 1 - \sqrt{8p + 1} \right)
\]

**Equation 4.4**

for any given data set of \( p \) experiments before over-fitting becomes likely. The FA model used for each trait, and the percentage of genotype and G × E variance accounted for by the model were obtained. Experiments with no genetic variance for the trait under analysis were excluded from the model, as it is not possible to include them in a FA analysis (discussed in Beeck et al., 2010).

Empirical Best Linear Unbiased Predictors [E-BLUPs, as described in Kelly et al. (2007) and Beeck et al. (2010)] were obtained from the FA models described according to the methods of Beeck et al. (2010) and Cullis et al. (2010). These E-BLUPs are related to the regression implicit in the FA model and are centred at zero. To give appropriate scale for each trait, experiment raw means were added to the E-BLUPs for predictions at each experiment, and the overall raw means were added to the E-BLUPs for across-experiments predictions. For each trait, cluster analysis following Cullis et al. (2010) was performed to elucidate the genetic correlation between experiments. The correlation matrix for each trait is presented graphically using a heatmap, so the rows and columns of this matrix are re-ordered to aid visualisation of the clusters. The order of environments within each heatmap is based on the dendrogram obtained using the agnes package (an agglomerative hierarchical clustering algorithm) in R (R Development Core Team, 2012). Therefore, environments that are highly correlated are located closer together on the heatmap, and less correlated environments will be further apart.
Simple genetic variance models were produced to estimate the proportion of variance due to genetic and environmental (site, year and irrigation treatment) factors, and residual variance, for each trait. These model terms are nested within the FA models described above, and are an estimation of the G × E interactions in the data.

Broad sense heritability ($H^2$) for each trait at each experiment was calculated following the generalised formula for unbalanced data in Cullis et al. (2006):

$$H^2 = 1 - \frac{PEV}{2\sigma_g^2}$$

**Equation 4.5**

where $PEV$ is the average pairwise prediction error variance of genotype obtained from the model predict statement, and $\sigma_g^2$ is the genetic variance component taken from the single site model. Heritability were estimated using only the genotypes set used for WSC analysis for traits where phenotypes were available for the full genotype set to ensure statistics were comparable. Because heritability was estimated from the single site model, heritability was calculated at sites which had no genetic variance in the across-experiments FA model, such as the heritability estimations for WSCA and anthesis biomass at the 09YANA_RFD experiment.

Estimated genetic gain (EGG) was also calculated from Cullis et al. (2006) using:

$$EGG = i\sigma_g H$$

**Equation 4.6**

Where $i$ is the selection intensity corresponding to the mean of the top x% of a population with normal distribution (Falconer & Mackay, 1996), $\sigma_g$ is the square root of the genetic variance component taken from the model, and $H$ is the square root of the heritability given by the generalised formula in Equation 4.5 above. This formula is
equivalent to the response to selection equation described in Falconer and Mackay (1996).

Correlated response, $CR$, of trait $k$ to selection of trait $j$ was calculated from Falconer and Mackay (1996) using:

$$CR_k = iH_jH_k r_g \sigma_{p_k}$$

Equation 4.7

Where $H_j$ and $H_k$ were obtained following the generalised formula given in Equation 4.5, $r_g$ was obtained from the variance components of a single site bivariate model with both of the traits in consideration following Holland (2006), and $\sigma_{p_k}$ was obtained from the square root of the sum of all the variance components including by reason of consistency the explained variance components of experimental blocking structures, as these were also part of the model used to estimate $H_k$. The relative selection efficiency (RSE) was calculated from Falconer and Mackay (1996) (their Equation 19.9) using CR as a ratio of EGG to derive a measure of the merit of indirect selection.

All data were analysed using the software package ASReml-R (Butler et al., 2009), in the R environment (R Development Core Team, 2012).

4.4 Results

4.4.1 Trait means for agronomic and quality characteristics

Table 4.1 summarises 95% confidence intervals for means in each experiment for the different traits measured. The variation in WSCC reflect site and year variation. However comparisons can be drawn between irrigation treatments. For example, in 2009, the rainfed and irrigated treatment WSCC means at each site were similar whereas in 2010, the rainfed treatment WSCC means were larger than the irrigated treatment means at each site. The means in WSCA were smaller for the rainfed
compared to the irrigated treatments in 2009, reflecting the higher anthesis biomass accumulation in the well-watered irrigated sites. For both WSCA and anthesis biomass, no significant genetic variance was observed at 09YANA_RFD and this experiment could not be included in the across-experiments FA model. In 2010, when both treatments did not experience water deficit, mean WSCA values were greater in the irrigated than the rainfed treatments. Dry matter content ratio was larger in the 2009 rainfed experiments than the 2009 irrigated and all 2010 experiments, reflecting differences between water deficit and well-watered environments. Tiller numbers per m² were greater in the irrigated than the rainfed for each treatment. Confidence intervals for the mean mature plant heights were consistent across the experiments. Mean lodging scores also reflect the difference between irrigated and rainfed treatments. No lodging was observed in the 2009 rainfed treatments at both sites, therefore this trait was not scored in those experiments. The irrigated treatments had higher mean lodging than the rainfed treatments in 2010. Relative maturity at flowering time, measured by Zadoks score when the experiments were approximately at mid-anthesis, did not change much over all experiments, ranging from Z52 to Z59. Thousand grain weight was lower in the rainfed treatments in 2009, and in 2010 showed no clear trend between treatments. Screenings results followed the same trends. Grain protein content differed across sites in 2009, and by irrigation treatment in 2010. Confidence intervals for mean grain yield were higher in the irrigated than the rainfed treatments in 2009, but the yields for all experiments in 2010 were similar, except for the Yanco irrigated experiment, which averaged over 8t/ha.

Histograms for selected traits including relative maturity (Figure 4.5 show the subset of genotypes analysed for WSC, Figure 4.6 shows all genotypes), anthesis
biomass (Figure 4.7), WSCC (Figure 4.8), WSCA (Figure 4.9), and grain yield (Figure 4.10) show the approximately normal distribution of trait observations.
Table 4.1: Predicted experiment mean confidence intervals (95%) for each trait. WSCC (mg/g), WSCA (g/m²), anthesis biomass (g/m²), dry matter content (DMC), tillers per m² at anthesis, height at maturity (cm), lodging score (scale 0-9 where 9 = completely lodged), relative maturity (ZAD) (Zadoks scale 0-99), thousand grain weight (g), screenings <2mm (%), grain protein content (%), and grain yield (t/ha.).

<table>
<thead>
<tr>
<th>Experiment</th>
<th>WSCC</th>
<th>WSCA</th>
<th>Anthesis biomass</th>
<th>DMC</th>
<th>Tillers</th>
<th>Plant height</th>
<th>Lodging score</th>
<th>Relative maturity (ZAD)</th>
<th>Thousand grain weight</th>
<th>Screenings</th>
<th>Grain protein</th>
<th>Grain Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>09COLE_IRR</td>
<td>175.97-</td>
<td>318.16-</td>
<td>1825.53-</td>
<td>0.346-</td>
<td>552.94-</td>
<td>89.92-</td>
<td>3.52-3.68</td>
<td>59.41-</td>
<td>42.11-42.60</td>
<td>NA</td>
<td>13.15-</td>
<td>7.34-</td>
</tr>
<tr>
<td></td>
<td>176.03</td>
<td>331.30</td>
<td>1840.27</td>
<td>0.348</td>
<td>560.29</td>
<td>90.61</td>
<td></td>
<td></td>
<td></td>
<td>60.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>09COLE_RFD</td>
<td>172.68-</td>
<td>266.09-</td>
<td>1556.66-</td>
<td>0.503-</td>
<td>501.44-</td>
<td>87.40-</td>
<td>NA*</td>
<td>57.03-</td>
<td>36.05-36.44</td>
<td>3.27-3.41</td>
<td>11.71-</td>
<td>5.10-</td>
</tr>
<tr>
<td></td>
<td>172.73</td>
<td>271.84</td>
<td>1565.76</td>
<td>0.505</td>
<td>506.87</td>
<td>87.97</td>
<td></td>
<td></td>
<td></td>
<td>57.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>09YANA_IRR</td>
<td>149.29-</td>
<td>265.29-</td>
<td>1776.16-</td>
<td>0.325-</td>
<td>611.14-</td>
<td>92.03-</td>
<td>7.43-7.57</td>
<td>52.12-</td>
<td>40.21-40.66</td>
<td>1.26-1.32</td>
<td>13.4-</td>
<td>6.73-</td>
</tr>
<tr>
<td></td>
<td>149.34</td>
<td>275.05</td>
<td>1786.88</td>
<td>0.327</td>
<td>618.52</td>
<td>92.67</td>
<td></td>
<td></td>
<td></td>
<td>52.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>09YANA_RFD</td>
<td>131.49-</td>
<td>NA</td>
<td>NA</td>
<td>0.490-</td>
<td>550.69-</td>
<td>83.22-</td>
<td>NA*</td>
<td>55.49-</td>
<td>33.27-33.62</td>
<td>3.84-4.10</td>
<td>14.07-</td>
<td>4.42-</td>
</tr>
<tr>
<td></td>
<td>131.54</td>
<td></td>
<td></td>
<td>0.492</td>
<td>554.85</td>
<td>83.76</td>
<td></td>
<td></td>
<td></td>
<td>56.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10COLE_IRR</td>
<td>166.15-</td>
<td>254.63-</td>
<td>1535.58-</td>
<td>0.311-</td>
<td>582.50-</td>
<td>92.51-</td>
<td>6.81-6.96</td>
<td>55.43-</td>
<td>40.99-41.50</td>
<td>3.01-3.11</td>
<td>14.04-</td>
<td>6.66-</td>
</tr>
<tr>
<td></td>
<td>166.20</td>
<td>264.18</td>
<td>1545.40</td>
<td>0.313</td>
<td>588.50</td>
<td>93.07</td>
<td></td>
<td></td>
<td></td>
<td>56.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10COLE_RFD</td>
<td>188.99-</td>
<td>275.20-</td>
<td>1469.25-</td>
<td>0.324-</td>
<td>523.25-</td>
<td>92.35-</td>
<td>5.44-5.62</td>
<td>54.20-</td>
<td>42.23-42.67</td>
<td>3.37-3.48</td>
<td>11.97-</td>
<td>6.71-</td>
</tr>
<tr>
<td></td>
<td>189.03</td>
<td>282.72</td>
<td>1478.56</td>
<td>0.326</td>
<td>529.41</td>
<td>92.90</td>
<td></td>
<td></td>
<td></td>
<td>54.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10YANA_IRR</td>
<td>180.78-</td>
<td>272.02-</td>
<td>1508.78-</td>
<td>0.288-</td>
<td>586.87-</td>
<td>97.79-</td>
<td>5.04-5.24</td>
<td>52.35-</td>
<td>40.55-41.03</td>
<td>3.23-3.35</td>
<td>14.11-</td>
<td>8.17-</td>
</tr>
<tr>
<td></td>
<td>180.83</td>
<td>280.99</td>
<td>1516.42</td>
<td>0.290</td>
<td>593.47</td>
<td>98.45</td>
<td></td>
<td></td>
<td></td>
<td>52.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10YANA_RFD</td>
<td>210.23-</td>
<td>323.31-</td>
<td>1542.46-</td>
<td>0.321-</td>
<td>560.38-</td>
<td>99.06-</td>
<td>1.91-2.01</td>
<td>52.01-</td>
<td>43.82-44.34</td>
<td>2.98-3.08</td>
<td>12.31-</td>
<td>6.70-</td>
</tr>
<tr>
<td></td>
<td>210.27</td>
<td>330.54</td>
<td>1551.95</td>
<td>0.322</td>
<td>567.15</td>
<td>99.80</td>
<td></td>
<td></td>
<td></td>
<td>52.56</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Lodging scores were not recorded for the two water deficit experiments 09COLE_RFD and 09YANA_RFD as no lodging was observed.
Figure 4.5 Histogram of Zadoks development score for the 365 genotypes identified for WSC measurement at each experiment.
Figure 4.6 Histogram of Zadoks development score for all of the genotypes at each experiment (990) including the lines grown but not measured for WSC.
Figure 4.7 Histogram of anthesis biomass for the genotypes identified for WSC measurement at each experiment.
Figure 4.8 Histogram of WSC concentration (WSCC) for the genotypes at each experiment.
Figure 4.9 Histogram of WSC amount (WSCA) for the genotypes at each experiment.
Figure 4.10 Histogram of grain yield for the genotypes identified for WSC measurement at each experiment.
4.4.2 Genotype × environment interactions

The genetic correlations between experiments are presented as heatmaps of line ranking correlations. For relative maturity, genetic correlations between experiments were very high, ranging from $r_G = 0.92$ to 0.99 (Figure 4.11). Anthesis biomass genetic correlations across experiments ranged from $r_G = -0.20$ to 0.99, with no clear environment clustering observed (Figure 4.12). Genetic correlations of experiments for WSCC are shown in Figure 4.13. The experiments clustered into two distinct environment groups, with the Yanco and Coleambally 2009 rainfed experiments which experienced terminal water deficit forming one cluster, and all of the other experiments that did not forming the other (clearly matching the water relations differences illustrated in Figure 4.4). Within these clusters, the genetic correlations were maintained at $r_G = 0.87$ between the two rainfed treatments that make up the water deficit environment cluster, and ranged from $r_G = 0.74 - 0.98$ in the well-watered environment cluster. Between the two clusters, genetic correlations ranged from $r_G = 0.02$ to 0.35. The WSCA experiment values formed two clusters. One large cluster contained the well-watered environment experiments observed for WSCC, and the other cluster consisted of the 09COLE_RFD experiment, which was the only water deficit environment experiment included in the across-experiments model. For the well-watered experiments, genetic correlations ranged from $r_G = 0.63$ to 0.99 (Figure 4.14). Figure 4.15 shows the genetic correlation groupings for grain yield. Here, the clustering between experiments is less defined than with WSCC, and the 2009 Yanco irrigated experiment clusters with the 2009 rainfed experiments, and the other experiments form the main cluster. Within the well-watered cluster, genetic correlations range from $r_G = 0.63$ to 0.94 and within the water deficit experiments and
the 2009 Yanco irrigated the genetic correlations range from $r_c = 0.72$ to 0.95. Between these two clusters genetic correlations ranged from $r_c = 0.35$ to 0.73.

### 4.4.3 Trait correlations

The heatmaps of genetic correlations between experiments for WSCC (Figure 4.13) and WSCA (Figure 4.14) show the experiments clustered into a water deficit environments cluster of the 09COLE_RFD and 09YANA_RFD experiments, and a well-watered environment for the other six experiments. Therefore, E-BLUPs for a range of traits were averaged across experiments within each environment cluster and plotted against WSCC (Figure 4.16 and Figure 4.17) and WSCA (Figure 4.18 and Figure 4.19). However, for WSCA, only 09COLE_RFD was used for the water deficit E-BLUPs as 09YANA_RFD could not be included in the across-experiments FA model due to lack of significant genetic variance.

For WSCC, the strongest correlations were observed with maturity, with a positive relationship in the well-watered environment cluster (0.507), and a negative relationship for the water deficit environment cluster (-0.428) (Table 4.2). Dry matter content ratio also had a positive relationship in the well-watered environment cluster (0.475) and a negative relationship in the water deficit environment cluster (-0.381). Lines were stratified by maturity in different colours in Figure 4.16 and Figure 4.17 to highlight any interactions between relative maturity and WSCC. A negative correlation between WSCC and tillers per m² in the well-watered environment cluster (-0.343) decreases to -0.214 in the water deficit environment cluster. Grain yield and anthesis biomass did not show significant ($p < 0.05$) correlations with WSCC in either environment cluster (Table 4.2). Grains per m² showed a weak but significant ($p < 0.05$) negative correlation in the well-watered environment cluster (-0.299), and this
increased to -0.369 in the water deficit environment cluster. Thousand grain weight showed a moderate and significant (p < 0.05) positive correlation (0.437) with a negative correlation for screenings (-0.299) in the well-watered environment cluster. These relationships were significant at (p < 0.05) in the water deficit environment cluster, although not as strong (0.361 for thousand grain weight, -0.333 for screenings). Grain protein and mature plant height had significant but weak correlations with WSCC in the water deficit environment cluster, but were not significant in the well-watered environment cluster.

Correlations between WSCA and other traits shown in Figure 4.18 and Figure 4.19 are detailed in Table 4.3. Relative maturity at flowering time was significantly (p < 0.05) positively correlated with WSCA in both well-watered (0.487) and water deficit environment clusters (0.449). Dry matter content ratio was positively correlated in the well-watered (0.540) and water deficit (0.323) environment clusters. Anthesis biomass was also significantly (p < 0.05) correlated in both well-watered (0.463) and water deficit environment clusters (0.359). Thousand grain weight was significantly (p < 0.05) correlated with WSCA in the well-watered (0.451) environment cluster and water deficit environment clusters (0.418) as well. Significant but weak correlations were observed in both environment clusters between WSCA and grains per m², tillers per m², grain protein and screenings. WSCA is significantly (p < 0.05) but weakly correlated (0.249) with grain yield for the water deficit environment but not the well-watered environment cluster. Mature plant height was not significantly (p < 0.05) correlated with WSCA at either environment cluster.
Figure 4.11 Heatmap of relative maturity at flowering time (Zadoks development score) genetic correlations between experiments
Figure 4.12 Heatmap of anthesis biomass per m² genetic correlations between experiments. The across-experiments FA model did not include 09YANA_RFD as this experiment had no significant genetic variance in the model.
Figure 4.13 Heatmap of genetic correlation of WSC concentration between experiments.
Figure 4.14 Heatmap of genetic correlation of WSC amount per m² between experiments. The across-experiments FA model did not include 09YANA_RFD as this experiment had no significant genetic variance in the model.
Figure 4.15 Heatmap of genetic correlation of grain yield between experiments.
Figure 4.16 Plots of WSC concentration E-BLUPs with grain yield, grains per m$^2$, anthesis biomass, tillers per m$^2$ and maturity E-BLUPs. Genotypes are coloured by maturity group, where blue = winter types, purple = facultative, green = longer maturity spring types, red = shorter maturity spring types, and yellow = short season genotypes.
Figure 4.17 Plots of WSC concentration E-BLUPs with grain protein, lodging, mature plant height, screenings and thousand grain weight E-BLUPs. Lodging was not measured at either of the two water deficit environment experiments (09YANA_RFD and 09COLE_RFD). Genotypes are coloured by maturity group following Figure 4.16
Figure 4.18 Plots of WSC amount E-BLUPs with grain yield, grains per m², anthesis biomass, tillers per m² and maturity E-BLUPs. Genotypes are coloured by maturity group following Figure 4.16. The water deficit environment cluster consists of only one experiment (09COLE_RFD) as shown in Figure 4.14.
Figure 4.19 Plots of WSC amount E-BLUPs with grain protein, lodging, mature plant height, screenings and thousand grain weight E-BLUPs. Genotypes are coloured by maturity group following Figure 4.16. The water deficit environment cluster consists of only one experiment (09COLE_RFD) as shown in Figure 4.14, and lodging was not scored at that experiment.
Table 4.2: Pearson’s product-moment correlation coefficient ($r$), coefficient of determination ($r^2$), correlation p-value, and correlation 95% confidence intervals (CI) for trait E-BLUPs with WSCC at both well-watered (WW) and water deficit (WD) environment clusters, following Figure 4.16 and Figure 4.17.

<table>
<thead>
<tr>
<th>Trait</th>
<th>WW $r$</th>
<th>WW $r^2$</th>
<th>WW p-value</th>
<th>WW 95% CI for $r$</th>
<th>WD $r$</th>
<th>WD $r^2$</th>
<th>WD p-value</th>
<th>WD 95% CI for $r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grain yield</td>
<td>-0.045</td>
<td>0.002</td>
<td>0.397</td>
<td>-0.146 - 0.058</td>
<td>-0.125</td>
<td>0.016</td>
<td>0.017</td>
<td>-0.225 - -0.023</td>
</tr>
<tr>
<td>Grains per m$^2$</td>
<td>-0.299</td>
<td>0.090</td>
<td>&lt;0.01</td>
<td>-0.390 - -0.203</td>
<td>-0.369</td>
<td>0.136</td>
<td>&lt;0.01</td>
<td>-0.455 - -0.277</td>
</tr>
<tr>
<td>Anthesis biomass</td>
<td>0.070</td>
<td>0.005</td>
<td>0.181</td>
<td>-0.033 - 0.172</td>
<td>-0.082</td>
<td>0.007</td>
<td>0.120</td>
<td>-0.183 - 0.021</td>
</tr>
<tr>
<td>Tillers per m$^2$</td>
<td>-0.343</td>
<td>0.118</td>
<td>&lt;0.001</td>
<td>-0.430 - -0.249</td>
<td>-0.214</td>
<td>0.046</td>
<td>&lt;0.001</td>
<td>-0.310 - -0.114</td>
</tr>
<tr>
<td>Relative maturity</td>
<td>0.507</td>
<td>0.257</td>
<td>&lt;0.001</td>
<td>0.426 - 0.579</td>
<td>-0.428</td>
<td>0.183</td>
<td>&lt;0.001</td>
<td>-0.508 - -0.340</td>
</tr>
<tr>
<td>Dry matter content ratio</td>
<td>0.475</td>
<td>0.226</td>
<td>&lt;0.001</td>
<td>0.392 - 0.551</td>
<td>-0.381</td>
<td>0.145</td>
<td>&lt;0.001</td>
<td>-0.465 - -0.289</td>
</tr>
<tr>
<td>Grain protein</td>
<td>0.094</td>
<td>0.009</td>
<td>0.074</td>
<td>-0.009 - 0.194</td>
<td>0.155</td>
<td>0.024</td>
<td>0.003</td>
<td>0.053 - 0.253</td>
</tr>
<tr>
<td>Lodging score</td>
<td>0.068</td>
<td>0.005</td>
<td>0.193</td>
<td>-0.035 - 0.170</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Mature plant height</td>
<td>-0.007</td>
<td>0.000</td>
<td>0.897</td>
<td>-0.109 - 0.096</td>
<td>0.192</td>
<td>0.037</td>
<td>&lt;0.001</td>
<td>0.091 - 0.289</td>
</tr>
<tr>
<td>Screenings</td>
<td>-0.299</td>
<td>0.089</td>
<td>&lt;0.001</td>
<td>-0.389 - -0.202</td>
<td>-0.333</td>
<td>0.111</td>
<td>&lt;0.001</td>
<td>-0.421 - -0.238</td>
</tr>
<tr>
<td>Thousand grain weight</td>
<td>0.437</td>
<td>0.191</td>
<td>&lt;0.001</td>
<td>0.350 - 0.517</td>
<td>0.361</td>
<td>0.131</td>
<td>&lt;0.001</td>
<td>0.268 - 0.447</td>
</tr>
</tbody>
</table>
Table 4.3: Pearson’s product-moment correlation coefficient (r), coefficient of determination (r²), correlation p-value, and correlation 95% confidence intervals (CI) for trait E-BLUPs with WSCA at both well-watered (WW) and water deficit (WD) environment clusters, following Figure 4.18 and Figure 4.19.

<table>
<thead>
<tr>
<th>Trait</th>
<th>WW r</th>
<th>WW r²</th>
<th>WW p-value</th>
<th>WW 95% CI for r</th>
<th>WD r</th>
<th>WD r²</th>
<th>WD p-value</th>
<th>WD 95% CI for r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grain yield</td>
<td>-0.064</td>
<td>0.004</td>
<td>0.221</td>
<td>-0.166 - 0.039</td>
<td>0.249</td>
<td>0.062</td>
<td>&lt;0.001</td>
<td>0.150 - 0.343</td>
</tr>
<tr>
<td>Grains per m²</td>
<td>-0.313</td>
<td>0.098</td>
<td>&lt;0.001</td>
<td>-0.403 - -0.218</td>
<td>-0.173</td>
<td>0.030</td>
<td>0.001</td>
<td>-0.271 - -0.071</td>
</tr>
<tr>
<td>Anthesis biomass</td>
<td>0.463</td>
<td>0.214</td>
<td>&lt;0.001</td>
<td>0.378 - 0.540</td>
<td>0.359</td>
<td>0.129</td>
<td>&lt;0.001</td>
<td>0.266 - 0.446</td>
</tr>
<tr>
<td>Tillers per m²</td>
<td>-0.178</td>
<td>0.032</td>
<td>0.001</td>
<td>-0.275 - -0.077</td>
<td>-0.213</td>
<td>0.045</td>
<td>&lt;0.001</td>
<td>-0.309 - -0.113</td>
</tr>
<tr>
<td>Relative maturity</td>
<td>0.487</td>
<td>0.238</td>
<td>&lt;0.001</td>
<td>0.405 - 0.562</td>
<td>0.449</td>
<td>0.201</td>
<td>&lt;0.001</td>
<td>0.363 - 0.527</td>
</tr>
<tr>
<td>Dry matter content ratio</td>
<td>0.540</td>
<td>0.292</td>
<td>&lt;0.001</td>
<td>0.463 - 0.609</td>
<td>0.323</td>
<td>0.105</td>
<td>&lt;0.001</td>
<td>0.228 - 0.412</td>
</tr>
<tr>
<td>Grain protein</td>
<td>0.073</td>
<td>0.005</td>
<td>0.163</td>
<td>-0.030 - 0.175</td>
<td>-0.346</td>
<td>0.120</td>
<td>&lt;0.001</td>
<td>-0.433 - -0.252</td>
</tr>
<tr>
<td>Lodging score</td>
<td>0.135</td>
<td>0.018</td>
<td>0.010</td>
<td>0.033 - 0.235</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Mature plant height</td>
<td>0.042</td>
<td>0.002</td>
<td>0.429</td>
<td>-0.061 - 0.144</td>
<td>0.103</td>
<td>0.011</td>
<td>0.050</td>
<td>0.000 - 0.203</td>
</tr>
<tr>
<td>Screenings</td>
<td>-0.268</td>
<td>0.072</td>
<td>&lt;0.001</td>
<td>-0.361 - -0.170</td>
<td>-0.223</td>
<td>0.050</td>
<td>&lt;0.001</td>
<td>-0.319 - -0.123</td>
</tr>
<tr>
<td>Thousand grain weight</td>
<td>0.451</td>
<td>0.203</td>
<td>&lt;0.001</td>
<td>0.365 - 0.529</td>
<td>0.418</td>
<td>0.174</td>
<td>&lt;0.001</td>
<td>0.329 - 0.499</td>
</tr>
</tbody>
</table>
4.4.4 Heritability, estimated genetic gain, and correlated response to selection

Line-mean heritability for WSCC ranged from $H^2 = 0.43$ to 0.79 across experiments, and for WSCA, ranged from $H^2 = 0.45$ to 0.57 (Table 4.4). This compared to a range in heritability of $H^2 = 0.27$ to 0.40 for anthesis biomass, $H^2 = 0.49$ to 0.58 for dry matter content, $H^2 = 0.59$ to 0.86 for grain yield and $H^2 = 0.84$ to 0.91 for maturity (all measured from the group of genotypes used for WSC measurement, Table 4.4). The estimated genetic gain, or response to selection, (EGG) of WSCC at each site ranged from 25.92 to 39.52mg/g for 10% selection intensity (which is indicative of the population mean increase if the top 10% of the population was selected), and 18.77 to 28.62mg/g for 25% selection intensity (Table 4.5). For WSCA, EGG ranged from 37.73 to 86.63g/m² for 10% selection intensity, and 27.32 to 62.74g/m² for 25% selection intensity (Table 4.6). Heritability and response to selection estimates did not differ markedly between environment clusters. Correlated response to selection (CR) showed that using dry matter content ratio (DMC) as a surrogate trait was between 52-92% as effective as selecting for WSCC directly, and between 59-98% as effective as direct selection for WSCA. Selection for higher dry matter content in the rainfed environment cluster sites reduced WSCC and WSCA rather than increased it, as selection in the irrigated environment cluster sites. Selecting for increased WSCC had a highly variable correlated response to selection on yield, ranging from -0.51 to 0.18t/ha at 10% selection intensity, to -0.37 to 0.13t/ha at 25% selection intensity. Similarly, the correlated response to selection of WSCA on grain yield was small and variable, ranging from -0.37 to 0.15t/ha at 10% selection intensity and -0.27 to 0.11t/ha at 25% selection intensity. The effect on relative maturity from selection for increased WSCC was consistent, ranging from 2.09 to 3.28 units for the well-watered environment cluster.
sites to -3.69 and -4.76 units for the water deficit sites at 10% selection intensity, and 1.51 to 2.38 units for the well-watered, and -2.67 and -3.45 units for the water deficit sites at 25% selection intensity. For WSCA, correlated response of selection on maturity also showed a consistent negative response at the water deficit environments (-2.59 to -1.98 at 10% selection intensity, and -1.88 to -1.43 at 25% selection intensity) and positive selection response at the well-watered environments (2.13 to 2.46 at 10% selection intensity, and 1.51 to 1.78 at 25% selection intensity, Table 4.6). The range at both selection intensities for WSCC and WSCA on maturity reflects 1-2 days maturity difference given average temperatures during anthesis.
Table 4.4: Line mean heritability of WSC concentration, total WSC amount, anthesis biomass, dry matter content ratio, grain yield and maturity.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>WSCC</th>
<th>WSCA</th>
<th>Anthesis biomass</th>
<th>Dry matter content ratio</th>
<th>Grain yield</th>
<th>Relative maturity</th>
</tr>
</thead>
<tbody>
<tr>
<td>09COLE_IRR</td>
<td>0.61</td>
<td>0.57</td>
<td>0.40</td>
<td>0.56</td>
<td>0.86</td>
<td>0.89</td>
</tr>
<tr>
<td>09COLE_RFD</td>
<td>0.62</td>
<td>0.54</td>
<td>0.38</td>
<td>0.53</td>
<td>0.59</td>
<td>0.89</td>
</tr>
<tr>
<td>09YANA_IRR</td>
<td>0.79</td>
<td>0.56</td>
<td>0.34</td>
<td>0.53</td>
<td>0.61</td>
<td>0.85</td>
</tr>
<tr>
<td>09YANA_RFD</td>
<td>0.74</td>
<td>0.47</td>
<td>0.27</td>
<td>0.58</td>
<td>0.64</td>
<td>0.84</td>
</tr>
<tr>
<td>10COLE_IRR</td>
<td>0.62</td>
<td>0.51</td>
<td>0.31</td>
<td>0.53</td>
<td>0.59</td>
<td>0.88</td>
</tr>
<tr>
<td>10COLE_RFD</td>
<td>0.43</td>
<td>0.49</td>
<td>0.35</td>
<td>0.49</td>
<td>0.67</td>
<td>0.87</td>
</tr>
<tr>
<td>10YANA_IRR</td>
<td>0.63</td>
<td>0.49</td>
<td>0.27</td>
<td>0.50</td>
<td>0.69</td>
<td>0.91</td>
</tr>
<tr>
<td>10YANA_RFD</td>
<td>0.75</td>
<td>0.45</td>
<td>0.27</td>
<td>0.53</td>
<td>0.78</td>
<td>0.88</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>0.65</td>
<td>0.51</td>
<td>0.32</td>
<td>0.53</td>
<td>0.69</td>
<td>0.88</td>
</tr>
</tbody>
</table>
Table 4.5: Estimated genetic gain (EGG, equivalent to response to selection) for WSCC at 10% and 25% (EGG 10 and EGG25) selection intensities. The correlated response to selection (CR) of dry matter content ratio on WSCC, and WSCC for selection on anthesis biomass, grain yield and maturity is given at 10% and 25% (CR10 and CR25) selection intensities, with the relative selection efficiency (RSE) for each correlated response.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>EGG WSCC (mg/g)</th>
<th>CR dry matter content ratio on WSCC (mg/g WSCC)</th>
<th>CR WSCC on anthesis biomass (g/m² biomass)</th>
<th>CR WSCC on grain yield (t/ha. grain yield)</th>
<th>CR WSCC on relative maturity (Zadoks score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>09COLE_IRR</td>
<td>38.00</td>
<td>27.52</td>
<td>29.65</td>
<td>21.47</td>
<td>0.56</td>
</tr>
<tr>
<td>09COLE_RFD</td>
<td>34.22</td>
<td>24.79</td>
<td>-31.34</td>
<td>-22.70</td>
<td>-0.92</td>
</tr>
<tr>
<td>09YANA_IRR</td>
<td>36.75</td>
<td>26.62</td>
<td>27.67</td>
<td>20.04</td>
<td>0.75</td>
</tr>
<tr>
<td>09YANA_RFD</td>
<td>34.55</td>
<td>25.02</td>
<td>-18.09</td>
<td>-13.10</td>
<td>-0.52</td>
</tr>
<tr>
<td>10COLE_IRR</td>
<td>39.52</td>
<td>28.62</td>
<td>28.17</td>
<td>20.40</td>
<td>0.71</td>
</tr>
<tr>
<td>10COLE_RFD</td>
<td>35.62</td>
<td>25.79</td>
<td>31.46</td>
<td>22.78</td>
<td>0.88</td>
</tr>
<tr>
<td>10YANA_IRR</td>
<td>35.21</td>
<td>25.50</td>
<td>21.13</td>
<td>15.30</td>
<td>0.60</td>
</tr>
</tbody>
</table>
Table 4.6: Estimated genetic gain (EGG, equivalent to response to selection) for WSCA at 10% and 25% (EGG 10 and EGG25) selection intensities. The correlated response to selection (CR) of dry matter content ratio on WSCA, and WSCA for selection on grain yield and maturity is given at 10% and 25% (CR10 and CR25) selection intensities, with the relative selection efficiency (RSE) for each correlated response.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>EGG WSCA (g/m²)</th>
<th>CR dry matter content ratio on WSCA (g/m² WSCA)</th>
<th>CR WSCC on grain yield (t/ha. grain yield)</th>
<th>CR WSCC on relative maturity (Zadoks score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>09COLE_IRR</td>
<td>86.63</td>
<td>62.74</td>
<td>80.39</td>
<td>58.22</td>
</tr>
<tr>
<td>09COLE_RFD</td>
<td>57.56</td>
<td>41.69</td>
<td>-36.37</td>
<td>-26.34</td>
</tr>
<tr>
<td>09YANA_IRR</td>
<td>73.30</td>
<td>53.08</td>
<td>61.96</td>
<td>44.87</td>
</tr>
<tr>
<td>09YANA_RFD</td>
<td>37.73</td>
<td>27.32</td>
<td>-22.31</td>
<td>-16.16</td>
</tr>
<tr>
<td>10COLE_IRR</td>
<td>66.08</td>
<td>47.86</td>
<td>47.60</td>
<td>34.47</td>
</tr>
<tr>
<td>10COLE_RFD</td>
<td>59.67</td>
<td>43.22</td>
<td>51.11</td>
<td>37.02</td>
</tr>
<tr>
<td>10YANA_IRR</td>
<td>66.26</td>
<td>47.99</td>
<td>50.26</td>
<td>36.40</td>
</tr>
<tr>
<td>10YANA_RFD</td>
<td>57.13</td>
<td>41.38</td>
<td>56.08</td>
<td>40.62</td>
</tr>
</tbody>
</table>
4.4.5 Variance partitioning in water-soluble carbohydrate concentration and total water-soluble carbohydrate basic models

Table 4.7 shows the proportions of variance attributable to genotype and G × E components for WSCC and WSCA compared to maturity and grain yield from a reduced mixed linear model with simplified variance components that are nested within the full FA models. For WSCC genotype effects of 14.1% compared to the relative G × E which ranged from 0% for Genotype × Location to 6.9% for Genotype × Irrigation treatment. For WSCA, the largest G × E component was the interaction term of Genotype × Location × Irrigation treatment × Year. Maturity showed little G × E relative to the large genetic variance. Grain yield also showed relatively less G × E than WSCC and WSCA. Residual error variance for all traits except maturity was high.
Table 4.7: Variance components for WSC concentration, total WSC amount, maturity and grain yield from a simplified mixed linear model showing variance partitioning between genotype (G), and the environment components site location (L), irrigation treatment (T) and year (Y).

<table>
<thead>
<tr>
<th>Variance (%)</th>
<th>WSCC</th>
<th>WSCA</th>
<th>Relative maturity</th>
<th>Grain yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype</td>
<td>14.1</td>
<td>11.7</td>
<td>76.8</td>
<td>22.1</td>
</tr>
<tr>
<td>G x L</td>
<td>0</td>
<td>0</td>
<td>1.0</td>
<td>2.8</td>
</tr>
<tr>
<td>G x T</td>
<td>6.9</td>
<td>3.7</td>
<td>0</td>
<td>2.7</td>
</tr>
<tr>
<td>G x Y</td>
<td>6.1</td>
<td>2.1</td>
<td>5.0</td>
<td>3.7</td>
</tr>
<tr>
<td>G x L x T x Y</td>
<td>4.8</td>
<td>7.9</td>
<td>2.0</td>
<td>0.5</td>
</tr>
<tr>
<td>Residual Error</td>
<td>68.1</td>
<td>74.6</td>
<td>17.2</td>
<td>68.2</td>
</tr>
</tbody>
</table>
4.4.6 Coefficient of determination between the well-watered and water deficit environment cluster predicted values

Figure 4.20 shows WSCC E-BLUPs from the well-watered environment cluster plotted against the water deficit environment cluster. Overall, the coefficient of determination between these two clusters is $r^2 = 0.05$ (shown also in Figure 4.13). However, when lines are stratified on maturity, here into winter wheats, facultative, longer spring wheats and shorter spring wheats, and short season lines, then significant relationships between the two clusters emerge. The two largest groups of long spring and early spring have $r^2 = 0.20$ and $r^2 = 0.40$ between the two environment clusters respectively. Figure 4.21 illustrates the same data for WSCA E-BLUPs, albeit with only 09COLE_RFD E-BLUPs making up the water deficit environment. The overall coefficient of determination was $r^2 = 0.35$, compared to $r^2=0.33$ for long spring and $r^2 = 0.46$ for early spring groups.
Figure 4.20 WSC Concentration E-BLUPs plotted by environmental group. Genotypes are coloured into maturity groups using maturity E-BLUPs to illustrate correlation between environment groups within groups of genotypes with comparable maturity.
Figure 4.21 WSC amount E-BLUPs plotted by environmental group. Genotypes are coloured into maturity groups using relative maturity E-BLUPs. The water deficit environment consists of E-BLUPs from 09COLE_RFD which was the only water deficit experiment in the across-experiments FA model.
4.5 Discussion

4.5.1 The target population of environments for water-soluble carbohydrate selection and genotype × environment interactions

The clustering of experiments in this study show two clear TPE forWSCC and WSCA – one of well-watered environment experiments and the other of water deficit environment experiments. Genetic correlations within the well-watered environment cluster are strong and consistent for both WSCC and WSCA. The water deficit environment cluster consisted of only two sites for WSCC and one site for WSCA, although correlation between the two sites for WSCC was high ($r_{c}=0.87$). This provides evidence that WSC measurements are repeatable across experiments within a TPE, particularly for WSCC. The large G×E interactions observed for WSCC and WSCA compare with almost no G×E for maturity, and a more complex pattern of clustering for grain yield. The clear clusters are consistent with previous studies showing genotype rankings maintained across experiments for WSC traits.

Characterisation of G×E interactions using the variance components from the simplified models shown in Table 4.7 showed that irrigation treatment explained more variance for WSCC and WSCA than grain yield, while for grain yield year was the most important environment variance component. Further investigation of environmental covariances, such as the approach of Bouffier et al. (2015) to identify important environmental parameters, may be required to understand the environmental complexities beyond water availability.

The well-watered experiments were notable for their high biomass and high grain yields, reflecting growing conditions that facilitate the maximum performance of genotypes. Table 4.1 showed that average yields in the well-watered experiments were comparable with potential commercial irrigated yields in southern NSW and northern
Victoria (Toohey & Chaffey, 2006). In contrast, the water deficit experiments produced yields that are on average two-thirds lower than the well-watered experiments at the same site and year. Selection under water deficit compared to well-watered conditions has been a strategy considered for water productivity traits for some time (Cooper et al., 1995; Rebetzke et al., 2009). Clearly the success of correlated selection depends on the genetic correlation between environments, and in this study it is evident that the absence of genetic correlation between water deficit and well-watered environment clusters means that selection in in one environment cluster is not going to translate to performance in the other environment cluster. The G × E interactions illustrate why indirect selection for greater WSC may have not occurred with selection for grain yield in south-eastern Australia as it did in Western Australian and CIMMYT breeding programs. If terminal droughts were not encountered each year, then the selection environment would be inconsistent and selection pressure for higher WSC not as strong.

4.5.2 Selection for water-soluble carbohydrate concentration and total water-soluble carbohydrates in well-watered and water deficit target populations of environments

For WSCC, moderate broad sense heritability (comparable to the heritability for grain yield) in this study (Table 4.4) and the genetic correlation between experiments within either the well-watered or water deficit environment cluster indicate potential for increase via selection. These heritability estimates are lower than those reported in the diversity panel study of Ruuska et al. (2006) and the biparental mapping study of Rebetzke et al. (2008), however it is important to note that these heritability estimates are more relevant to improvement of this trait given they relate to the germplasm that would be used in breeding and selection. The heritability for WSCA was similar, but
lower due to the low heritability of anthesis biomass measurement. Unsurprisingly, estimated genetic gain (EGG) for WSCC was higher in the well-watered environments than the water deficit environments, a phenomenon also observed in wheat by Chapman (2008). Most complex traits have improved broad sense heritability under well-watered conditions, as genotypes are able to perform closer to their potential (Cooper et al., 1995), and WSCC is no different, as previously shown by Ruuska et al. (2006). However the suitability of selection under well-watered conditions for performance under water deficit conditions depends on the correlation in performance between the two environment types. This has been demonstrated successfully for a number of other complex traits, such as transpiration efficiency and early vigour (Rebetzke et al., 2009). In this study, there is no genetic correlation between water deficit and well-watered environment clusters, although Figure 4.20 shows that the lack of genetic correlation between the environment clusters can be partially explained by differences in maturity inferred in part by sampling differences between the well-watered and water deficit experiments. Samples were taken at the same day-degrees for each experiment, nevertheless water deficit phenology was further advanced than well-watered treatment. Within groups of lines that flower at a similar time, there was a positive correlation between performances in the two environments. This indicates that selection in well-watered environments within maturity groups may be correlated with selection for performance in water deficit environments. Accounting for such maturity differences may explain some of the variability in performance for WSCC seen between studies and between treatments. This implies precision should be improved by sampling strictly according to phenology as was done by Ruuska et al. (2006) and Rebetzke et al. (2008). This is supported Rattey et al. (2009) and McIntyre et
al. (2010) who report rankings forWSCC across environments were observed to be stable, and the genotypes were especially selected to be a similar maturity.

4.5.3 Using dry matter content as an indirect method of selection for water-soluble carbohydrate accumulation

This study investigated the potential of using the ratio of fresh weight to dry weight (dry matter content ratio) at anthesis as an indirect means to select for WSC, as a strong correlation between these traits has been previously identified by Ehdaie et al. (2008) and Xue et al. (2009), with the latter advocating that dry matter content was a useful indirect measure of WSC content in tillers. As observed in these previous studies, this study identified significant (p = 0.05) phenotypic coefficients of determination between dry matter content and both WSCC (up to $r^2 = 0.22$, Table 4.2) and WSCA (up to $r^2 = 0.29$, Table 4.3). However heritability for dry matter content was lower at every site for WSCC and comparable to heritability for WSCA (Table 4.4). This is reflected in the relative selection efficiency estimations of dry matter content on WSCC in water deficit experiments was -0.92 to -0.52 while the well-watered experiments was 0.56 to 0.88 (Table 4.5). Relative selection efficiency (RSE) estimations reflected the lower heritability for WSCA than WSCC, RSE for dry matter content on WSCA in the water deficit environments was -0.59 to -0.63 and for the well-watered environments ranged from 0.85 to 0.98 (Table 4.6). From the RSE estimations it should be noted that the selection was negative in the water deficit environments, and positive in the well-watered environments, suggesting that using dry matter content as a measure of WSC will require additional measures of crop development stage to inform breeders selection methods. These results show that the direction of selection will depend on what stage of WSC accumulation or remobilisation the crop is at when phenotyping is conducted.
An accurate indirect measure of WSC would be a useful tool for plant breeders looking to characterise their germplasm for this trait, given the difficulties posed in direct phenotyping. Measurement of WSC needs to be carefully timed relative to the crop growth stage, and requires access to substantial resources dedicated to sample processing and NIRS facilities if more than a small number of genotypes are to be screened efficiently. As well as dry matter content, other alternative methods of selecting for WSC should be further investigated, such as the suggestion by Blum (1998) that chemical desiccants be used to defoliate field experiments, to facilitate a measure of carbohydrate storage and remobilisation capacity.

4.5.4 Correlated response of grain yield and maturity from selection for water-soluble carbohydrates

In well-watered environment clusters there was no significant correlation between WSCC and grain yield, despite the range of variation for both traits (Table 4.2 and Table 4.3). In the water deficit environments, a significant correlation with yield (p < 0.05) was identified, but was weakly positive for WSCA and weakly negative for WSCC ($r^2 = 0.02$). It therefore follows on that the correlated response to selection of WSCC and WSCA on yield is negligible and inconsistent in this study. Other studies have shown relationships with yield under both high input well-watered environments (for example the UK, Snape et al. (2007)) as well as water deficit environments (Rebetzke et al., 2008). While a small number of lines were included in this study from WA and CIMMYT breeding programs that are known to favour selection for WSC, the majority of the genotypes have been selected for yield in south-eastern Australia target environments in a number of different breeding programs. The lack of relationship between yield and WSCC or WSCA could possibly be related to the
lack of relationship between the year of release of new varieties and variety WSCC content for eastern Australian breeding programs observed in Rebetzke et al. (2009).

While the correlation with WSCC and WSCA was strong, it is interesting that correlated response to selection estimations within the genotypes sampled for WSC in this study were small. At selection intensities of 10% a correlated response of around two on the Zadoks scale for both WSCA and WSCC is anticipated, translating to approximately one day difference in flowering time. Figure 4.20 plots WSCC E-BLUPs coloured by maturity group, indicating that despite an overall lack of correlation between well-watered and water deficit environment clusters, within maturity groupings there appears to be a more consistent relationship between the substantially different environment types.

4.5.5 Correlation of water-soluble carbohydrate concentration with other component traits

Phenotypic correlations with traits other than grain yield were weaker in this study relative to other comparable studies. Broadly, this maybe because there was a much larger number of genotypes tested than in most other studies, and the genotypes used were potentially less genetically diverse than other studies, given they consisted of advanced breeding material and varieties that had all undergone selection for yield and agronomic performance. However the range in WSCC values was large, and comparable with other studies. This may be partially due to non-genetic variation, but also would be because these lines were unselected for WSCC. For agronomic traits, there is a clear correlation between WSCC and maturity, and in this study a contributing factor to this could be samples for WSCC measurement being collected at one time, rather than samples being taken relative to the development stage of each genotype, such as the methods of Ruuska et al. (2006) and Rebetzke et al. (2008). As
Figure 4.5 shows, the earliness induced by drought stress in water deficit environments means that these genotypes were sampled later in their growth cycle. The accumulation peak and decline as remobilisation increases appears to have been captured in the accumulation phase in the well-watered environments with a positive relationship between maturity and WSCC, and in the decline phase in the earlier maturing water deficit environments. It is therefore unsurprising that genotypes re-rank for WSCC between the two environment groups (Figure 4.13). However, the other agronomic traits that do show a relationship with WSCC have the same sign relationship between water deficit and well-watered environment clusters, although with weaker correlations in the water deficit environments. In general, genotypes higher for WSCC seem to be higher in both water deficit and well-watered environments when maturity is taken into account (Figure 4.20). Dreccer et al. (2009) showed that genotypes maintained their WSC ranking when sampled for WSCC prior to and after anthesis. This study provides similar support for genotypes sampled in different environment clusters providing relative maturity of the different environments are taken into consideration.

Additionally, no relationship was observed between mature plant height and WSCC. As previously demonstrated in the work of Rattey et al. (2009), all of the genotypes in this study were semi-dwarf, consistent with previous studies, where relationships between height and WSCC are observed only if tall and semi-dwarf genotypes are contrasted, such as Ehdaie et al. (2008) and Rebetzke et al. (2008). As Blum (1998) suggests, the relationship between height and WSCC has been mainly attributed to the increased storage capacity of the taller lines.

In grain quality traits, there appeared to be no relationship between grain protein and WSCC. In contrast to the strong negative correlation between nitrogen
uptake and WSC accumulation indicated by a number of studies (Gebbing & Schnyder, 1999; Rebetzke et al., 2008; Ruuska et al., 2008; McIntyre et al., 2011) grain protein does not appear to have been previously considered. The 2010 rainfed experiments at both locations had higher WSCC than the irrigated experiments. The only difference between them would be nitrogen availability, as the high rainfall meant that both treatments had full moisture profiles for the growing season. This may show some effect of lower nitrogen uptake leading to higher WSCC.

The interrelated traits of thousand grain weight and screenings show a moderate relationship with WSCC in both clusters of environments, with higher WSCC correlated to higher thousand grain weight and lower screenings. This relationship is widely reported in other studies, and appears to be durable across environments and genotypes.

While weaker than most other studies, overall these correlations support previous reports of trait correlations with WSCC. Interestingly the correlations reduced in water deficit compared to well-watered environment clusters for all traits except for grains per m². These results also show the interrelationship between reduced tiller number, increased WSCC, and higher grain weight reported elsewhere (Rattey et al., 2009; Dreccer et al., 2013). There is some evidence that an ideotype with reduced tiller numbers and increased capacity for accumulation and lack of other post anthesis sinks for WSCC, may be the basis of the relationship between WSCC and performance under water deficit conditions observed in other studies. It follows that fewer grains per m² to fill may be due to fewer tillers, with the resulting heavier, plumper grain indicated by higher thousand grain weight and lower screenings due to a larger pool of assimilate (from WSC) available to fill a smaller number of florets. Alternatively, a plant with fewer tillers will have less photosynthetic area for transpiration and use less water in
the vegetative growth phase, leaving more soil water for grain filling under terminal drought. This plant type may be advantageous under consistent terminal drought conditions however the pathways to yield are plastic, so this combination of traits is not the only way to increased water use efficiency. South eastern Australia does not experience consistent terminal drought every season, so this combination of traits would probably not be selected for by default in the target population of environments for this region. Notwithstanding, it appears that WSCC can be selected for independently to tiller number and thousand grain weight. It is not clear from this study if maturity is related to the proposed ideotype, as this study sampled all genotypes in an experiment at one time, confounding the effect of differences in maturity with differences in the accumulation or remobilisation phases of the carbohydrate accumulation cycle.

4.5.6 Potential to utilise water-soluble carbohydrate accumulation as a drought avoidance trait

The results of this study provide evidence of the potential to select for both WSCA and WSCC. However, the results also illustrate the challenges involved in characterising this complex trait. Phenotyping is intensive and not easily mechanised, although the use of NIRS is a substantial reduction in time and complexity compared to previous wet chemistry methods of determination. Augmenting conventional selection approaches with the application of marker-assisted selection may be a useful avenue for improvement of WSC in south-eastern Australia, especially if the TPE for selection are not consistent from year to year.

4.6 Conclusions

This study revealed strong G × E interactions for both WSCC and WSCA across environments differing in water availability. Within these environments, the genetic
correlation between experiments was large, reflecting clear groupings of TPE. In both TPE, heritability for WSCC was higher than WSCA, reflecting the lower heritability of anthesis biomass sampling. Because heritability for WSC is low relative to other traits like grain yield, greater genetic gain could be realised selecting on WSCC rather than WSCA. The lower heritability of WSCC and WSCA observed in this study may be due in part to the use of elite germplasm, which also means the heritability estimates are more relevant to trait improvement. Using indirect methods of selection for WSC such as dry matter content ratio is worthy of further consideration, however dry matter content heritability was lower than WSCC so other surrogate measures of WSC should be investigated as well.

The little or no correlation of WSCC and WSCA with grain yield under well-watered or water deficit conditions is in contrast with other studies reporting substantial correlations between these traits. This may reflect the nature of the genotypes in this study, which as a cohort of advanced breeding lines may not have included the range of plant types present in other studies. Despite weaker correlations than other studies, relationships with other traits are notable, especially with lower tillering, grains per m², and higher thousand grain weight. Molecular markers for WSCC may be a useful tool for breeding programs seeking to incorporate this trait into breeding germplasm, and may ameliorate some of the difficulties involved in large-scale phenotyping and selection of WSCC.
5. Genome-Wide Association Studies for Water-Soluble Carbohydrate Concentration

5.1 Introduction

The thorough investigation of G × E and related traits in Chapter 3 showed that WSCC is a very complex character to select, although moderate broad-sense heritability indicates useful genetic variance that could be exploited for genetic gain. Identifying QTL through marker-trait associations may assist genetic improvement for WSCC by identifying molecular markers that explain a significant amount of the genetic variation for WSCC. Molecular markers can increase the rate of genetic gain in a breeding program by enabling selection for favourable genotypes early in the breeding cycle, and genotyping could be more resource-effective than the complex phenotyping required to measure WSCC.

Previous QTL studies indicate the potential for identifying loci explaining a significant part of WSCC genetic variance that could be useful in marker-assisted selection based breeding strategies. A genome-wide association approach in a diverse set of breeding lines has not been used before to our knowledge, and may uncover new insights into the genetic architecture of this complex trait that would be of direct relevance for use of selection in wheat breeding programs in south-eastern Australia. GWAS are complex, with many factors affecting the outcome of marker-trait associations. These factors include marker type, marker number, analysis method, and the genetic structure (diversity and relatedness) of the population used in analysis. Given the nature of WSCC phenotype expression, where the accumulation and remobilisation cycle depends on timing of plant development and associated environmental conditions, covarying traits such as flowering time and G × E interactions should also be considered in an informative GWAS.
5.2 Aims

The methodology of GWAS can be used to explore the underlying genetic architecture of a complex trait as well as identify candidate markers. This study sought to identify candidate markers for WSC Concentration (WSCC) by GWAS using software and methods commonly available to plant breeders. Determining if a GWAS has identified biologically significant loci is notoriously difficult, as there are many factors in the analysis that have a bearing on the results (Attia et al., 2009). Here, we sought to identify candidate WSCC markers using a combination of two molecular marker sets (SNP and DArT markers) and two analysis pipelines (WGAIM and GAPIT). Each GWAS was conducted separately for each site of a MET, given the importance of conducting individual environment association analyses in the presence of significant G × E interaction (Oldmeadow et al., 2011; Zila et al., 2013). Additionally, we use these GWAS marker sets and methods to identify candidate markers for relative maturity around flowering time (ZAD) as a positive control process, to ascertain if the methods used can detect the well validated, biologically significant loci of the major flowering time genes.

5.3 Materials and methods

5.3.1 Genetic material and phenotyping for water-soluble carbohydrate concentration and relative maturity at flowering time

The experiments and genotypes used for phenotypic measurement in this study were detailed previously in Chapters 4.3.1 and 4.3.2. Also the phenotyping methods for measurement and statistical analysis of WSCC and ZAD are detailed in Chapter 4.3.6 and 4.3.7. For ZAD, phenotypic information from all genotypes was used, while for WSCC the subset detailed in Chapter 4.3.1 was used. For WSCC, the nature of G × E interaction effects indicating the presence of two main environment clusters for
experiments is described in Chapter 54. These environment clusters corresponded to well-watered environment and water deficit environment experiments, respectively.

5.3.2 Marker sets and statistical analysis of marker information

Two separate marker sets were used: 985 lines from the overall experiment were genotyped using the Illumina 9k Infinum® iSelect beadchip array (Cavanagh et al., 2013) resulting in 4,883 polymorphic SNPs across the population. Similarly, 955 lines were genotyped with Diversity Arrays technology (DArT) (Jaccoud et al., 2001; Akbari et al., 2006) resulting in 2,013 polymorphic markers across the population. Genotype information for SNP and DArT marker datasets were prepared separately for analysis using the R software package Synbreed (Wimmer et al., 2012). Imputation of missing values was done using the software package Beagle (Browning & Browning, 2009). This package imputes missing marker values based on haplotype block (linkage disequilibrium) information. For the SNP marker dataset, 537 duplicated markers and 88 monomorphic markers were removed, as well as 432 markers with minor allele frequency (MAF) of less than 5%. For the DArT marker set, 50 duplicated markers and 190 markers with MAF of less than 5% were removed. The resulting 3,826 SNP markers and 1,773 DArT markers were used to compute an IBD relationship matrix after Endelman and Jannink (2012), Habier et al. (2007b) and VanRaden (2008) for each marker dataset. In the SNP dataset, 12 clones and closely-related lines were removed to ensure that the relationship matrix was positive definite, meaning genotype information for 973 lines were included in the analysis. From the phenotypic models, a total of 365 genotypes had WSCC phenotype information and a total of 1,314 genotypes had ZAD phenotype information (Section 4.3.1). The SNP dataset had matching phenotype information on WSCC for 358 genotypes and ZAD for 802.
genotypes, while the DArT dataset had matching phenotype information on WSCC for 350 genotypes and ZAD for 773 genotypes. Only genotypes with both marker and phenotype information were able to be used in the analysis.

Consensus maps were used to provide physical positions for the marker information. For the DArT dataset this study used the Wheat Interpolated Maps (version 6) as a reference to locate the positions of DArT markers (Diversity Arrays Pty Limited pers. comm.) and for the SNP dataset the 9K Consensus Map (version 4) was used (Dr Matthew Hayden (Victoria DPI). pers. comm.).

### 5.3.3 Linkage disequilibrium analysis

The LD in the SNP and DArT marker sets was estimated using the methods of Breseghello and Sorrells (2006). Pairwise LD estimates ($r^2$) were calculated with the software package PLINK (Purcell et al., 2007) for unlinked loci pairs (loci pairs not on the same consensus chromosome) and for syntenic loci (loci pairs on the same chromosome) separately. Syntenic $r^2$ was plotted against pairwise marker distance for all chromosomes on each genome following Edae et al. (2014), with a second degree locally weighted polynomial regression (LOESS) curve fitted to each scatter plot (Cleveland, 1979). All of the unlinked $r^2$ estimates were square root transformed to approximate a normal distribution, and the parametric 95th percentile of that distribution was determined as a critical value of $r^2$ above which LD is likely to be caused by genetic linkage (Breseghello & Sorrells, 2006). The intersection of the LOESS curve and the $r^2$ critical value was taken as an estimate of the extent of LD decay within each genome following Laidò et al. (2014).
5.3.4 Genome-wide association study methods for GAPIT

Associations for bothWSCC and ZAD were undertaken separately for each marker dataset, and associations were performed using phenotype data from each experiment.

For the GAPIT GWAS pipeline, E-BLUPs generated following the methods in Chapter 4.3.7 were used as the phenotype information. Association studies forWSCC and ZAD were conducted using the compressed mixed linear model approach (Zhang et al., 2010) that accounts for multiple degrees of relatedness (population structure and cryptic relationships) implemented in the R software package Genome Association and Prediction Integrated Tool (GAPIT) (Lipka et al., 2012) according to

\[ \hat{y} = X\tau + Z_g u_a + \eta \]

Equation 5.1

where \( \hat{y} \) is the vector of phenotypic values predicted from a single-experiment version of Equation 4.1, \( \tau \) is a \((t \times 1)\) vector of fixed effects for the corresponding \( n \times t \) design matrix \((X)\), including a SNP or DArT marker, depending on the analysis, the \( m \times 1 \) vector of overall genetic line effects \( g \) (with associated design matrix \( Z_g \)) is partitioned into a vector of additive line effects \( u_a \), and the \( n \times 1 \) residual vector \( \eta \) represents local stationary variation. To account for inflated false positives (type I error) the false discovery rate (FDR) procedure was used following Benjamini and Hochberg (1995) with a nominal threshold of 10% given the stringencies of this test (Edae et al., 2014).

5.3.5 Genome-wide association study methods for WGAIM

Following the GAPIT procedure, associations for bothWSCC and ZAD were performed separately for each marker dataset, and associations were performed using phenotype data from each experiment.
For the WGAIM GWAS, whole genome average interval mapping (WGAIM) was implemented following the approach of Verbyla et al. (2007) using the R software package WGAIM (Taylor & Verbyla, 2011). The WGAIM method can use either an interval approach or a marker-by-marker approach, the details of which are described in Taylor (2013). This study used the latter approach, as interval mapping is not appropriate for GWAS using a consensus map. The WGAIM method is a one-step analysis pipeline so individual models for each experiment after Equation 4.1 were used for the GWAS. This linear mixed model for $m$ varieties ($i = 1, 2, 3, \ldots m$) is given by

$$y = X\tau + Z_g g + Z_u u + \eta$$

Equation 5.2

where $y$ is the ($n \times 1$) data vector of the response variable; $\tau$ is a ($t \times 1$) vector of fixed effects, including experiment main effects, designed based effects and other extraneous field variation with associated design matrix $X$. The term $Z_u u$ is a random component and associated design matrix and contains experiment-specific terms used to capture design based extraneous variation, such as the blocking structure of the field experiments (replicate and irrigation bay), and the interaction of experiment range and row effects. The residual error is $\eta$ and was assumed to have distribution $\eta \sim N(0, \sigma^2 R)$ where $\sigma^2$ is the residual variance for the experiment and $R$ is a matrix that for the field-only trait models contains a parameterization for a separable autoregressive AR1 $\otimes$ AR1 process to model potential correlation of the observations due to the neighbouring plots of the experiment, and $\otimes$ represents the Kronecker product. For the multiphase analysis of water soluble carbohydrate concentration and content, residual variance for each year was modelled.
The genetic effects vector $g$ from Equation 4.1 represent a vector of random effects with associated design matrix $Z_g$. The $i$th genetic component of the genetic effects model is given by

$$g_i = \sum_{k=1}^{c} \sum_{j=1}^{m_k-1} q_{i,k,j} a_{k,j} + p_i$$

Equation 5.3

where $m$ is the total number of markers, $c$ is the number of chromosomes, $m_k$ is the number of markers on chromosome $k$, $(k = 1, \ldots, c)$, and $q_{i,k,j}$ is the parental allele type for line $i$ in interval $j$ on chromosome $k$. $q_{i,k,j} = \pm 1$, reflecting two possible genotypes. Further, $a_{k,j}$ is QTL effect size assumed to have distribution $a_{k,j} \sim N(0, \sigma^2 \gamma_a)$ and $p_i \sim N(0, \sigma^2 \gamma_p)$ represents genetic effects not captured by the QTL effects. As the vector of QTL allele types are replaced by the expectation of the QTL genotype, the full vector of genotypic effects is

$$g = Ma_M + p$$

Equation 5.4

where the marker QTL sizes are $a_M = \Lambda a$. This suggests that the marker QTL sizes have an assumed distribution of the form $a_M \sim N(0, \sigma^2 \gamma_a \Lambda \Lambda^T)$ and are correlated. The full working model for GWAS is therefore written as

$$y = X\tau + Z_g Ma + Z_g p + Z_u u + \eta$$

Equation 5.5

Once the full model has been fitted to the data with and without the random regression QTL effects $Z_g Ma$, the log likelihood ratio test following Stram and Lee (1994) is applied to the hypothesis $H_0: \gamma_a = 0$. If $\gamma_a$ is significant, a putative QTL is determined based on the alternative outlier model described in Gogel (1997) and Verbyla et al. (2007). Once a marker has been identified as a QTL it is moved into the
fixed effects of the working model (Equation 5.5) and the process is repeated until $\gamma_a$ is not significant. After the selection process is complete the QTL markers appear as fixed effects and the final model is written as

$$y = X\tau + Z_g M_s a_s + Z_g M_{-s} a_{-s} + Z_g p + Z_u u + \eta$$

Equation 5.6

where $a_s$ are the identified QTL as fixed effects, and $a_{-s}$ are the unselected markers as a set of random effects. In WGAIM the false positives are controlled naturally by assuming a background level of QTL variation through a single variance component associated with a contiguous set of QTL across the whole genome.

### 5.4 Results

#### 5.4.1 Heritability, population structure and linkage disequilibrium

For ZAD, no G x E was observed across the experiments. Line mean heritability at each experiment for ZAD was high, ranging from $H^2 = 0.84$ to 0.91, compared to heritability estimations for WSCC, which ranged from $H^2 = 0.43$ to 0.79 (Table 4.4). Phenotype values were predicted for each experiment from the models described in Chapter 4.3.7 and these estimated best linear unbiased predictors, or E-BLUPs, were used in the GAPIT GWAS models. The single experiment model described in Equation 4.1 and the experiment raw phenotypic data was used for the WGAIM GWAS models.

Heatmaps of the IBD relationship matrices calculated from the SNP and DArT marker datasets are shown in Figure 5.1 and Figure 5.2 respectively. For the SNP markers, genotyped lines fall largely into two clusters, corresponding generally to a grouping of Wagga Wagga and Horsham breeding program lines and high rainfall zone (HRZ) breeding program lines, with substantial crossover between clusters. The DArT marker set shows stronger population structure with three main clusters, corresponding to Wagga Wagga and Horsham breeding lines, HRZ breeding lines, and
a cluster containing genotypes from all breeding programs. For both marker types, CIMMYT-bred material grouped with the HRZ breeding lines clusters.

The plots of LD estimates in Figure 5.3 and Figure 5.4 show a decay of LD with genetic distance. The 97,593 DArT marker and 485,448 SNP pairwise comparisons for syntenic loci showed LD estimates were largely determined by their genetic distance. Pairwise estimates of syntenic LD ($r^2$) ranged from 0 to 0.999, with a median value of 0.0135 and mean of 0.0864 for the DArT marker set. A range of 0 to 0.999 with a median of 0.0204 and mean of 0.0104 was observed for the SNP marker set. The inter-chromosomal analysis of LD for unlinked markers consisted of 1,458,368 pairwise comparisons for the DArT marker set, and of 6,655,714 pairwise comparisons for the SNP marker set, and provided a critical value of $r^2= 0.0456$ for the DArT markers and $r^2= 0.0470$ for the SNP markers ($r^2$ values greater than these critical values indicate physical linkage). In turn, 26.76% of syntenic pairwise comparisons for DArT markers and 33.90% of SNP markers were greater than the respective critical values for each marker dataset. As described in section 5.3.3, the point at which the smoothing LOESS curve for LD estimates on each genome intersected the critical value for $r^2$ was determined to be the average decay of LD in this population. In the DArT marker set, this intersect was at approximately 18cM for the A genome, 16cM for the B genome and 25cM for the D genome. Correspondingly, the intersect for the SNP marker dataset occurred at 24cM for the A genome, and 21cM for both the B and D genomes.
Figure 5.1: Kinship relationship matrix heatmap and dendrogram output from GAPIT for the SNP data set.
Figure 5.2: Kinship relationship matrix heatmap and dendrogram output from GAPIT for the DArT data set.
Figure 5.3: Pairwise LD estimates ($r^2$) plotted against Euclidian pairwise marker distances for markers on the same consensus chromosome for the SNP marker set. Chromosomes are plotted for each genome, and a second degree LOESS curve fitted to each scatter plot. The dashed line indicates the critical value of $r^2$, above which marker pairs are assumed to be physically linked. For the SNP marker set the critical $r^2$ value ($r^2=0.0470$) intersects the LOESS curve at 24cM for the A genome and 21cM for both the B and D genomes.
Figure 5.4 Pairwise LD estimates ($r^2$) plotted against Euclidian pairwise marker distances for markers on the same consensus chromosome for the DArT marker set. Chromosomes are plotted for each genome, and a second degree LOESS curve fitted to each scatter plot. The dashed line indicates the critical value of $r^2$, above which marker pairs are assumed to be physically linked. The critical value of $r^2$ for the DArT marker set ($r^2 = 0.0456$) intersects the LOESS curve at approximately 18cM for the A genome, 16cM for the B genome and 25cM for the D genome.
5.4.2 Genome-wide association study results for relative maturity at flowering time using GAPIT

For the ZAD GAPIT GWAS using the DArT markers, only one marker was identified, and was detectable in multiple experiments (W9A10COLE_IRR, W9A10COLE_RFD and W9A10YANA_IRR) on chromosome 2D (Figure 5.7). The GAPIT GWAS for ZAD using the SNP marker set identified 18 unique markers across all experiments, with 17 of those markers identified at more than one experiment. Of these, 11 markers were located within 3cM of each other on the consensus map on chromosome 5A (Figure 5.24) and three markers collocated on chromosome 5B (Figure 5.25). Additionally a marker each was identified for loci on chromosomes 2D, 4B and 5D (Figure 5.18, Figure 5.22 and Figure 5.26). Four of the markers on 5A were recorded at all eight experiments, as was the marker identified on 2D. Association results are presented in detail in the Appendix (Table A.4).

5.4.3 Genome-wide association study results for water-soluble carbohydrate concentration using GAPIT

For the GAPIT GWAS for WSCC no markers were significant using the SNP marker set. For the DArT marker set, nine unique markers were detected, and of these seven markers were detected at both Coleambally and Yanco in 2009 rainfed experiments. Of these markers, two were located on chromosome 1A (Figure 5.5), one on 1B (Figure 5.5), three on 1D (Figure 5.6) and one on 2D (Figure 5.7). On the consensus map, the marker on 2D was located over 30cM away from the flowering locus identified in the ZAD GWAS described above. Association results are presented in detail in the Appendix (Table A.3).
5.4.4 Genome-wide association study results for relative maturity at flowering time using WGAIM

Many more markers were identified using WGAIM than the GAPIT approach. The WGAIM GWAS for ZAD using the SNP marker set identified a total of 112 SNP markers across all experiments associated with the trait. Of these, 55 were detected in more than one experiment. The WGAIM GWAS for the DArT marker set identified 99 unique DArT markers associated with ZAD across all experiments, and of these, 48 were discovered in more than one experiment (Table A.2). The SNP and DArT markers for ZAD using WGAIM were identified across all 21 chromosomes of the respective consensus maps. Association results are presented in detail in the Appendix (Table A.2).

5.4.5 Genome-wide association study results for water-soluble carbohydrate concentration using WGAIM

For the WGAIM WSCC GWAS, 36 unique marker-trait associations were identified in the eight individual experiment analyses in the SNP marker set. Of these, three markers were significant in more than one experiment on chromosomes 1A (Figure 5.13), 2B (Figure 5.17) and 5B (Figure 5.25). A total of 42 unique DArT marker trait associations were found across experiments, with nine markers detected at more than one site. Interestingly, SNP and DArT markers that were on the same chromosome according to the respective consensus maps did not appear to be close to each other. For example, on chromosome 1A, the significant SNP marker is located on the short arm, while both significant DArT markers are on the long arm (Figure 5.5 and Figure 5.13). Similarly both DArT markers on chromosome 2B are on the short arm, while the significant SNP for chromosome 2B is located on the long arm (Figure 5.6 and Figure 5.17). One DArT marker maps to the short arm of chromosome 3B while the map location of the SNP on the same chromosome is on the long arm (Figure 5.8 and
In Figure 5.20. Of the DArT and SNP markers identified as significant for WSCC, SNP marker \textit{wsnp}\textunderscore\textit{Ex}\textunderscore\textit{c6420\_11162653} on chromosome 1A (Figure 5.13) is also significant for flowering time, as was DArT marker \textit{wPt\_8404} on chromosome 2B (Figure 5.6). On chromosome 3B, DArT marker \textit{wPt\_1158} (significant for ZAD) mapped within 1cM of DArT marker \textit{wPt\_4209} which was significant for WSCC (Figure 5.8), and SNP marker \textit{wsnp\_CAP11\_c948\_571287} (significant for WSCC) on 5B mapped within 1cM of SNP marker \textit{wsnp\_BE495277B\_Ta\_2\_5} which was significant for ZAD (Figure 5.25).

Association results are presented in detail in the Appendix (Table A.1).

### 5.4.6 Markers for water-soluble carbohydrate concentration and relative maturity common between both GAPIT and WGAIM analyses

Here, this study details only the loci that were identified at more than one of the individual experiment GWAS for each trait, and were detected using both GAPIT and WGAIM GWAS methods. The ZAD GWAS identified four loci on the consensus maps in both the GAPIT and WGAIM analysis pipelines with the following markers: \textit{wsnp\_CAP12\_c812\_428290} on 2D (Figure 5.18), the co-locating markers \textit{wsnp\_BE422566B\_Ta\_1\_1} and \textit{wsnp\_BE422566B\_Ta\_1\_2} on 4B (Figure 5.22), \textit{wsnp\_Ex\textunderscore\textit{c31799\_40545376}} on 5A and \textit{wsnp\_Ex\textunderscore\textit{c508\_1008029}} on 5D (Figure 5.26). For WSCC, two DArT markers were common to both GAPIT and WGAIM analyses, \textit{wPt\_9592} on chromosome 1A (Figure 5.5) and \textit{wPt\_3743} on chromosome 1D (Figure 5.6). The discussion will focus on these markers detected in both analysis pipelines in an analogous approach to Neumann et al. (2011) and Azadi et al. (2015) who also report only marker trait associations common to two modelling methods.
Figure 5.5 Chromosome maps showing DArT markers for consensus chromosomes 1A and 1B. Significant loci forWSCC and ZAD at each experiment are shown for both WGAIM (blue) and GAPIT (green) analyses.
Figure 5.6: Chromosome maps showing DArT markers for consensus chromosomes 1D, 2A and 2B. Significant loci for WSCC and ZAD at each experiment are shown for both WGAIM (blue) and GAPIT (green) analyses.
Figure 5.7 Chromosome maps showing DArT markers for consensus chromosomes 2D and 3A. Significant loci forWSCC and ZAD at each experiment are shown for both WGAIM (blue) and GAPIT (green) analyses.
Figure 5.8 Chromosome maps showing DArT markers for consensus chromosomes 3B, 3D and 4A. Significant loci for WSCC and ZAD at each experiment are shown for both WGAIM (blue) and GAPIT (green) analyses.
Figure 5.9 Chromosome maps showing DArT markers for consensus chromosomes 4B, 4D and 5A. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis.
Figure 5.10 Chromosome maps showing DArT markers for consensus chromosomes 5B and 5D. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis.
Figure 5.11 Chromosome maps showing DArT markers for consensus chromosomes 6A, 6B and 6D. Significant loci for ZAD at each experiment are shown for the WGAim (blue) analysis.
Figure 5.12 Chromosome maps showing DArT markers for consensus chromosomes 7A, 7B and 7D. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis.
Figure 5.13 Chromosome maps showing SNP markers for consensus WGAfM (blue) analysis.

<table>
<thead>
<tr>
<th>Chromosome 1A</th>
<th>Significant loci for ZAD</th>
<th>Each experiment are shown for the</th>
</tr>
</thead>
<tbody>
<tr>
<td>ws_Ex_c1255_2411550</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_c1255_2411977177.3</td>
<td>ZAD(Y10_Rf)</td>
<td></td>
</tr>
<tr>
<td>ws_Ku_c40759_48907151</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ku_rep_c71113_70810693</td>
<td>ZAD(C09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ra_rep_c97840_85062318</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ra_rep_c97840_85062404</td>
<td>ZAD(C10_Rf)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_c11939_19147514</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_c1557_2973321</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_c1861_3506202</td>
<td>ZAD(Y10_Rf)</td>
<td></td>
</tr>
<tr>
<td>ws_Ku_c12441_20092792</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ku_c26281_36243489</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_c48407_53323483</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_c15377_23637176</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_rep_c103474_88421290</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_c271_52142916</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_c9343_15514687</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_c38849_46284348</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ku_c816_16843541</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_c1970_3707739</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_c7684_13109335</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ra_c19637_28840370</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ku_rep_c109724_94227136</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_CAP11_c146_160903</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_CAP12_rep_c8598_3608465</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_CAP7_rep_c5385_2417464</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_CAP7_rep_c5385_2417467</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_c10903_17716723</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_c19353_28290393</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ra_c12759_20345987</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_rep_c68171_66944702</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_c1359_26042981</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_rep_c102067_87313597</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_c1427_2736441</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_c2273_4259708</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_c2814_5203467</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_c2840_5247386</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_c31983_40710608</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_c7965_13519586</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ku_rep_c101175_88380491</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_BF200531A_Ta_2_127.5</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_c10783_17554146</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_Ex_c2868_52934852</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
<tr>
<td>ws_RFL_Contig4735_56739990</td>
<td>ZAD(Y09_Irr)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 5.14 Chromosome maps showing SNP markers for consensus chromosome 1B. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis.
Figure 5.15 Chromosome maps showing SNP markers for consensus chromosome 1D. No markers on this chromosome were associated with either ZAD or WSCC at multiple experiments.
**Figure 5.16** Chromosome maps showing SNP markers for consensus WGAIM (blue) analysis. Significant loci for ZAD at each experiment are shown for each chromosome 2A. Significant loci for ZAD at each experiment are shown for each chromosome 2A.
Figure 5.17 Chromosome maps showing SNP markers for consensus chromosome 2B. Significant loci for WSCC and ZAD at each experiment are shown for the WGAIM (blue) analysis.
Figure 5.18 Chromosome maps showing SNP markers for consensus chromosome 2D. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis.
Figure 5.19 Chromosome maps showing SNP markers for consensus chromosome 3A. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis.
Figure 3B. Chromosome maps showing SNP markers for consensus WGA-M (blue) analysis. Significant loci for ZAD at each experiment are shown for the chromosome 3B.
Figure 5.21 Chromosome maps showing SNP markers for consensus chromosome 3D and 4A. No markers on this chromosome were associated with either ZAD or WSCC at multiple experiments.
Figure 5.22 Chromosome maps showing SNP markers for consensus chromosome 4B. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis.
Figure 5.23 Chromosome maps showing SNP markers for consensus chromosome 4D. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis.
Figure 5.24 Chromosome maps showing SNP markers for consensus chromosome 5A. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis.
Figure 5.25 Chromosome maps showing SNP markers for consensus chromosome 5B. Significant loci for WSCC and ZAD at each experiment are shown for the WGAIM (blue) analysis.
Figure 5.26 Chromosome maps showing SNP markers for consensus chromosome 5D. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis.
Figure 5.27 Chromosome maps showing SNP markers for consensus chromosome 6A. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis.
Figure 5.28 Chromosome maps showing SNP markers for consensus loci for ZAD at each experiment are shown for chromosome 6B and 6D. Significant loci for ZAD at each experiment are shown for the WCAIM (blue) analysis.
Figure 5.29. Chromosome maps showing SNP markers for consensus chromosome 7A. No markers on this chromosome were associated with either ZAD or WSCC at multiple experiments.
Figure 5.30 Chromosome maps showing SNP markers for consensus chromosome 7B. Significant loci for ZAD at each experiment are shown for the WGAIM (blue) analysis.
Figure 5.31 Chromosome maps showing SNP markers for consensus chromosome 7D. No markers on this chromosome were associated with either ZAD or WSCC at multiple experiments.
5.5 Discussion

5.5.1 Comparison of analysis at individual experiments

The GWAS results for both ZAD and WSCC show that significant associations can be experiment specific, and relatable to the G × E relationships between experiments for each trait. For ZAD, where genetic correlations between all experiments were very high, ranging from $r_c = 0.92$ to $0.99$ (Figure 4.11) significant loci were detected in all experiments. For WSCC, environment clustering analysis in (Figure 4.13) showed two distinct environments corresponding to well-watered and water-deficit environments. The significant associations for WSCC were detected in the two experiments in the water deficit environment cluster, and no significant loci were detected in the well-watered environment cluster of experiments.

Differences in individual experiment GWAS results for WSCC could be related to sampling differences between the experiments. Given both rainfed and irrigated experiments were sampled at the same time, and rainfed experiments were on average at a later development stage than the irrigated experiments, the WSCC sampling would have been conducted at a later point in the WSCC accumulation and remobilisation cycle in the rainfed experiments. Average Zadoks score in the well-watered environments cluster of experiments was 55.04 compared to 57.22 for the water deficit environments, corresponding to approximately one day later flowering in the well-watered experiments (Figure 4.5 and Table 4.1). However, genotypes with higher peak WSCC are known to maintain greater WSCC levels for longer (McIntyre et al., 2011) so the sampling method may still capture the higher-accumulating WSCC genotypes.
Combinations of well-watered and water deficit environments have been used for WSCC QTL detection previously, and in some studies such as Yang et al. (2007) and Pinto et al. (2010), QTL for WSCC were detected in well-watered, or water deficit environments but not in both. This illustrates the importance of environmental characterisation in QTL analysis, and the value in understanding the TPE each QTL analysis is performed in. Once this has been undertaken, experiment-by-experiment GWAS should allow a means to replicate QTL detection in comparable environments, and as both Oldmeadow et al. (2011) and Zila et al. (2013) indicate, understand possible QTL × environment interactions.

5.5.2 Comparison of WGAIM and GAPIT analysis pipelines

The range of GWAS software pipelines available to plant breeders illustrates the advances made in QTL analysis, with most applying somewhat different approaches to marker trait association. It is beyond the scope of this study to provide more than cursory comment on the relative usefulness of the two options tested, although the difference in results suggests that the effectiveness of the different models deserves further attention by plant breeders seeking to use them.

WGAIM in this study was used as an alternative to GAPIT which is the R package equivalent to the popular software TASSEL (Lipka et al., 2012). Additionally, WGAIM was developed for use in biparental and backcross populations (Verbyla et al., 2007). However, the ability to use a ‘marker’ rather than ‘interval’ approach in the software means that a GWAS marker analysis can be performed following methods set out in Taylor (2013). The one step model that encompasses extraneous field and environmental information provides a contrast to the standard approach of GAPIT where associations are modelled on the predicted phenotype from a separate model.
used to account for extraneous phenotypic variation. It is difficult to say if one approach is better than the other given this study does not use a simulation approach but instead real data with ZAD, a trait where many significant loci are known and have been utilised by plant breeders over many years (Pugsley, 1983). For ZAD, WGAIM identified 38 DArT marker loci and 37 SNP marker loci, while GAPIT identified 1 DArT marker and 5 SNP marker loci. Possibly, false discovery rates are higher in WGAIM (Verbyla et al., 2007), and overly stringent in GAPIT (Edae et al., 2014), but this requires a different approach to quantify.

Here, we reiterate the conclusions of Oldmeadow et al. (2011) and suggest that agreement between results of multiple analysis approaches can be used as a way to identify signal from noise in statistical associations. We use the presence of the same loci in the results of both GAPIT and WGAIM as a measure of the robustness of the association, as loci that have a strong enough association across the different software approaches and across environments may be less likely to be artefacts of the methods used, even though the extra stringency of the GAPIT procedure may have biased results towards this method.

### 5.5.3 Comparison of DArT and SNP molecular marker sets

Molecular marker types are rarely contrasted in association studies, which is notable given there are fundamental genomic differences between the most common marker types. For the marker types used in this study, SNP variation is more likely to be functional (Hamblin et al., 2007), while DArT markers capture an additional degree of genetic information compared to SNPs, as they are concentrated in the hypermethylated parts of the genome (pers. comm. Dr Andrzej Kilian, Diversity Arrays Pty. Ltd.) thereby providing both genetic and epigenetic information. It is also
possible for DArT markers to be located in insertion/deletion sites (indels) although approximately 80% are SNPs (Kilian et al., 2003). Early association studies were constrained by limited numbers of markers and the high costs of genotyping, so comparisons between marker types would be more likely to be distorted by LD differences in genome coverage rather than the markers themselves (Flint-Garcia et al., 2003). As marker types have increased and the cost of genotyping decreased, LD has in theory been accounted for by an increasing genome wide coverage of markers. In this study, there are more than twice as many SNP markers compared to DArT markers, and both Figure 5.3 and Figure 5.4 show together that LD measured across the genome varies between SNPs and DArT markers. Across the A and B genomes, DArT markers have 25% and 24% shorter linkage blocks, and approximately 119% larger linkage blocks in the D genome. The longer linkage blocks in the D genome for both SNP and DArT markers compared to the A and B genome is consistent with previous reports (Wang et al., 2014). Figure 5.5 through to Figure 5.31 provide a graphical visualisation of the marker coverage, and highlights the limited coverage of the D genome by both DArT markers and SNP markers. This illustrates that the marker sets used are different, providing justification for the separate analysis of the different marker sets. This is in contrast to similar recent studies which have used genetic maps developed for a combination of marker types (Wenzl et al., 2006; Crossa et al., 2007; McIntyre et al., 2010; Cui et al., 2014b). Given LD underlies the ability to perform statistical associations between trait and genotype, it is interesting to note that the loci detected across experiments and analysis methods for ZAD were only found in the SNP marker set, while loci for WSCC meeting the same criteria were only found in the DArT marker set.
Marker differences are also evident in the genetic structure results in this study. The IBD relationship matrix between lines shows that the genotypes used group into two main clusters for the SNP marker set (Figure 5.1), which contrast with the DArT marker set which produced three main clusters of genotypes (Figure 5.2). A number of reports have compared differences in genetic structure between SSRs and SNPs in wheat, barley and maize (Hamblin et al., 2007; Van Inghelandt et al., 2010) and (Remington et al., 2001). Matthies et al. (2012) compares SSR and DArT markers in barley. They too demonstrate a difference in population structure results from the marker types, and also in GWAS results, although this appears to be a trait-specific outcome. The results from this study do not explore the implications of marker differences between the SNP marker set and the DArT marker set to the same extent as Matthies et al. (2012) did with SSR and DArT markers. However, the association results show that the marker types are one factor driving the outcome of association analyses. As far as we are aware, there are no previous reports comparing SNP and DArT marker genetic structure in a population.

5.5.4 Loci associated with relative maturity at flowering time

While the results between marker types and analysis pipelines differed somewhat, the analyses used in this study identified a range of the known major flowering time loci, including the main genes under selection in the Australian wheat breeding germplasm pool (Eagles et al., 2009). Both DArT and SNP markers identified the photoperiod sensitivity locus Ppd-D1 on chromosome 2D, with a range of DArT and SNP markers on the 2D consensus maps from 43cM to 84cM that could possibly be linked with this gene, although only the SNP marker wsnp_CAP12_c812_428290 was identified in both WGAIM and GAPIT analyses in all experiments. To our knowledge
this marker (corresponding to the standard marker designation IWA989) is not reported elsewhere as being associated with ZAD. Given the importance of the \textit{Ppd-D1} locus to selection of growth duration and adaptability (Kamran \textit{et al.}, 2014), this SNP marker may prove useful to supplement other markers for this gene such as those outlined in Cane \textit{et al.} (2013), especially as it can be utilised in conjunction with the array of SNPs generated as part of a whole genome. Further work is needed to determine which alleles of \textit{Ppd-D1} correspond to the different alleles of this SNP.

Both WGAIM and GAPIT analyses were able to detect the \textit{Vrn-A1} and \textit{Vrn-D1} loci across multiple experiments, although only SNP markers were identified in the GAPIT analysis including 11 markers for \textit{Vrn-A1}, and three markers for \textit{Vrn-B1}. One marker was detected for \textit{Vrn-D1}. In contrast, the WGAIM analysis detected eight loci on chromosome 5A, with the marker \textit{wsnp_Ex_c31799_40545376} common to both GAPIT and WGAIM analyses and the rest distributed across the consensus map for that chromosome. One marker identified in the GAPIT analysis only (\textit{wsnp\_AJ612027A\_Ta\_2\_1}) was also reported to be associated with the \textit{Vrn-A1} locus by Lopes \textit{et al.} (2015). A total of 13 loci were detected across both marker sets on chromosome 5B with WGAIM, these were also spread across the consensus chromosome map, with the closest to the three markers identified with GAPIT on chromosome 5B at least 15cM away according to the respective consensus maps. These three GAPIT markers are within 10cM of the \textit{Vrn-B1} locus reported by Guedira \textit{et al.} (2014). The single marker on chromosome 5D associated with flowering time (\textit{wsnp\_Ex\_c508\_1008029}) at all sites with WGAIM and at both 10YANA\_IRR and 10YANA\_experiments with GAPIT corresponds to the vicinity of the \textit{Vrn-D1} locus (Eagles \textit{et al.}, 2009). Eagles \textit{et al.} (2009) noted that \textit{Vrn-D1} occurs relatively infrequently
in Australian germplasm, so the inclusion of CIMMYT germplasm in this study may have helped detection of this locus.

An additional marker associated with flowering time at both the 09YANA_IRR and 09YANA_RFD experiments in GAPIT and the 09COLE_RFD, 09YANA_RFD, 09YANA_IRR and 10COLE_RFD experiments in WGAIM was located on chromosome 4B (wsnp_BE422566B_Ta_1_2). This encompasses both the water deficit environments in this study as well as the well-watered environments. The allelic effect from the WGAIM analysis ranged from 2.1 to 3.6%. There is no well characterised major gene for flowering time or earliness *per se* on 4B mentioned in the literature, although QTL for heading date on this chromosome have been previously reported by Hanocq *et al.* (2007); Griffiths *et al.* (2009); and Le Gouis *et al.* (2012), and photoperiod sensitivity QTL on this chromosome have been reported by Shindo *et al.* (2003) and Sourdille *et al.* (2003).

### 5.5.5 Loci associated with both water-soluble carbohydrate concentration and relative maturity at flowering time

Markers significant for WSCC in the WGAIM analysis included two markers also identified in the WGAIM flowering time analysis on chromosomes 1A and 2B, as well as five markers mapping close to other loci associated with ZAD. The marker *wPt-6531* on chromosome 2D in the WGAIM analyses has also been associated with grain yield by Edae *et al.* (2014), and is located between the markers *wPt-9780* and *wPt-6847* on the consensus map, which were both significant for ZAD. This places marker *wPt-6531* somewhere in the vicinity of *Ppd-D1*. The large distances between significant markers may be something to do with the *Ppd-D1* locus being heavily selected by breeders, so that there is a higher degree of linkage drag for genomic regions nearby (Kamran *et al*., 2014). It is also likely that the maker *wPt-797974* which is significant for
WSCC in the GAPIT analysis on chromosome 2D is associated with the flowering time marker \textit{wPt}-730744 which in turn is close to the marker \textit{wPt}-9780, even though according to the consensus map they are approximately 30cM apart (Figure 5.7). These results support the findings of Rebetzke \textit{et al.} (2008) that flowering time loci are key genes associated with WSCC. This is unsurprising given that the WSCC accumulation and remobilisation cycle is driven by phenology (Gebbing, 2003).

\textbf{5.5.6 Loci associated with water-soluble carbohydrate concentration previously associated with other traits}

From the WGAIM analysis the marker \textit{wPt}-2751 on chromosome 1B is significant for WSCC in three irrigated experiments and is located close to \textit{wPt}-7359 on the consensus map, which is significant in the two water deficit environment experiments 09YANA\_RFD and 09COLE\_RFD in the GAPIT analysis. Both of these markers are reportedly close to the \textit{Glu-3Bh} locus (Cui \textit{et al.}, 2014a), and \textit{wPt}-2751 has been previously associated with kernel weight (Zhang \textit{et al.}, 2014b). The marker \textit{wPt}-4209 was also significant in the WGAIM analysis, and is reported to be associated with grain yield under drought conditions (Crossa \textit{et al.}, 2007) and kernel diameter (Edae \textit{et al.}, 2014). Marker \textit{wPt}-5316 was also associated with yield by Crossa \textit{et al.} (2007). For the GAPIT analysis, marker \textit{wPt}-800147 has been previously reported to be associated with plant height (Yu \textit{et al.}, 2014) and seedling shoot dry weight under normal and saline conditions (Masoudi \textit{et al.}, 2015).

\textbf{5.5.7 Loci associated with water-soluble carbohydrate concentration across experiments and analysis pipelines}

The GWAS at multiple experiments in both WGAIM and GAPIT analysis pipelines detected markers on chromosomes 1A and 1D associated with WSCC. The marker \textit{wPt}-3743 on chromosome 1D was associated with WSCC in the two water
deficit experiments in both the WGAIM and the GAPIT analyses, with two other markers in the GAPIT analysis located within 5cM on the consensus map (Figure 5.6). This marker does not appear to be associated with any flowering time loci, and instead has been previously reported to be located in a region close to the high molecular weight glutenin Glu-D1 locus, and the storage protein activator gene locus SPA-D (Plessis et al., 2013; Deng et al., 2015). The Glu-D1 locus is an important trait for selection as, along with the Glu-A1 and Glu-B1 loci, it is responsible for a large percentage of the phenotypic variation for dough quality. It is notable that glutenin subunits encoded by the Glu-1D locus are thought to have the greatest impact on bread-making quality (Payne, 1987). Associations for a range of grain quality traits with this locus have been reported by Plessis et al. (2013), including grain protein concentration, quantities of grain storage protein, quantities of each gliadin class per grain and quantities of each glutenin subunit per grain. The combination of glutenin alleles present will largely determine the end use and marketing class of wheat varieties to meet grain quality standards (Whiting, 2004). Glutenin protein complexes play an important role in conferring elasticity and strength in wheat dough (Plessis et al., 2013), and the Glu loci have been shown to collocate with QTL for nitrogen and dry matter accumulation in grain (Charmet et al., 2005). According to Whiting (2004), there is variation for the major Glu alleles in the genotypes that make up this GWAS population.

This locus could affect WSCC in a number of ways. As such an important trait for wheat breeders targeting bread-making quality (Payne, 1987), it is possible that selection for Glu-D1 has resulted in the inclusion of a genetic background of other chromosome blocks in LD with this locus during selection that favours the accumulation of WSC. It is also possible that the locus containing Glu-D1 and SPA-D
genes and the wPt-3743 marker could have pleotropic effects on WSCC accumulation and remobilisation, and as such it is logical that loci known to be important for grain dry matter accumulation would also be associated with WSC. Gebbing and Schnyder (1999) measured the contribution of pre anthesis carbon assimilates to grain protein at 30-47\%, and the source strength of Glu-D1 could influence the remobilisation of WSC reserves and therefore WSCC levels, especially as the water deficit environments were sampled for WSCC later in crop development terms than the well-watered environment experiments (section 5.5.1). Different Glu-D1 alleles have been reported to have different onset of glutenin subunit polymerisation between anthesis and maturity (Irmak et al., 2008), which could be responsible for temporal differences in sink demand for WSC.

A third possibility is the association of WSCC accumulation and Glu-D1 with a third trait or traits, such as plant nitrogen concentration. WSCC is influenced by nitrogen content, as higher nitrogen availability in the plant drives sink demand for assimilate (van Herwaarden et al., 1998a; Ruuska et al., 2008), and WSC tends to accumulate in the absence of sink demand (Gebbing, 2003). It is interesting to note Glu loci are associated with WSC content per tiller in Rebetzke et al. (2008) but not with WSCC. Potentially the Glu loci could be involved with the inheritance of WSC through an interaction between nitrogen use, tiller number and grain weight. The Glu-D1 and SPA-D loci also contribute to phenotypic variation for grain yield and grain number through the plant response to nitrogen (Bordes et al., 2013). The SPA-D gene expression changes are induced by signals that indicate the availability of nitrogen, a regulatory network that maintains homeostasis for the total protein per grain (Ravel et al., 2009).

Fourthly it is possible that this gene rich area of the genome contains a separate locus with genes influencing WSC accumulation or remobilisation, which is plausible
given the marker *wPt-3743* has been reportedly associated with a range of other traits including grain yield and resistance to yellow rust, powdery mildew, and leaf rust (Crossa *et al.*, 2007), grain yield and spike length under salt stress conditions (Azadi *et al.*, 2015), and spike number (tiller number) per plant (Cui *et al.*, 2014b).

The *wPt-9592* marker was also identified in the water deficit environments using both GAPIT and WGAIM methods. This marker has been identified in a number of association studies. Crossa *et al.* (2007) found this marker associated with grain yield under water deficit conditions (in particular, terminal drought similar to the conditions experienced in this study) as well as yellow rust resistance. Le Gouis *et al.* (2012) found this marker corresponding to a QTL for heading date after vernalisation. Singh *et al.* (2010) identified associations with this marker and seed dormancy.

Neither of the markers *wPt-3743* or *wPt-9592* appear to be associated with ZAD, in contrast to the results of Rebetzke *et al.* (2008). It is possible that by constraining the population of breeding lines used for WSCC measurement based on ZAD (see section 4.3.1) this study has succeeded in identifying loci for WSCC independent of genetic variation for ZAD. However, it is also possible that methods used for measurement of WSCC were not sensitive enough to detect associated variation in ZAD or, ZAD loci could possibly be in LD with the markers identified.

5.5.8 **Genome-wide association study methodology limitations**

This study illustrates the difficulty with which any certainty can be put to identified loci as being biologically significant in a statistical genetics association. Significant variation was observed between the different marker sets and the different analysis pipelines for both measured traits (see sections 5.5.2 and 5.5.3). It is possible that the larger number of markers identified with WGAIM compared to GAPIT reflect
the different way each method accounts for false discovery rates. The number of
markers used will alter the results, as marker association tests are not independent.
Within each marker set, using different (or reduced) subsets of markers in the GWAS
for either method used will change the GWAS results dramatically (data not shown).
The population structure kinship analysis (Figure 5.1 and Figure 5.2) show a relatively
weak clustering in the population, which suggests that these GWAS was not
substantially confounded by population structure.

The two markers identified (wPt-3743 and wPt-9592) accounted for 7- 8% and 5
- 6% of phenotypic variation for WSCC respectively in the WGAIM analysis (Table A.1)
compared to a broad sense heritability for WSCC of $H^2 = 0.62$ for 09COLE_RFD and $H^2$
= 0.74 for the 2009 09YANA_RFD experiments (Table 4.4). No major additive effect loci
were identified in the GWAS, with the difference in genetic variance explained
between the markers identified and line-mean heritability indicating that there is likely
to be other loci, or combinations of loci, involved in the inheritance of WSCC that are
not detected in these GWAS. It is clear from the GWAS results that WSCC is a complex
trait controlled by many different loci, and the ‘missing heritability’ effect noted by
Manolio et al. (2009) and Zila et al. (2013) is relevant here. The limitations of GWAS are
evident in this study, which seeks to identify loci associated with a particularly
phenotypically quantitative trait, using methodology that is designed to detect
relatively simply inherited QTL. In fact, this study strongly suggests that WSCC is not
controlled by major additive gene effects.

With these limitations evident, GWAS still has an important role in the plant
breeders toolbox. However, GWAS results are of considerably greater utility if the trait
of interest is under relatively simple genetic control with limited G × E interaction, or
Gx E interactions that can be well characterised. Certainly, many such traits can often
be selected easily and effectively through conventional phenotypic selection, for example wheat flowering time. However for traits where phenotyping is expensive, for example cooking and dough quality in wheat (Plessis et al., 2013), or the phenotype intermittent, for example pest and disease resistance (Crossa et al., 2007; Joukhadar et al., 2013), GWAS is an ideal tool to identify useful marker trait associations for application to marker-assisted selection.

The GWAS for flowering time in this study showed that the methods used are robust, and can detect major known loci. In this sense, the flowering time GWAS proved an effective positive control for the WSCC GWAS, although the WSCC GWAS may not have been effective for a number of reasons related to the population and experimental design rather than the analysis methodology. There may have been limited LD with the WSCC trait, purely by chance with the set of polymorphic markers used. Possibly the population utilised did not contain relatively common alleles for WSCC – and given this population was a selected set of advanced breeding lines, it is possible that WSCC is selected against in south-eastern Australian breeding programs. Ogut et al. (2015) showed that multi-family analysis was not as effective at identifying rare QTL (even of large effect) as single family analysis. Hence it is possible that a biparental population analysis will yield more insight into the genetic control of this trait (Rebetzke et al., 2008).

5.6 Conclusions

This study can add some loci to those previously identified in other studies for the accumulation of WSCC in water deficit south-eastern Australian TPE, including QTL on chromosomes 1A and 1D. These QTL have both previously been associated
with performance under water limited conditions. The marker on 1D co-locates with the \textit{Glu-D1} locus, which may have some pleiotropic effect on WSCC.

The results of this study show that GWAS can provide insight into the genetic architecture of a trait, although the identified loci need to be treated with circumspection. Comparing GWAS across sites and years is a more robust method to identify significant loci, and more confidence can be given if the significant loci are identified in other studies. GWAS is particularly useful when the trait in question is highly heritable and controlled by few loci with additive effects. Although traits that fulfil these criteria are often also easy to phenotype, such as ZAD, but this is not always the case and there are many traits (such as disease resistance or baking quality) where GWAS could identify predictive markers to enhance selection efficiency in a breeding program.

This study adds further evidence to the hypothesis that WSCC inheritance is not controlled by additive major gene effects (Rebetzke \textit{et al.}, 2008). WSCC appears to be polygenically inherited, and is therefore not well suited to analysis with GWAS. WSCC may well be influenced by epistatic or pleiotropic interactions between networks of loci, however the independent testing methods and limited power of GWAS means these methods are unlikely to detect such complex genetic effects.

The analysis of ZAD confirmed the location of key flowering time loci on chromosomes 2D, 5A, 5B and 5D. The SNP marker \textit{wsnp\_CAP12\_c812\_428290} on chromosome 2D was associated with ZAD at every experiment in both WGAIM and GAPIT analysis, and could prove a useful marker for breeders selecting for \textit{Ppd-D1}. Additionally, a locus for flowering time was identified on chromosome 4B, consistent with a number of reports for flowering time loci on this chromosome.
It is interesting to note that the loci associated with WSCC across experiments and analysis pipelines were not also associated with flowering time, which is in contrast to Rebetzke et al. (2008). This may be due to the sampling methods used for phenotyping WSCC, or because the association population lines for WSCC were selected to be constrained for development, or because major flowering time loci covaried with the identified WSCC loci.
6. Genomic Selection for Water-Soluble Carbohydrate Accumulation

6.1 Introduction

The previous chapter of this thesis investigated the possibility that GWAS could be used to identify genetic loci significant for WSCC, so that they could be utilised in MAS. From the results of this chapter, which largely concur with the results of previous studies such as Rebetzke et al. (2008), it is clear that there is no major loci independent of other key traits that account for a large proportion of genetic variance in the population used in this study.

Studies such as Makowsky et al. (2011) suggested that GWAS was more suited to traits where only a few loci control the trait of interest, but perform poorly for more complex traits affected by many loci. If this assumption can be applied to the inheritance of WSCC in wheat, then plant breeders will need to look beyond GWAS to use molecular markers to improve this trait, and use statistical techniques more suited to the complex genetic architecture of the trait.

While the previous chapter suggested that there are no major loci responsible for WSCC inheritance, results from Chapter 4 showed that broad-sense heritability within the study was moderate, indicating the potential for genetic improvement for this trait. The combination of moderate genetic variance and no large-size major loci indicate that WSCC might be controlled by many genetic loci of modestly small effect. If this is the case, then this trait is a suitable candidate for genomic selection where the marker effects across the whole genome are used to predict the breeding value of a line.
6.2 Aims

This study sought to investigate the potential of using genomic selection to breed for improved WSCC genotypes, by modelling GEBVs for WSCC, and generate a measure of the prediction accuracy of GEBVs by using cross validation. This will provide baseline information on the suitability of applying genomic selection to the genetic improvement of WSCC, so that conclusions can be drawn as to the potential usefulness of genomic selection for WSCC in breeding programs.

6.3 Materials and methods

6.3.1 Genetic material, phenotyping methods and the molecular marker information used in this study

This investigation of genomic selection for WSCC utilises data from Chapter 4 for phenotype values of WSCC at each of the experiments. The phenotyping methods for WSCC are described in Chapter 4.3.6. The molecular marker data used in genomic prediction models is described in Chapter 5, in particular, the SNP marker set and the kinship matrix calculated using those SNP markers described in Chapter 5.3.2.

6.3.2 Genomic selection mixed linear models

The linear mixed model for variable \( y = (y_1, \ldots, y_p) \), for \( p \) experiments \( (j = 1, 2, 3, \ldots p) \) and \( m \) varieties \((i = 1, 2, 3, \ldots m) \) and combining the variety main effects \((g)\), split into marker (additive, \( u_a \)), non-additive effects \((u_q)\) and the \( G \times E \) effects are given in Beeck et al. (2010), Oakey et al. (2007) and Oakey et al. (2006) as:

\[
y = X\tau + Z_g u_a + Z_g u_{\tilde{a}} + Z_u u + \eta
\]

Equation 6.1

where \( y \) is the \((n \times 1)\) data vector of the response variable across \( p \) experiment with \( N_j \) plots per experiment \( j; \tau \) is a \((t \times 1)\) vector of fixed effects for the corresponding \( n \times t \) design matrix \((X)\), including experiment main effects, design-based effects and other
extraneous field variation. The $n \times 1$ residual vector $\eta$ represents local stationary variation. The $m \times 1$ vector of overall genetic line effects $g$, with corresponding design matrix $Z_g$ is partitioned into a vector of additive line effects $u_a$ and non-additive effects $u_d$ such that $g = u_a + u_d$. Each $u_a$ and $u_d$ are assumed to be independent, and are jointly Gaussian with zero mean, so the variance model for additive and the residual non-additive (interaction) genetic effects, is given by

\[ \text{var}(u_a) = A \otimes G_{ea} \]

Equation 6.2

\[ \text{var}(u_d) = I_m \otimes G_{e\bar{a}} \]

Equation 6.3

where $I_m$ is used hereafter to denote an $m \times m$ identity matrix, and the corresponding design matrix $Z_g$ is $n \times m$ and relates observations to lines. The symbol $\otimes$ denotes the Kronecker product. $A$ is the known $m \times m$ additive relationship matrix computed from the marker effects (see Chapter 5.3.2 for details). For each variance model above,

\[ G_{es} = \left( \Lambda_{es}^{T \otimes s} \Lambda_{es} + \psi_{es} \right), s = a, \bar{a}. \]

Here, $\psi$ is given by a diagonal matrix of the $p$ environment specific variances, thus accommodating the variance-covariance heterogeneity between experiments. This means that factor analytic models of different orders may be required to model each of the two genetic components.

The variance parameters from this model (Equation 6.1) were used as starting values for a factor analytic model of order 1, where the vector of random variety effects in the factor analytic model can be written as

\[ g = (\Lambda \otimes I_m)f + \delta \]

Equation 6.4

where $\Lambda$'s are the factor loadings for each experiment $j$, $f$ contains the scores for each variety $i$ and $\delta$ is the residual term of the multiplicative model.
6.3.3 Factor analytic model selection for genomic prediction

A sequence of factor analytic models incorporating phenotype and genotype information, as well as the kinship matrix [after Beeck et al. (2010) and Oakey et al. (2007)] were fitted to the data. These models were assessed for suitability using both AIC and Chi-squared tests (Table 6.1). Care was taken to ensure that the models converged at the maximum log likelihood, as complex FA models can sometimes falsely indicate convergence at lower log likelihoods. Model E in which an FA model of order two was used for both additive and non-additive variance matrices provides the best fit compared to combinations of FA1: additive and FA2: non-additive, and FA2: additive and FA1: non-additive models (Table 6.1). Model combinations with three factors proved too computationally difficult to fit, even though it is within the maximum number of factors possible for this dataset described by Smith (1999) in Equation 4.4. Model E was chosen as the preferred G-BLUP model for genomic prediction and used to obtain GEBVs.
Table 6.1 Factor analytic models fitted to the dataset for genomic prediction. Increasing order factor models were assessed using AIC and Chi-squared tests. Model E in this table (additive:FA2/non-additive:FA2) shows a significant improvement in fit from both additive:FA1/non-additive:FA1 and additive:FA1/ non-additive:FA2. Higher order FA models were not possible to fit with the computing resources available.

<table>
<thead>
<tr>
<th>Model</th>
<th>FA(k) Additive</th>
<th>FA(k) Non-additive</th>
<th>Log Likelihood</th>
<th>AIC</th>
<th>Parameters</th>
<th>Full / reduced model parameters difference</th>
<th>Chi squared test model comparison:</th>
<th>Critical value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>DIAG</td>
<td>DIAG</td>
<td>4037.480</td>
<td>-8042.956</td>
<td>16</td>
<td>-</td>
<td>to DIAG/DIAG:</td>
<td>736.32</td>
<td>2.389×10^{-146}</td>
</tr>
<tr>
<td>B</td>
<td>FA1</td>
<td>FA1</td>
<td>4405.638</td>
<td>-8747.276</td>
<td>32</td>
<td>16</td>
<td>to DIAG/DIAG:</td>
<td>736.32</td>
<td>2.389×10^{-146}</td>
</tr>
<tr>
<td>C</td>
<td>FA1</td>
<td>FA2</td>
<td>4465.921</td>
<td>-8853.842</td>
<td>39</td>
<td>7</td>
<td>to FA1/FA1</td>
<td>120.57</td>
<td>5.840×10^{-23}</td>
</tr>
<tr>
<td>D</td>
<td>FA2</td>
<td>FA1</td>
<td>4453.568</td>
<td>-8829.136</td>
<td>39</td>
<td>7</td>
<td>to FA1/FA1</td>
<td>120.57</td>
<td>5.840×10^{-23}</td>
</tr>
<tr>
<td>E</td>
<td>FA2</td>
<td>FA2</td>
<td>4473.418</td>
<td>-8854.836</td>
<td>46</td>
<td>7</td>
<td>to FA1/FA2</td>
<td>95.86</td>
<td>7.708×10^{-18}</td>
</tr>
<tr>
<td>E*</td>
<td>FA2</td>
<td>FA2</td>
<td>4473.418</td>
<td>-8854.836</td>
<td>46</td>
<td>7</td>
<td>to FA1/FA1</td>
<td>135.56</td>
<td>5.399×10^{-22}</td>
</tr>
<tr>
<td>F</td>
<td>FA3</td>
<td>FA3</td>
<td>(Convergence failed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 6.1 Biplot of REML estimates of rotated loadings and percentage variance from the G-BLUPs model for additive effects. This shows the cluster of well-watered environment experiments weighted towards loading one, and the two water-deficit environment experiments weighted towards loading two.
Figure 6.2 Biplot of REML estimates of rotated loadings and percentage variance from the G-BLUPs model for non-additive effects. Similar to the additive effects biplot in Figure 6.1, this figure shows the cluster of well-watered environment experiments weighted towards loading one, and the two water-deficit environment experiments weighted towards loading two.
6.3.4 Cross validation of the G-BLUP model genomic estimated breeding value predictions

The ‘fold’ method following Ogut et al. (2015), Crossa et al. (2014), Burgueño et al. (2012) and Lorenz et al. (2011) was used, which consisted of allocating the n genotypes of the MET dataset randomly to five subsets for a ‘five-fold’ cross validation scheme across separate experiments in the model. Each approximately 20% subset of genotypes has the phenotype observations removed once from the overall dataset, so that each genotype was included in exactly four training sets and one validation set. The model was run five times with each ~80% training set to obtain a full validation set of GEBVs. The average of 10 model runs was taken, and for each dataset the predicted values from the genotypes without phenotypes were compiled to yield a full set of cross validation predictions that were correlated to the observations from the original model, which was then run on the whole dataset to assess predictive ability.

Random allocation to the cross-validation bins resulted in the percentage of genotypes allocated to each bin varying slightly. However each genotype is excluded from one bin and present in the other four only once. Cross validation was not extended to experiment-by-experiment comparisons because of the computational time required.

Relative accuracy (RA) from the fivefold cross validation was calculated following Lorenz et al. (2012) and Sallam et al. (2015) by

\[
RA = \frac{corr(GEBV, observed)}{\sqrt{H^2}}
\]

Equation 6.5

where RA is calculated as the Pearson’s rank correlation of the genomic predictions (lines without phenotypes, otherwise called the validation set) and the observed true breeding values (lines with phenotype as well as genotype in the model, otherwise
called the training population). This value is divided by the square root of the heritability estimate for the validation population to give an accuracy value relative to phenotypic selection (Legarra et al., 2008; Chen et al., 2011).

The heritability for the RA estimation in the water deficit and well-watered environment clusters was calculated following Zila et al. (2013), who used the generalized line-mean heritability formula of Cullis et al. (2006), and estimated $\sigma^2_g$ from the average of the pairwise covariance estimates between experiments for the experiments in each environment cluster from the genotype plus genotype × environment covariance matrix, obtained in this study from the FA G-BLUP model.

6.4 Results

6.4.1 Genotypic and genotype × environment interaction variance explained by the model

The REML estimates of multiplicative genotype plus genotype × environment variance following Beeck et al. (2010) showed that Model E explained 99% of the total additive genetic and G × E variance and 86% of total non-additive genotypic and G × E variance. The proportion of genotypic and G × E variance explained by marker effects (additive variance) ranged from 15 to 50% across experiments. When calculated across experiments in each environment cluster, the proportion of additive variance was 36% for the well-watered environment cluster, and 23% for the water deficit environment cluster (Table 6.2). Experiments clustered into two distinct groups consisting of the same well-watered experiments and water deficit experiments identified in the phenotype-only analysis in Chapter 4. The environment cluster variances calculated across experiments in the well-watered environment cluster, and the water-deficit environmental cluster (Table 6.2) showed that additive genetic variance due to the
marker effects was small, and explained less genetic variance in the water-deficit cluster of experiments (23%) than in the well-watered cluster of experiments (36%).

6.4.2 Genomic estimated breeding values for water-soluble carbohydrate concentration

Modelled GEBVs for additive genetic effects were obtained from the model for all 973 genotypes, while a selection index for non-additive and combined additive and non-additive (total selection index) were obtained from the model for the 358 varieties that had WSCC phenotype values. The phenotype-only FA3 heatmap from the E-BLUPs model described in Chapter 4 provided a comparison with the heatmaps that include marker effects (Figure 4.13). Each heatmap summarises the correlation of genotype ranking between experiments (i.e. the genetic correlation between experiments). There are two distinct clusters of water deficit (the drought affected 09COLE_RFD and 09YANA_RFD experiments) and well-watered environments (09COLE_IRR, 90YANA_IRR and all 2010 experiments). The rotated loadings for the FA2 modelled additive and non-additive effects are shown in Figure 6.1 and Figure 6.2.

The additive-only GEBVs heatmap showed that the water deficit experiments and well-watered experiments maintained the same cluster structure with some additional correlation between the 09YANA_RFD experiment and the experiments in the well-watered cluster (Figure 6.3). The non-additive selection index indicated the same clustering as the additive GEBVs, but with a slightly weaker correlation between 09COLE_IRR and the other well-watered environment experiments (Figure 6.4). The 09COLE_IRR experiment was the only experiment to be grown on raised beds with all other experiments utilising a flat irrigation layout. The total genetic effects correlations between experiments is shown in Figure 6.5.
6.4.3 Genomic predictions for carbohydrate accumulation

The combined G-BLUP model closely matched the results of the phenotype only E-BLUP model developed in Chapter 4. Correlation between total genetic selection index and the phenotype only model are shown in Figure 6.6. For the water deficit cluster of experiments, \( r = 0.978 \) (\( r^2 = 0.958 \)), and for well-watered environment experiments \( r = 0.993 \) (\( r^2 = 0.987 \)). In Figure 6.7 the predicted additive effect GEBVs are regressed on the combined additive and non-additive selection index. For water deficit, \( r = 0.761 \) (\( r^2 = 0.579 \), \( p<0.05 \)), and for well-watered \( r = 0.833 \) (\( r^2=0.694 \), \( p<0.05 \)). There is some difference between environments in the percentage of phenotypic variance explained by genomic predictions. The predicted non-additive effects were plotted against total selection index in Figure 6.8. For the water deficit cluster, \( r = 0.863 \) (\( r^2 = 0.744 \), \( p<0.05 \)). For the well-watered cluster, \( r = 0.819 \) (\( r^2 = 0.671 \), \( p<0.05 \)). This indicates that the residual genetic variance from the additive part of the model is explaining the phenotype better than the additive part of the model (Table 6.2), and that the relationship is more pronounced in the water deficit rather than the well-watered environment experiment cluster.

Correlation between the additive GEBVs and non-additive selection index are shown in Figure 6.9. For the water deficit cluster, \( r = 0.329 \) (\( r^2 = 0.108 \), \( p<0.05 \)). For the well-watered cluster, \( r = 0.365 \) (\( r^2 = 0.133 \), \( p<0.05 \)).
Table 6.2: Genetic variances calculated from the average of the covariances from each environment cluster of experiments in the genotype and G × E covariance matrix from the G-BLUPs model following Zila et al. (2013).

<table>
<thead>
<tr>
<th>Experiment Cluster</th>
<th>Additive Genetic Variance</th>
<th>Non-additive Variance</th>
<th>Proportion of Additive Genetic Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well-watered</td>
<td>0.01297</td>
<td>0.02299</td>
<td>36.07%</td>
</tr>
<tr>
<td>Water deficit</td>
<td>0.00914</td>
<td>0.03028</td>
<td>23.19%</td>
</tr>
<tr>
<td>All experiments</td>
<td>0.00824</td>
<td>0.01537</td>
<td>34.90%</td>
</tr>
</tbody>
</table>
Figure 6.3 Additive GEBV genotypic and $G \times E$ correlations between experiments. The experiments cluster strongly into a well-watered environment grouping of all of the 2010 experiments, 09COLE_I RR and 09YANA_I RR, and a water deficit cluster consisting of the 09COLE_RFD and 09YANA_RFD experiments.
Figure 6.4 Non-additive selection index genotypic and $G \times E$ correlations between experiments for the G-BLUP model. The clustering of well-watered and water deficit experiments is maintained compared to the equivalent heatmap for the GEBVs, with some change of experiment clustering in the well-watered environment cluster.
Figure 6.5 Combined additive and non-additive genotypic and G × E correlations between experiments for the GBLUP model. This heatmap shows a similar clustering of experiments to both the additive and non-additive only heatmaps with a cluster of well-watered and water deficit environments.
Figure 6.6 Total selection index (consisting of combined additive GEBVs and non-additive selection index) from the G-BLUPs model regressed on the phenotype-only E-BLUPs from the model used in Chapter 4 (Equation 4.1) for each environment cluster. For the water deficit cluster, $r = 0.979$, $r^2 = 0.958$, significant at $p<0.05$. For the well-watered cluster, $r = 0.993$, $r^2 = 0.987$, significant at $p<0.05$. 
Figure 6.7 Additive GEBVs regressed on the total (additive and non-additive) selection index for each environment cluster. For the water deficit cluster, \( r = 0.761, r^2 = 0.579 \), significant at \( p<0.05 \). For the well-watered cluster, \( r = 0.833, r^2 = 0.694 \), significant at \( p<0.05 \).
Figure 6.8 Non-additive selection index regressed on the total additive and non-additive total selection index for each environment cluster. For the water deficit cluster, $r = 0.863$, $r^2 = 0.744$, significant at $p<0.05$. For the well-watered cluster, $r = 0.819$, $r^2 = 0.671$, significant at $p<0.05$. 
Figure 6.9 Additive GEBVs regressed on the non-additive selection index for each environment cluster. For the water deficit cluster, $r = 0.329$, $r^2 = 0.108$, significant at $p<0.05$. For the well-watered cluster, $r = 0.365$, $r^2 = 0.133$, significant at $p<0.05$. 
6.4.4 Cross-validation of genomic estimated breeding values

The average correlation, or the predictive ability (Sallam et al., 2015) of GEBVs from the cross validation models is shown in Table 6.3. Here, the diagonal elements of the table indicate the predictive ability of the cross validation GEBVs with the observed, true breeding values of the full genotype and phenotype model at each environment cluster and each individual experiment. Predictive ability across the whole of the well-watered environment cluster experiments was moderate-to-strong at \( r = 0.502 \), and across both water deficit experiments was \( r = 0.455 \). The predictive ability within individual experiments ranged from \( r=0.461 \) to \( 0.535 \) for the well-watered experiments and was \( r=0.400 \) and \( r=0.447 \) for the predictive ability of each water deficit experiment. Between environment clusters, predictability was poor. Use of the water deficit phenotypes to predict performance in the well-watered environments was poorly correlated (\( r = 0.196 \)) and similarly for the well-watered environments to predict performance in the water deficit environment (\( r = 0.211 \)).

Relative accuracy for both water deficit and well-watered cluster environments is shown in Table 6.4. The relative accuracies for the well-watered environment cluster were high with \( RA = 0.65 \), and for the water deficit environments \( RA = 0.70 \). This indicates genomic selection could achieve genetic gain per breeding cycle equivalent to approximately 65-70% of the genetic gain from phenotypic selection in the respective TPE. An important feature of genomic selection is the facilitation of a reduction in the duration of the breeding cycle (Equation 2.1). Following the assumptions of Heffner et al. (2010); Lorenz et al. (2011); Desta and Ortiz (2014) and, genomic selection is predicted to facilitate a reduction in the breeding cycle time of the average wheat breeding program of at least half. Therefore relative to phenotypic selection, genetic
gain per unit of time from genomic selection in this study is estimated to be $2 \times 65\% = 130\%$ for the well-watered environment cluster and $2 \times 70\% = 140\%$ for the water deficit environment cluster (reviewed in section 2.20 and illustrated in Figure 2.2).
Table 6.3: Cross-environment predictive ability of GEBVs from each experiment, and from each environment cluster of experiments. Predictive ability is calculated as the correlation between the additive GEBV from the cross validation model (without phenotype values) and true breeding value (total selection index from the full model with marker and phenotype values). The training population should be predictive of breeding values in the validation population, and the correlation between the two provides an estimation of how accurate each set of additive-only GEBVs predict the total additive and non-additive selection index.

<table>
<thead>
<tr>
<th>Validation Population</th>
<th>Training Population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Well-watered</td>
</tr>
<tr>
<td>Well-watered</td>
<td>0.502</td>
</tr>
<tr>
<td>Water deficit</td>
<td>0.211</td>
</tr>
<tr>
<td>09COLE_IRR</td>
<td>0.497</td>
</tr>
<tr>
<td>09COLE_RFD</td>
<td>0.128</td>
</tr>
<tr>
<td>09YANA_IRR</td>
<td>0.495</td>
</tr>
<tr>
<td>09YANA_RFD</td>
<td>0.313</td>
</tr>
<tr>
<td>10COLE_IRR</td>
<td>0.499</td>
</tr>
<tr>
<td>10COLE_RFD</td>
<td>0.503</td>
</tr>
<tr>
<td>10YANA_IRR</td>
<td>0.491</td>
</tr>
<tr>
<td>10YANA_RFD</td>
<td>0.503</td>
</tr>
</tbody>
</table>
Table 6.4 Cross validation relative accuracy by experiment and environment cluster. The predictive ability of the GEBVs model at each experiment and environment cluster is divided by the mean line heritability to provide an accuracy measure relative to phenotypic selection.

<table>
<thead>
<tr>
<th>Cluster or Experiment</th>
<th>Line mean heritability</th>
<th>Predictive ability</th>
<th>Relative accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well-watered</td>
<td>0.595</td>
<td>0.502</td>
<td>0.651</td>
</tr>
<tr>
<td>Water deficit</td>
<td>0.425</td>
<td>0.455</td>
<td>0.698</td>
</tr>
<tr>
<td>09COLE_IRR</td>
<td>0.608</td>
<td>0.503</td>
<td>0.646</td>
</tr>
<tr>
<td>09COLE_RFD</td>
<td>0.619</td>
<td>0.471</td>
<td>0.598</td>
</tr>
<tr>
<td>09YANA_IRR</td>
<td>0.790</td>
<td>0.535</td>
<td>0.602</td>
</tr>
<tr>
<td>09YANA_RFD</td>
<td>0.738</td>
<td>0.445</td>
<td>0.518</td>
</tr>
<tr>
<td>10COLE_IRR</td>
<td>0.624</td>
<td>0.474</td>
<td>0.600</td>
</tr>
<tr>
<td>10COLE_RFD</td>
<td>0.428</td>
<td>0.466</td>
<td>0.712</td>
</tr>
<tr>
<td>10YANA_IRR</td>
<td>0.626</td>
<td>0.520</td>
<td>0.657</td>
</tr>
<tr>
<td>10YANA_RFD</td>
<td>0.745</td>
<td>0.481</td>
<td>0.558</td>
</tr>
</tbody>
</table>
6.5 Discussion

6.5.1 Integration of genomic selection with genotype × environment effects

The results provide a synthesis of the Target Population of Environments (TPE) selection methods of Comstock (1977) and Cooper et al. (1997) with genomic prediction methods. Clearly, the same principles apply to genomic selection as phenotypic selection, where predictive performance values, be they genomic or phenotypic, are relevant to the TPE underlying the prediction model. These results show that predictive ability depends on the extent and nature of the G × E relationship between the training and the validation populations. These results are comparable to Heslot et al. (2013), who show empirically in their study of wheat genomic selection that G × E patterns for phenotypic selection are similar and can be dealt with in the same way. Heslot et al. (2013) also observed that the main driver of prediction accuracy between environments were G × E effects. This may be a cause of the low predictive ability reported in some other genomic selection studies [such as those reviewed by Lin et al. (2014)] where G × E is not accounted for. These results show it is not necessary to model genomic predictions in a homogenous environment such as reported in Sallam et al. (2015), or by modelling each environment separately such as reported in Crossa et al. (2010), and G × E can be accounted for in an MET genomic prediction model. In fact, the responses of genotypes in genomic selection models provide a way of characterising target environments, still following the ideas on exploiting G × E to make genetic gains outlined by Byth (1981). Ly et al. (2013) goes part of the way to identifying G × E patterns in a large MET dataset but the authors focus on locations as the source of G × E, as the authors viewed the year-by-location effect as too difficult to capture (and interpret). The environment clusters in this study show that year effects
are more important than location for WSCC (Figure 6.3 and Figure 6.5), and this observation is borne out by other G \times E studies with similar findings, particularly for traits where expression is significantly affected by seasonal conditions (Chenu et al., 2011; Milgate et al., 2015; Smith et al., 2015).

This study can be contrasted to the approach by Asoro et al. (2011) where G \times E terms are included, but the authors suggest [with reference to the Falconer and Mackay (1996) selection equation] that they detract from heritability because genetic variance components are reduced when the environmental interaction terms are added to the model. It can be suggested that G \times E needs to be more appropriately addressed in genomic selection models, as clearly such interactions are statistically significant effects and cannot be ignored. After all, it is not the genotype-phenotype association which is important to plant improvement, it is the genotype effect-phenotype association that matters (Eichler et al., 2010). In the case of Asoro et al. (2011) and Storlie and Charmet (2013) it would suggest that these studies encountered multiple target environments, (highly probable given the former study covered 388 experiments in a MET over 13 years, and the latter used a data set that included six locations over 10 years) which individually require assessment of heritability and selection response. Such an approach can be found in Heslot et al. (2013) who note that genomic predictions are specific to the TPE they are predicted for.

This study employs the factor analytic (FA) models of Smith (1999) as implemented by Kelly et al. (2007) and Cullis et al. (2010) to model G \times E effectively. Heslot et al. (2014) suggests that the FA approach is not appropriate for genomic predictions, as FA models do not allow prediction of G \times E in unobserved environments. However, use of GEBVs for selection in an environment outside the TPE previously characterised and targeted by the breeding program would be a rare
situation in either phenotype-only selection or genomic selection. Indeed, it is unlikely
to be a successful approach for a breeding program to follow anyway (Cooper et al.,
1995; Cooper et al., 1997; Basford & Cooper, 1998). The covariances between correlated
environments in FA models are an important consequence of the analysis, as they can
contribute to potentially better predictions than single-site analysis, because of
information in the model borrowed from across environments (Cullis et al., 2010; Guo
et al., 2013). This can identify those environments where the expression of the trait in
question is genetically correlated (where lines rank similarly), which means that
predictions are likely to hold across environments. Burgueño et al. (2012) utilised such
FA models in a similar fashion to this study. However, the authors considered four
environments and made cross-environment predictions rather than determine the
nature of G × E in their study. This approach has limitations, as the average of
environment effects can be misleading, given genetic correlations between
environments vary greatly, and sometimes line ranking between sites can be negatively
correlated (Cooper et al., 1997). Rutkoski et al. (2015) also utilised FA models and
predicted across 16 environments without a detailed consideration of the G × E
interactions between the experiments in the model. This study followed a different
approach and used the clustering approach of Beeck et al. (2010), and although this
approach is not without limitations in determining the extent of clusters, it is an
established method for using genotype rankings to determine groups of trials where it
is appropriate to compute an average predicted value. If predicted values were
averaged across all environments in this study, given the poor genetic correlation and
predictive ability between well-watered and water deficit environment experiments,
then the average predicted values are likely to be intermediate to both environment
clusters and relevant to neither (Smith et al., 2001a; 2001b).
6.5.2 Practical application of genomic selection for water soluble carbohydrate improvement

This study provided some evidence that genomic selection for WSCC could be implemented in Australian wheat breeding programs to provide more rapid genetic gain than phenotypic selection alone. Implementing genomic selection in a standard wheat breeding program where the relative accuracy is around 65-70% could see genetic gains of 30-40% more per unit time if the breeding cycle can be reduced by half compared to phenotypic selection (reviewed in section 2.20). With reference to Equation 2.1 and Figure 2.2, genomic selection may facilitate a reduction in the time between identification of superior genotypes, and the crossing process (the breeding cycle). Standard phenotypic selection means that the breeding cycle (between crosses) is at least four years, as lines need to reach at least the F5 generation to be sufficiently genetically fixed (homozygous) for a stable phenotype to be obtained. Genomic selection uses the phenotypes measured in the line evaluation phases of the breeding program (the training population) in a genomic prediction model to predict genotypes which are more likely to have the phenotype of interest in the early generations of a breeding program. Identified superior genotypes can be utilized in crossing, therefore reducing the breeding cycle time from four years to two years (Heffner et al., 2010; Lorenz et al., 2011; Sallam et al., 2015).

In addition to a reduction in breeding cycle duration, there may be further resource savings given WSCC is not a straightforward trait to phenotype. Levels of WSCC increase and decrease with crop development and growing conditions, meaning that full expression of the phenotype is not easily captured, and both wet chemistry and NIRS systems for measuring water soluble carbohydrates are time consuming and expensive. On the other hand whole genome marker profiles are decreasing in cost and
may continue to do so (Heffner et al., 2010; Lorenz et al., 2011). The challenge in practical terms, as noted by Cooper et al. (2014), may be in implementing the enabling technologies needed to make genomic selection work and readily integrated into a commercial breeding program. Plant breeding database systems are currently not equipped to handle whole genome scale marker information, and data visualisation and analysis tools for whole genome profiles are only starting to be developed and implemented for use.

A key question for plant breeders considering implementation of genomic selection is how to maintain the relevance of a training population relative to genomic selection progress in their breeding programs. As genetic gain is made in a breeding program, the training population needs to be kept relevant to the early-generation material that is undergoing selection (Smith, 2014; Rutkoski et al., 2015). Isidro et al. (2015) explored various aspects of training population development, and concluded that training populations should be optimised by both maximising the phenotypic variation and minimising the genetic relationship between genotypes as much as possible, without making the training population unrepresentative of the diversity in the prediction population. Typically, a wide range of germplasm will be included in addition to the advanced lines selected from within the breeding program (Figure 2.2). This could possibly require retention of some germplasm that is phenotypically inferior in the later stages of the breeding process in order to provide the phenotypic variance needed for the training population model.

The approach of genomic selection differs from conventional phenotypic selection in that the expression of a successful phenotype in a complex trait such as yield, or in this case WSCC, could be the result of any combination of component traits and biological pathways, and conventional phenotypic selection does not favour any
particular final combination. Indeed, conventional plant breeding selection is not about ‘picking winners’ but shifting the population distribution and therefore the mean for the trait in question by discarding poor performers (pers. comm. Prof Major Goodman, North Carolina State University: “you can always pick the dogs”). In contrast, genomic selection prescribes the genetic variation to be selected, as a strategy of ‘genomic enrichment’ for alleles and genome regions more likely to result in a favourable phenotype – although only within the range of phenotypes and genotypes included in the training population. A potential challenge to plant breeders arising from this genetic bottleneck is the effect of genomic selection methods on long term genetic gain in a breeding program (Heffner et al., 2010). Genomic selection is expected to more rapidly fix many loci, by more frequent co-selection of relatives (Jannink et al., 2010). Genetic variance will decrease more rapidly than with phenotypic selection therefore limiting the longer-term genetic potential of the breeding population and increasing the probability of losing favourable alleles that occur at low frequency through genetic drift. This has been considered in dairy cow improvement (Goddard, 2009; Hayes et al., 2009b), and Jannink (2010) has discussed the implication to crop improvement. In the simulation study of Jannink (2010), genetic gain from genomic selection reached a maximum after around 12 cycles, regardless of the training population used or the heritability of the trait.

In a breeding program, this may lead to short term breeding success but limit longer term breeding gains. Studies on crop genetic diversity have shown that there are ‘genome deserts’, and selective sweeps of low polymorphism around major domestication loci (Wang et al., 2009; Cavanagh et al., 2013; Krishnan S et al., 2014; Jordan et al., 2015), and on a much smaller scale, genomic selection will be also contributing to increased disequilibrium and lack of recombination around early fixed
alleles, possibly more rapidly than with phenotypic selection. This issue provides
another reason to keep updating a ‘dynamic’ training population for a breeding
program strategy, somewhat akin to the ‘mapping as you go’ strategy in Podlich et al.
(2004), or weighting favourable alleles in the genomic selection model, at some cost to
short term gain (Hayes et al., 2009b). At any rate, if the genomic selection is in the
direction of the breeding target, bottlenecks in genetic diversity can be accommodated
especially if genomic selection results in the identification and aggregation of
combinations of large- and small-effect alleles in elite germplasm.

Genomic selection is somewhat analogous to the ideotype concepts of
phenotypic selection, and possibly susceptible to some of the same criticisms.
Combinations of particular plant traits and architecture are not always the only way to
achieve a superior phenotype. For example, expression of grain yield as a phenotype
could potentially reflect numerous trait combinations (Richards, 1991; Rebetzke et al.,
2009). This can be compared to genomic selection, where regions of the genome are
selected based on importance to the relationship matrix in the training population
model. Although genomic selection essentially treats the genetic architecture (and
therefore the biology) of a trait as a black box, it could be used to select an ideotype of
QTL combinations that would be difficult to assemble through marker-assisted
selection (Nakaya & Isobe, 2012; Guo et al., 2013). Understanding the physiology of the
trait in question may help, however assembling physiological traits together to design
the ideal crop plant has had a poor track record in delivering new varieties (Yang &
Zhang, 2010).
6.5.3 Accuracy of genomic predictions

The prediction accuracies for GEBVs in these data indicate that within each cluster of experiments, GEBVs predicted for one experiment can reliably predict performance in another experiment (Table 6.3). While it is not possible to know if these predictions will also apply to environments encountered in the future, the approach used in this study of characterising environmental interactions will go some way to predict future performance if the target population of environments does not change. Indeed, these methods have been applied successfully for many years in wheat breeding and varietal evaluation in Australia (Smith et al., 2015). In terms of the relative accuracy of GEBV predictions, the cross validation results show that for each environment cluster, GEBVs explain a moderate percentage of the predicted phenotypic variance (Table 6.3 and Table 6.4). The slightly smaller relative accuracy of GEBVs is indicative of the higher mean line heritability in the well-watered cluster of environments.

Accuracy of GEBVs is essentially a measure of persistence of LD between markers and the trait loci between generations (Asoro et al., 2011). However, preserving this LD relationship depends on the genetic similarity between the training population and the validation population. Because genomic selection does not utilise patterns of markers across the genome and is actually measuring the numbers of markers (and therefore genome) that is shared by line pairs, LD patterns (which will drive accuracy) can be hidden by other factors. This is where the relationship between the training population and the prediction population is very important to the prediction accuracy – LD patterns will need to be similar otherwise comparing the numbers of markers shared between line pairs may not be reflective of marker-trait association.
Diverse training populations have been frequently advocated for genomic selection (Heffner et al., 2009; Lorenz et al., 2011; Nakaya & Isobe, 2012; Windhausen et al., 2012), although with diversity often comes issues with population structure, or the occurrence of more closely related sub-populations. If these are left uncorrected, the GEBVs from the training population will have artificially high prediction accuracy and essentially predict population structure (Isidro et al., 2015). As with phenotypic selection, maximising genetic variance is more important that the range of phenotypic variation. Maximising genetic variance in genomic selection is critical to success (Lin et al., 2014), as it has always been in conventional selection (Cooper et al., 2014). The interaction terms in the prediction model are also important for prediction accuracy.

The issue around line ranking with G × E interaction (discussed in section 6.5.1 above) and genotype × genotype interactions (epistasis) may also be important but characterising these interactions, and particularly epistatic interactions will be challenging (Holland, 2007; Lorenz et al., 2011). In order to maintain prediction accuracy, training populations will need to evolve as LD between markers and traits decay over generations in a breeding program. Heffner et al. (2009) reported that it took three cycles of meiosis and recombination to break down LD in maize, although for inbreeding crops like wheat, LD will likely be conserved longer (Breseghello & Sorrells, 2006). A smaller training population with less diversity and comparable LD between markers and trait to the prediction population may be a better solution (Lorenz et al., 2012), with the extreme being the use of biparental populations as training populations (Heffner et al., 2011a).

While the aim is to maximise prediction accuracy of GEBVs, the reduction in breeding cycle time in a genomic selection system means that lower accuracies can still outperform phenotypic selection with regards to genetic gain over time, shown
theoretically by Heffner et al. (2011b) and empirically in a more recent study by Longin et al. (2015). However, the comparison with phenotypic selection gains may not be reflective of the methods available to a plant breeder selecting families in a breeding program. An estimation of progeny performance can be given simply by the family phenotypic mean (mid-parent mean), with no need for marker data or analysis (Ogut et al., 2015), and this would be a better base level of accuracy for genomic selection comparisons.

6.5.4 Correlations between the additive and non-additive genetic variance components in the model

The correlation between additive GEBVs and non-additive selection index in Figure 6.9 is interesting from the stand point of genetic architecture, as it is indicative that the additive genetic effects have not been completely absorbed into the additive genetic term in the model. This is in contrast with the results of Oakey et al. (2007) who used very similar models, and showed no correlation between additive genetic and interaction term predictions (Oakey et al., 2007 Figure 1). It is possible that this correlation could be a result of the factor analytic modelling process, whereby if the additive genetic effects are not fully absorbed into the additive genetic component of the model, the balance will be absorbed into the non-additive component. However given the similarity of the models used by Oakey et al. (2007), where no such concerns were apparent, this is unlikely. While the correlation in this study could be considered modest ($r = 0.329$ for the water deficit environment and $r = 0.365$ for the well-watered environment), it has significant implications for selection of WSCC as there is an interaction between the estimated breeding values (GEBVs) and the interaction term genetic effects in the model; in other words some level of epistasis.
Correlation between the additive and interaction genetic variance terms is possibly indicative of the additive × additive epistasis described in Schnell and Cockerham (1992). While individual traits may be largely explained by additive genetic variance, the expression of a complex phenotype with multiple underlying physiological mechanisms can result in multiplicative effect – termed phenotypic epistasis by Holland (2001b). This concept is also explored by Cheverud and Routman (1995) who show that phenotypic epistasis contributes to additive as well as interaction genetic variances and values, and could hence be measured as statistical epistasis. This fits well with the results of Chapter 4.4.3 which reported on the physiological nature of the accumulation of WSCC, where essentially a combination of plant architecture traits and traits that control source-sink relations interact to produce a phenotype that generate assimilates that can be remobilised. As detailed by Cheverud (1988), there is evidence that phenotypic correlations can suggest genetic correlations between traits, which may be the case with WSCC. These include correlations between traits involved in the supply of carbohydrates, such as flowering time, which is an indicator for plant phenology, and traits that control alternative sinks for assimilate, such as tiller number, and grain-filling through thousand grain weight.

It is important at this point to revisit the definitions of additivity and epistasis covered in this study. Rather than the effects of interactions of genes at individual loci, which is sometimes termed biological epistasis, this study instead refers to the concept of statistical additivity and epistasy introduced by Fisher (1918) and developed by Cockerham (1954), where the observed variance from the modelled population can be partitioned into an additive and an interaction (dominance or epistasis) term. The Fisher (1918) additive infinitesimal model assumption is that genetic effects are “cumulative Mendelian factors” and epistasis, as a deviation from the model.
assumptions is treated essentially as a residual genetic effect. That approach extends through to the models used in this study, following the methods of Oakey et al. (2007) but replacing a pedigree relationship matrix with the use of an IBD kinship matrix, which is assumed to represent additive genetic effects (Visscher et al., 2008; Habier et al., 2013).

In the statistical models used for genomic selection, what is assumed to be additive genetic effects are actually the first linear approximation of the marker effects in the model (pers. comm. Dr Mark Cooper, DuPont Pioneer, Iowa). Biologically these additive effects may arise from both additive gene action and also from dominance or epistatic effects of loci given the additive genetic variance is a function of squared average statistical effects (Cheverud & Routman, 1995; Holland, 2008; Zuk et al., 2012). As reviewed by Hill (2012), a high proportion of additive genetic variance does not imply that gene action is additive, which is compatible with the knowledge that many major genes are known to be dominant. The biological basis may also extend to statistical measures of additivity, given the results of Muñoz et al. (2014), who showed that additive variance terms can capture genetic interaction effects, because in breeding populations additive and non-additive genetic components are not independent – a conclusion largely borne out in this study. The results of Spiliopoulou et al. (2015) revealed that similar to pedigree based relationship matrices, the marker-based relationship matrix can also capture predictive signal that is due to factors other than additive genetic effects. In this context, it is also worth considering the nature of the SNP markers used to compute the additive genetic effects matrix (relationship matrix), as Vitezica et al. (2013) observes, SNP markers can include both additive and dominance biological effect markers. It is therefore likely that different SNP marker arrays may highlight different statistical genetic effects.
The pervasiveness of epistasis and interaction (i.e. non-additive) effects in genetic architecture of complex traits discussed by Carlborg and Haley (2004) and Zuk et al. (2012) is in contrast to the conclusions of Hill et al. (2008), whose review of mammalian systems shows evidence that genetic variance across complex traits is mostly additive. However, comparative modelling studies have shown that where pedigrees are used to estimate additivity (as in Hill et al., 2008) the partitioning of genetic variance is less precise, leading to the overestimation of additive genetic variance (Muñoz et al., 2014). Zuk et al. (2012) goes further, suggesting that ‘missing heritability’ reported in GWAS studies may in part be due to inadequately estimated genetic interaction terms. As Gianola et al. (2015) conclude, molecular markers are an imperfect instrument for interpreting the statistical genetic architecture of complex traits, and inferences from the marker effects in statistical genetic models should be treated with circumspection.

In terms of practical plant breeding and selection, the working assumption is that additive genetic variance reflects the breeding value of the line, given that genetic interaction terms are assumed to be lost due to recombination (except for G × E effects, as discussed above in section 6.5.1). Clearly not all of the measured additive genetic variance is additive, and it is likely that some of the interaction term genetic variance may be additive, depending on the nature of LD in the population sampled. It is possible that correlation between the additive and non-additive genetic terms in the model (Figure 6.9) may actually represent the potential for increased selection gains, since selection using GEBVs alone will also result in a correlated response to selection on the non-additive genetic selection index. The extent to which this blurring of the genetic variance terms has an impact on selection is difficult to predict, although Cooper et al. (2014) opine that a linear association is enough for genetic ranking and
therefore selection and genetic gain. In the end, a detailed knowledge of the trait and more complex modelling may not necessarily produce a better result.

Further work may consider ways of incorporating more detailed measures of epistasis in the modelling, even though doing so adds substantial complexity to the models used (Holland, 2007). Jonas and de Koning (2013) suggested that non-additive effects have not been adequately addressed in genomic selection research thus far, given the methods are largely drawn from dairy cattle breeding efforts where non-additive effects are assumed to be less relevant. If selection over generations is to be considered, the ‘mapping as you go’ concepts of Podlich et al. (2004) could be explored. Akdemir and Jannink (2015) discussed ways of partially including epistasis effects, with models that incorporate both additivity and ‘locally epistatic’ effects. Crossa et al. (2014) suggested that epistasis and interaction terms can be accounted for with results indicating that models accounting for non-linear genetic terms produced better predictions than those without. Oakey et al. (2006) and Muñoz et al. (2014) extended the methods used here by modelling both a dominance and additive effect relationship matrix. Together these would go some way to understanding the complex genetic architecture of this trait, although the difficulty of doing so in an inbreeding species means that specifically designed populations would need to be developed to implement this analysis.

6.6 Conclusions

This study provides empirical evidence that genomic selection methods could improve the rate of genetic gain forWSCC over phenotypic selection, providing the TPE are carefully characterised and understood. Genomic selection allows selection of alleles rather than fixed lines so reduced levels of accuracy in selection maybe offset if
there is also a significant reduction in the duration of the breeding cycle. Although genomic selection does not require a full understanding of the genetic architecture of a trait, it may be better suited to polygenic traits than those largely explained by few major loci. Such oligogenic traits may be better suited to conventional marker assisted selection.

The so-called additive effects characterised in genomic selection are assumed to be additive but are in fact the first linear approximation of the realised marker effects. They are most likely additive, but some dominance or epistatic interaction effects are also possibly captured. Correlations between the additive GEBVs and non-additive selection index in the G-BLUPs model indicate some complex genetic interactions in the WSCC trait architecture. Some possible explanations include genetic correlation between traits that need to be selected together, or are functionally related. Alternatively it may be evidence for phenotypic epistasis, where component traits that underlie complex characters have a multiplicative effect. However, this is of little consequence to the plant breeder using the genomic predictions, providing those predictions are relatively accurate and are able to provide a superior response to selection compared to phenotypic selection methods. The predictive ability of the GEBVs shown here, as well as the relative accuracies of those GEBVs indicate that genomic selection could well be a useful breeding technology to apply to selection of WSCC to increase the genetic gain for this trait.
7. General Discussion

7.1 Genetic architecture of water-soluble carbohydrate accumulation

This thesis explored the genetic architecture of WSC with a succession of empirical breeding methods. From the overall results, it was clear that WSC is complex genetically, as well as physiologically. These experiments showed moderate broad sense heritability for both WSCC and WSCA (Table 4.4), comparable to the broad sense heritability observed for grain yield. The heritability for both WSCA and WSCC observed here are consistent with Ruuska et al. (2006), although somewhat lower, perhaps due to the elite breeding germplasm sampled in this study. This population has undergone selection for yield performance and other characteristics and would not represent the full scope of wheat genetic diversity. The lower heritability observed for WSCA compared to WSCC was attributable to the lower repeatability of the biomass measurements that were used to calculate WSCA, so for the molecular marker analysis, investigations were based on WSCC.

The GWAS for WSCC demonstrated a key challenge in association studies: that of the difficulty in generating repeatable marker-trait associations. Statistical associations depended on a number of factors, including linkage disequilibrium patterns, marker types, and the type of analysis used, and need careful interpretation. The loci repeatedly identified in this study consisted of two QTL, on chromosomes 1A and 1D (Figure 5.5 and Figure 5.6). Together, these QTL explained around 10% of the heritability in each of the two water deficit environment experiments (section 5.5.7). These observations are consistent with the trend of ‘missing heritability’ in GWAS studies where the identified marker-trait associations only explain a small proportion of the overall heritability (Attia et al., 2009; Manolio et al., 2009). In this case however,
many possibly significant loci were discounted because of their lack of repeatability across experiments and analysis pipelines which was the criterion used in this study as a measure of repeatability. The lack of strong QTL is indicative of the genetic control of WSC, where there were no detectable major loci that could be exploited to improve this trait. This is consistent with the conclusions of Rebetzke et al. (2008). Interestingly, loci associated with both relative maturity at flowering time (ZAD) and WSCC were not consistently identifiable, which is in contrast to the results of Rebetzke et al. (2008). This may be due to the constrained flowering time range for the genotypes used in this study, compared to the variability in that study. Further evidence for some independence of ZAD and WSCC inheritance was seen in the correlated response to selection results in Table 4.5 and Table 4.6 where it appeared that ZAD was not significantly changed by selection for WSCC or WSCA in this population.

The absence of QTL detected by GWAS is in contrast to the predictive ability of GEBVs calculated from the whole genome marker profiles in Chapter 6, where within experiments in the same environment cluster, breeding values computed from the marker effects are approximately 65-70% as effective as phenotypic selection (Table 6.4). This compares to the much lower variance explained by the two QTL identified through GWAS discussed above. An interesting consequence of the computation of GEBVs is the observation of correlation between the additive and non-additive genetic variance components in the genomic selection model. This is indicative of either phenotypic epistasis (Holland, 2001a), genetic correlation between the component traits that give rise to WSC accumulation (Cheverud & Routman, 1995), or possibly the marker effects not explaining some of the additive genetic effects in the model used.
7.2 Selection for increased water-soluble carbohydrate accumulation

This study took a different approach to many of the studies of WSC genetic control preceding this work by focusing part of the investigation on conventional selection for WSC (Yang et al., 2007; McIntyre et al., 2010; Pinto et al., 2010; McIntyre et al., 2012). The findings of Chapter 4 indicate that there is substantial genetic variation for WSC in Australian wheat breeding germplasm, and conventional selection for this trait is likely to be possible, in line with the work of Ruuska et al. (2006) on diverse germplasm. In contrast to the conclusions of Rebetzke et al. (2008), Chapter 4 of this study found that WSCA may be a less useful trait to select on than WSCC, as repeatable measurement of WSCA is hindered by the lower accuracy of biomass estimations.

The correlations identified in other studies (including Rebetzke et al., 2008; and Dreccer et al., 2009) between higher WSC accumulation and reduced tillering, fewer grains per m² and greater grain weight were confirmed in this study, with significant correlations between these traits within both the well-watered and water deficit environment clusters of experiments. With consideration to flowering time varying with WSC levels, which was identified in Rebetzke et al. (2008), this appears to be accounted for in the sampling methods used in this study. Chapter 4 shows that WSC was more strongly related to ZAD than the other traits measured, however the correlated response to selection estimates indicated that the actual change in ZAD after selection for WSCC or WSCA was not be high (Table 4.5 and Table 4.6). Despite the correlation between WSCC and WSCA and other traits, overall this study did not find evidence that WSC accumulation is correlated with grain yield under well-watered or water deficit environments. This is in contrast to a body of literature investigating WSC
accumulation both under well-watered and water deficit conditions (Bidinger et al., 1977; Schnyder, 1993; Gebbing & Schnyder, 1999), and should be taken into account by wheat breeding programs seeking to select for this trait. The use of dry matter content (DMC) as a surrogate trait for WSC accumulation phenotyping was investigated in this study after Xue et al. (2009) highlighted the relationship between DMC and WSCC and the potential for selection to improve WSC accumulation. While the correlated response to selection was high (Table 4.5) heritability for DMC was lower than WSCC (Table 4.4), and the direction of selection differed between well-watered and water deficit environments.

If the use of genetic markers is more resource-effective than phenotypic selection, which will be difficult and expensive for WSC accumulation, then the two markers identified in Chapter 5 may prove useful in Australian breeding germplasm when selection is performed in water-limited environments. However as individual markers, these loci do not explain a large percentage of the variation in either of the water deficit experiments in which they were identified. Further, emphasis on selection for particular alleles at the Glu-1D locus, which is important for baking quality (Payne, 1987), may preclude the use of the marker wPt-3743 for WSCC which is located close to Glu-1D. The marker wPt-9592 found on chromosome 1A has been previously associated with yield in a water-limited environment, and deserves consideration by wheat breeders for marker-assisted selection of WSCC. As an alternative to individual markers, genomic selection may prove to be a useful tool for wheat breeders with access to whole genome marker profiles (Lorenz et al., 2011). The GEBVs modelled from the whole genome marker profiles in this study indicate around 65-70% accuracy relative to phenotypic selection, however using GEBVs rather than phenotype for selection allows selection to be performed much earlier than phenotypic selection,
possibly reducing the breeding cycle duration by half (Heffner et al., 2010; Lorenz et al., 2011). If this is the case, then genomic selection may result in 30-40% greater genetic gain per breeding cycle than phenotypic selection. If genomic selection for WSCC were implemented in wheat breeding programs, care should be taken with the development of an appropriately designed training population that encompasses the phenotypic variation for WSCC and, at the same time, restricts the genotypic variation to the scope of the material under selection in order to maximise prediction efficiency (Isidro et al., 2015).

7.3 Genotype × environment interactions in quantitative trait analysis

One of the central themes of this thesis is the critical need to characterise $G \times E$ interactions if useful conclusions are to be drawn from multi-environment trial data, be they for molecular genetics studies or for classical quantitative genetics models. Chapter 4 showed that characterisation of the TPE under which WSC is to be selected is essential, if genetic gains are to be made, given the strong $G \times E$ interactions involved in the expression of this trait. The nature of $G \times E$ should be readily exploitable, given the consistent performance of genotypes within environmental groupings. Previous studies have investigated WSC under well-watered (Rebetzke et al., 2008) and water deficit (Bennett et al., 2012) conditions, as well as limited trials in contrasting well-watered and water-deficit environments (Yang et al., 2007; McIntyre et al., 2010). This study showed contrasting genotypic performance under well-watered and water deficit environments, which represents the most comprehensive analysis of $G \times E$ interactions for WSC to date.

This study used the factor analytic mixed-model methods of Smith et al. (2001b) and Beeck et al. (2010) to partition environmental variance into hypothetical covariates.
Then, the clustering approach of Cullis et al. (2010) was used to identify clusters of experiments that performed similarly by determining the consistency of genetic correlations (line rankings) between experiments, and therefore represented similar environments. The methodology proved useful in exploring the G × E interactions present in the dataset, and influenced the approach to both the GWAS and genomic selection aspects of the thesis.

The computational difficulties of modelling QTL with factor analytic methods precluded using the clustering approach in the GWAS component of this study. Instead, the criterion of repeatability across experiments was used. It is interesting to note that the two QTL identified for WSCC were identified in the two water deficit environment experiments, and that these QTL were not found at any of the well-watered environment experiments. This adds weight to the necessity of characterising environments before embarking on marker-trait association studies as repeatability is key for the usefulness of associations within TPE, but not so important outside those target environments.

Genomic selection studies have approached G × E interactions in a range of ways. Some, including Sallam et al. (2015) attempted to minimise G × E interactions. Others treated G × E as a single variance component (including Asoro et al., 2011), and a number of studies have attempted to include specific covariates to explain some portion of G × E (Ly et al., 2013; Heslot et al., 2014). Factor analytic models have been applied in genomic selection studies previously (Burgueño et al., 2012; Rutkoski et al., 2015), however these studies average variety effects across experiments without reporting any consideration of differential performance between experiments. This is the first study that reports GEBVs for different target environments within the same model in an attempt to exploit G × E variation rather than negate it, as advanced
breeding programs do with phenotypic selection methods (Smith et al., 2015). Further work is needed to determine how best to utilise explainable G × E interactions to maximise the predictive ability of GEBVs for genomic selection.

### 7.4 Considerations on the methodology used in this study

Resource limitations and the vagaries of field trial experimentation often dictate compromise in the scope of the data set any plant geneticist has to work with, and this study is no different. Despite the attempts in these experiments to have terminal drought treatments in both years, the unseasonably wet conditions in 2010 (Figure 4.4) resulted in the characterisation of only two water deficit environments in this data set. A stronger data set would have a larger number of environments to ensure that the treatments used reliably performed as expected. This would involve further experimentation across sites and years.

The results of this study indicate the importance of the timing of measurement of WSC in the plant lifecycle, and highlight ways that measurement of WSC in future experiments can be refined. This study used measurements of WSC taken 180°Cd post-anthesis following Rebetzke et al. (2008), using the mean anthesis date in a population constrained for flowering time. Indeed, the results obtained indicate that sampling methods were successful in ensuring a limited impact of ZAD on WSCA and WSCC, justifying simultaneous sampling of whole experiments rather than plot-by-plot sampling. However the 2009 rainfed experiments experienced terminal drought beginning after anthesis (Figure 4.4), with concomitant acceleration of plant development compared to the irrigated experiments. Future work should seek further precision in the measurement of WSC relative to plant growth stage in contrasting well-watered and water deficit environments. A corollary to this is the need for more
detailed comparisons between WSC measurements taken simultaneously for a whole experiment and those taken relative to the development stage of individual plots to confirm that similar results are obtained.

An important interaction observed in the literature (van Herwaarden et al., 1998a; Ruuska et al., 2006), is the interaction between WSC accumulation and N content of the plant. This study did not investigate N content due to resource limitations. Further studies should seek to explore the relationship between N content, tillering and WSC accumulation, to determine the extent of genotypic control of sink development (other than the developing grain) that may prove a proxy for the genetic control of WSC accumulation, and the implications for grain size and grain yield. In particular, this study identified an association between WSCC and the DArT marker wPt-3743 close to the Glu-1D locus which is involved with the expression of glutenin grain storage proteins. This suggests that loci that are involved in the control of N relations in the plant may also play some role in determining the expression of WSC accumulation due to the interaction between these traits.

Genomic selection is a relatively new methodology that remains to be fully explored by plant breeders. Unfortunately, most genetics studies do not have the scope to extend over multiple breeding cycles and this study was no different. This constrained the validation methodology to the use of cross validation, where the same dataset is used to represent both the training and the prediction populations. A more robust approach would be to test GEBVs in progeny of genotypes used in the training population, to empirically determine the predictive ability of GEBVs. This principle should also apply to GWAS for complex traits if possible to assess the usefulness of individual markers in selection as predictors of phenotype.
7.5 Implications and applications of this thesis

This thesis should be seen in the context of some significant contributions to the understanding of the genetic control of WSC accumulation and complex traits in general, and some findings that highlight future areas of research. Based on the phenotypic correlations between WSCA and WSCC and other traits measured in Chapter 4 it is likely that there is much more to understand in the relationship between WSC accumulation, subsequent WSC remobilisation, grain number per m², grain size and grain yield. To better understand how WSC accumulation and remobilisation can contribute to yield under terminal water deficit it is important to consider WSCA and WSCC at a number of points in plant development. One point in plant development is when grain number per m² is determined, and then when grain size is determined so that the relative contribution by WSC to each of these yield determinants can be measured. This may go some way to explaining why this study did not identify any contribution to grain yield from WSC accumulation in either the well-watered or water deficit environments.

This study would be one of the first to report on the application of a positive control methodology in a crop GWAS. Much has been made of the limitations of GWAS, such as the nature of ‘missing heritability’ in GWAS studies (Manolio et al., 2009), and the many influencing factors such as population structure and linkage disequilibrium patterns (Myles et al., 2009). This study took an empirical approach to determining the credibility of the GWAS results used by testing the GWAS methods for the ability to detect major known flowering time loci as well as WSCC loci. Although the traits under consideration (ZAD and WSCC) differed, a key result is that no major loci for WSCC were detectable in this population. Further, two separate whole genome marker sets in GWAS are rarely used together with no previous reports
of SNP markers and DArT markers being separately and simultaneously applied to the same dataset. The differences noted in this study confirm marker type as a factor for consideration in the interpretation of future GWAS studies. The marker loci identified in this study in multiple experiments and across two analysis pipelines warrant further investigation, to confirm their suitability for application to marker-assisted selection for WSCC. In particular, the possibly pleiotropic effects of the locus Glu-1D near marker wPt-3743 need to be investigated, as does the contribution of the locus near wPt-9592 to yield performance under terminal drought.

This study is the first report describing the application of genomic selection methods in an Australian wheat dataset, and the first to investigate WSCC with genomic selection models anywhere. Genomic selection looks to be a promising means to efficiently select for traits such as WSC accumulation where many loci of minor effect are involved in genetic control and may prove more effective than phenotypic selection alone. Moreover, genomic selection could reduce the need for expensive and difficult phenotypic characterisation of traits like WSCC. The continuing decrease in the cost of whole genome marker profiles is likely to facilitate growing interest in genomic selection in coming years (Heffner et al., 2010), although comprehensive research will need to precede the application of this technology to breeding programs. A key consideration of genomic selection, as with phenotypic selection, is the characterisation of the TPE that the GEBVs will be utilised in. This study identified two distinct TPE for well-watered and water deficit environments and modelled GEBVs specific to each TPE. Future studies that explore the incorporation of G × E effects into genomic selection models will need to go further to assess the accuracy of GEBVs modelled for a specific TPE, preferably through actual progeny testing rather than cross validation.
8. Conclusions

This thesis reveals the benefits of using a combined approach with a range of tools available to the contemporary plant breeder to reveal the genetic architecture of WSC accumulation. Selection for quantitative traits is a challenging undertaking, evidenced by the complexity of the results from each of the experimental chapters. However the rewards from making incremental progress are substantial and improving drought avoidance in wheat is critically important to productivity gains in water-limited environments, including south-eastern Australia. The methods used in this thesis turn full circle in terms of the progression from classical quantitative genetics, to molecular marker based genome-wide association studies, and finally to the incorporation of molecular information into statistical genetics models for genomic selection.

Phenotyping for WSC is complex, and is one reason why study into this drought avoidance trait has been limited. Chapter 4 shows that characterisation of the TPE under which WSC is to be selected is critical if genetic gains are to be made, given the strong G × E interactions involved in the expression of this trait. The nature of G × E should be readily exploitable, given the consistent performance demonstrated within environmental groupings. The difficulty in measuring anthesis biomass (and subsequent reduction in heritability) offsets the usefulness of using WSCA as a selection target over WSCC. Surrogate traits such as DMC could be used to select for WSC, however DMC heritability is moderate, so doing so may not be justified in terms of genetic gain.

This study identified no correlation between grain yield and WSCC or WSCA, however consistent relationships were observed with other traits, such as lower
tillering, lower grains per m² and higher thousand grain weight with increasing WSCC. The challenges in moving from limited diversity studies to a larger-scale investigation of diversity for WSC in breeding germplasm is highlighted, with relationships in this study being relatively weaker than other reports (Ruuska et al., 2006; Rebetzke et al., 2008; Dreccer et al., 2009).

The investigation of GWAS for WSCC described in Chapter 5 identified loci associated with the expression of WSCC on chromosomes 1A and 1D that have not previously been reported. However, the markers identified in this study have been reported to be associated with other traits, in particular the marker $wPt$-3743 on 1D is close to the locus Glu-1D which controls expression of grain glutenin proteins (Payne, 1987). The marker $wPt$-9592 identified on 1A has been associated with yield performance in terminal drought environments (Crossa et al., 2007). The analysis of ZAD had two roles in this study. First, it provided a positive control by identifying the major known loci for flowering time as a positive control for the GWAS methods also used to identify WSCC loci, given the major loci controlling ZAD in Australian germplasm have been well characterised (Eagles et al., 2010). Secondly, the extent of colocation of markers for ZAD and WSCC was investigated, to determine if the major flowering time loci were responsible for variation in WSCC as reported by (Rebetzke et al., 2008). In this study several major flowering time loci were identified, but few ZAD loci co-located with WSCC loci. The GWAS results varied between experiments, and differences were observed between marker type and the analysis pipeline used. The GWAS results add further evidence to claims that WSCC is polygenically inherited (Rebetzke et al., 2008), with a clear absence of any major loci for the genetic control of WSCC.
The polygenic nature of WSCC makes this trait an interesting candidate for genomic selection (as detailed in Chapter 6). The inclusion of whole-genome marker effects in selection models has the potential to improve selection gains in breeding programs, as selection can be conducted early in the breeding cycle, and the actual breeding cycle can be shortened (Lorenz et al., 2011). Estimated relative accuracies of GEBVs calculated for WSCC show that genomic selection may out-perform conventional phenotypic selection by 30-40% if it can be coupled with a reduction in the breeding cycle time, and if the TPE for selection is well defined. Interestingly, the additive GEBV component of the genomic selection model was correlated with the non-additive genetic variance, so genetic selection gains may be enhanced by this interaction. The correlation between the additive and non-additive genetic effects for WSCC may indicate the presence of phenotypic epistasis, or genetic correlation between the traits that together result in the expression of WSCC (Cheverud & Routman, 1995; Holland, 2001a). Conclusions about the nature of the additive and non-additive genetic variance for WSCC need to be viewed in the context of the limitations around the interpretation of genome-wide marker effects. The additive variance in genomic selection models is an approximation from the marker effects, and not additive genetic effects in the biological sense (Zuk et al., 2012; Muñoz et al., 2014).

The expression of WSC accumulation is dependent on the balance between assimilate supply and demand in the plant and is related to other traits and environmental conditions (Rebetzke et al., 2009). By measuring WSC accumulation and related traits in a MET, this study was able to identify relationships between WSC and other traits and G × E interactions determining WSC expression. This facilitated the application of marker technologies to investigate marker-trait associations with WSC accumulation. The absence of detectable major loci for WSC in GWAS suggested that
genomic selection models could provide further insight into the genetic architecture of WSC, and the results of this study show that GEBVs may prove useful for selection of higher WSC-accumulating genotypes.

It is the use of multiple approaches to genetic analysis in this study that highlight the nature of complex trait investigation and improvement. In real terms, this study shows that a successful pathway is likely to lie in building genetic understanding of a trait through application of a suite of complimentary methods, which incorporates both attempts to understand the genetic architecture of the trait in question, and methods that can still result in genetic gain if detailed understanding of the nature of inheritance of the trait remains elusive. The corollary to this view is that major breeding breakthroughs have been historically underpinned by a comprehensive understanding of the physiological processes involved, so that robust and targeted selection methods can be implemented. This thesis is evidence that there is no single superior method for complex trait improvement, although greater understanding of the genetic control of a trait can come through the exploration of multiple approaches.
## Appendix

**Table A.1: Markers significant for WSCC at more than one site from WGAIM analysis for both the SNP and DArT marker sets.** Chromosome and position are from the consensus maps for each marker set. Nearby genes are from CMAP GrainGene database (http://wheat.pw.usda.gov/cmap/) searches

<table>
<thead>
<tr>
<th>Experiments</th>
<th>Chr</th>
<th>Marker</th>
<th>Dist(cM)</th>
<th>Size Range</th>
<th>P-value Range</th>
<th>% Var Range</th>
<th>LOD Range</th>
<th>Nearby Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>W9A09YANA_RFD, W9A10YANA_IRR</td>
<td>1A</td>
<td>wsnp_Ex_c6420_11162653</td>
<td>11.78</td>
<td>0.067 - 0.082</td>
<td>0 - 0</td>
<td>7 - 8.9</td>
<td>3.604 - 6.182</td>
<td></td>
</tr>
<tr>
<td>W9A09YANA_RFD</td>
<td>1A</td>
<td>wPt-9592</td>
<td>68.25</td>
<td>0.061 - 0.063</td>
<td>0 - 0</td>
<td>5 - 5.8</td>
<td>3.087 - 3.409</td>
<td></td>
</tr>
<tr>
<td>W9A09YANA_RFD, W9A10YANA_IRR</td>
<td>1A</td>
<td>wPt-5316</td>
<td>120.16</td>
<td>0.061 - 0.086</td>
<td>0 - 0.001</td>
<td>4 - 10.9</td>
<td>2.427 - 3.321</td>
<td>Glu-A1</td>
</tr>
<tr>
<td>W9A09COLE_IRR, W9A09COLE_IRR, W9A10YANA_IRR</td>
<td>1B</td>
<td>wPt-2751</td>
<td>2.81</td>
<td>-0.128 - -0.079</td>
<td>0 - 0.001</td>
<td>6.8 - 16.4</td>
<td>2.418 - 3.578</td>
<td>Glu-3Bh</td>
</tr>
<tr>
<td>W9A09COLE_IRR</td>
<td>1D</td>
<td>wPt-3743</td>
<td>83.27</td>
<td>0.074 - 0.083</td>
<td>0 - 0</td>
<td>7.1 - 8</td>
<td>3.654 - 3.785</td>
<td>Glu-D1</td>
</tr>
<tr>
<td>W9A10COLE_RFD</td>
<td>2A</td>
<td>wPt-9432</td>
<td>19.91</td>
<td>-0.099 - -0.08</td>
<td>0.001</td>
<td>7.7 - 9.2</td>
<td>2.422 - 2.545</td>
<td></td>
</tr>
<tr>
<td>Genes</td>
<td>Chromosome</td>
<td>Start/End</td>
<td>Description</td>
<td>Meiotic Frequency</td>
<td>Spatial Frequency</td>
<td>p-Value</td>
<td>q-Value</td>
<td>Correlation Coefficient</td>
</tr>
<tr>
<td>---------------</td>
<td>------------</td>
<td>-------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>------------------</td>
<td>---------</td>
<td>---------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>W9A09YANA_IRR, W9A10COLE_RFD</td>
<td>2B</td>
<td>21.07</td>
<td>tPt-4246</td>
<td>0.09 - 0.107</td>
<td>0.001 - 0.004</td>
<td>8.9 - 9</td>
<td>1.758 - 2.24</td>
<td></td>
</tr>
<tr>
<td>W9A09YANA_IRR, W9A10COLE_RFD</td>
<td>2B</td>
<td>47.74</td>
<td>wPt-8404</td>
<td>0.079 - 0.107</td>
<td>0 - 0.004</td>
<td>5 - 9</td>
<td>1.758 - 3.361</td>
<td></td>
</tr>
<tr>
<td>W9A09YANA_IRR, W9A10YANA_RFD</td>
<td>2B</td>
<td>165.61</td>
<td>wsnp_Ex_c17127_25756019</td>
<td>-0.066 - -</td>
<td>0 - 0</td>
<td>4.3 - 6.9</td>
<td>3.057 - 3.432</td>
<td></td>
</tr>
<tr>
<td>W9A09YANA_IRR, W9A10YANA_RFD</td>
<td>2D</td>
<td>67.32</td>
<td>wPt-6531</td>
<td>-0.083 - -</td>
<td>0 - 0</td>
<td>4 - 10.2</td>
<td>4.206 - 5.096</td>
<td></td>
</tr>
<tr>
<td>W9A09YANA_IRR, W9A10COLE_RFD</td>
<td>3B</td>
<td>57.14</td>
<td>wPt-4209</td>
<td>0.05 - 0.085</td>
<td>0 - 0.003</td>
<td>2.8 - 14.2</td>
<td>1.871 - 2.703</td>
<td></td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A09YANA_IRR, W9A10YANA_RFD</td>
<td>5B</td>
<td>125.89</td>
<td>wsnp_CAP11_c948_571287</td>
<td>0.087 - 0.106</td>
<td>0 - 0</td>
<td>6.1 - 12.9</td>
<td>4.349 - 6.254</td>
<td></td>
</tr>
</tbody>
</table>
Table A.2: Markers significant for ZAD at more than one site from WGAIM analysis for both the SNP and DArT marker sets. Chromosome and position are from the consensus maps for each marker set. Nearby genes are from CMAP GrainGene database (http://wheat.pw.usda.gov/cmap/) searches.

<table>
<thead>
<tr>
<th>Experiments</th>
<th>Chr</th>
<th>Marker</th>
<th>Dist(cM)</th>
<th>Size Range</th>
<th>P-value Range</th>
<th>% Var Range</th>
<th>LOD Range</th>
<th>Nearby Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>W9A09COLE_RFD,</td>
<td>1A</td>
<td>wsnp_Ex_c6420_11162653</td>
<td>11.78</td>
<td>0.568 - 0.993</td>
<td>0 - 0.001</td>
<td>0.7 - 2.4</td>
<td>2.362 - 7.986</td>
<td>W9A09YANA_IRR, W9A09YANA_RFD</td>
</tr>
<tr>
<td>W9A09YANA_I RR,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10COLE_IRR,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10COLE_RFD,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10YANA_IRR,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10YANA_RFD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09YANA_IRR,</td>
<td>1A</td>
<td>wPt-731843</td>
<td>16.69</td>
<td>0.716 - 0.937</td>
<td>0.001 - 0.005</td>
<td>1.2 - 1.5</td>
<td>1.75 - 2.305</td>
<td>W9A09YANA_RFD</td>
</tr>
<tr>
<td>W9A09YANA_RFD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09COLE_IRR,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09COLE_RFD,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09YANA_IRR,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10YANA_IRR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09YANA_RFD,</td>
<td>1A</td>
<td>wPt-4177</td>
<td>19.83</td>
<td>-0.1011 - -</td>
<td>0 - 0.595</td>
<td>1.1 - 1.9</td>
<td>2.107 - 4.462</td>
<td>W9A09YANA_RFD, W9A10COLE_RFD</td>
</tr>
<tr>
<td>W9A09YANA_RFD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09COLE_IRR,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09COLE_RFD,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09YANA_IRR,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10YANA_IRR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09YANA_RFD,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10COLE_RFD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09YANA_RFD,</td>
<td>1A</td>
<td>wsnp_Ra_c9209_15425473</td>
<td>91.52</td>
<td>0.479 - 0.572</td>
<td>0 - 0.017</td>
<td>0.4 - 1.1</td>
<td>1.227 - 3.318</td>
<td>W9A10COLE_RFD</td>
</tr>
<tr>
<td>W9A10COLE_RFD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

230
<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Amino Acid Level</th>
<th>Protein Level</th>
<th>Uncertainty</th>
<th>Protein Level</th>
<th>amino acid Level</th>
<th>Protein Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>W9A09YANA_RFD, W9A10COLE_RFD, W9A09COLE_IRR, W9A10YANA_RFD</td>
<td>1A</td>
<td>wsnp_Ex_c43228_49605281</td>
<td>104.02</td>
<td>0.641 - 1.311</td>
<td>0 - 0</td>
<td>1.2 - 2.3</td>
</tr>
<tr>
<td>W9A10COLE_RFD, W9A09COLE_IRR</td>
<td>1A</td>
<td>wsnp_Ex_c3258_6004611</td>
<td>104.35</td>
<td>1.133 - 1.387</td>
<td>0 - 0</td>
<td>2.7 - 3.0</td>
</tr>
<tr>
<td>W9A09YANA_RFD, W9A09COLE_RFD</td>
<td>1A</td>
<td>wPt-2847</td>
<td>117.16</td>
<td>-0.891 - -</td>
<td>0 - 0</td>
<td>0.8 - 2.4</td>
</tr>
<tr>
<td>W9A09YANA_RFD, W9A10YANA_RFD</td>
<td>1A</td>
<td>wsnp(ID) _rep_c49006_33254974</td>
<td>150.14</td>
<td>-0.758 - -</td>
<td>0 - 0</td>
<td>0.7 - 1.2</td>
</tr>
<tr>
<td>W9A09YANA_RFD, W9A10COLE_RFD</td>
<td>1B</td>
<td>wPt-2575</td>
<td>23.09</td>
<td>0.612 - 1.059</td>
<td>0 - 0</td>
<td>1.0 - 1.9</td>
</tr>
<tr>
<td>W9A10YANA_RFD, W9A09COLE_IRR</td>
<td>1B</td>
<td>wPt-741274</td>
<td>24.4</td>
<td>-2.177 - -</td>
<td>0 - 0</td>
<td>3.4 - 8.0</td>
</tr>
<tr>
<td>W9A10YANA_RFD, W9A09COLE_IRR</td>
<td>1B</td>
<td>wsnp_Ex_c5947_10431109</td>
<td>36.48</td>
<td>-1.480 - -</td>
<td>0 - 0</td>
<td>3.5 - 6.4</td>
</tr>
<tr>
<td>W9A10YANA_RFD, W9A09YANA_RFD</td>
<td>1B</td>
<td>wsnp_Ex_c5388_9526777</td>
<td>50.26</td>
<td>-0.651 - -</td>
<td>0 - 0</td>
<td>0.7 - 0.9</td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A10COLE_RFD</td>
<td>1B</td>
<td>wPt-5034</td>
<td>83.82</td>
<td>-0.605 - -</td>
<td>0.006 - 0.006</td>
<td>0.6 - 0.6</td>
</tr>
<tr>
<td>Genes</td>
<td>Chromosome</td>
<td>Marker</td>
<td>FDR 0.05</td>
<td>FDR 0.1</td>
<td>FDR 0.5</td>
<td>FDR 1</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>------------</td>
<td>--------</td>
<td>----------</td>
<td>---------</td>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>W9A09COLE_IRR, W9A09COLE_RFD,</td>
<td>1D</td>
<td>wPt-4671</td>
<td>66.98</td>
<td>0.716 - 1.247</td>
<td>0 - 0</td>
<td>1.5 - 2.7</td>
</tr>
<tr>
<td>W9A09YANA_IRR, W9A09YANA_RFD,</td>
<td>2A</td>
<td>tPt-1041</td>
<td>20.09</td>
<td>0.385 - 0.655</td>
<td>0.001 - 0.014</td>
<td>0.5 - 0.9</td>
</tr>
<tr>
<td>W9A09YANA_IRR, W9A10YANA_RFD,</td>
<td>2A</td>
<td>wsnp_Ex_rep_c93362_82371891</td>
<td>33.43</td>
<td>-0.888 - -</td>
<td>0.001 - 0.014</td>
<td>1.7 - 1.7</td>
</tr>
<tr>
<td>W9A09YANA_IRR, W9A10COLE_RFD,</td>
<td>2A</td>
<td>wPt-7011</td>
<td>86.14</td>
<td>-0.889 - -</td>
<td>0 - 0.001 - 0.014</td>
<td>0.7 - 0.8</td>
</tr>
<tr>
<td>W9A10COLE_RFD, W9A09COLE_RFD,</td>
<td>2A</td>
<td>wsnp_Ra_c107797_91270622</td>
<td>150.73</td>
<td>0.495 - 0.887</td>
<td>0 - 0.001 - 0.014</td>
<td>0.7 - 1.1</td>
</tr>
<tr>
<td>W9A10COLE_RFD, W9A10COLE_IRR,</td>
<td>2A</td>
<td>wsnp_Ex_c15681_24015303</td>
<td>190.67</td>
<td>-0.666 - -</td>
<td>0.001 - 0.014</td>
<td>0.6 - 0.6</td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A09YANA_RFD,</td>
<td>2A</td>
<td>wsnp_Ex_c10555_17236072</td>
<td>202.49</td>
<td>-1.363 - -</td>
<td>0 - 0.001 - 0.014</td>
<td>1.3 - 2.6</td>
</tr>
<tr>
<td></td>
<td>Allele</td>
<td>GPAC</td>
<td>PIV</td>
<td>95% CI</td>
<td>t</td>
<td>P</td>
</tr>
<tr>
<td>----</td>
<td>------------</td>
<td>------</td>
<td>------</td>
<td>------------</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>2B</td>
<td>wPt-3949</td>
<td>14.59</td>
<td>0.558</td>
<td>0.7 0 - 1.0</td>
<td>2.307 -</td>
<td>2.792</td>
</tr>
<tr>
<td>2B</td>
<td>wPt-8404</td>
<td>47.74</td>
<td>0.58</td>
<td>1.018 0 - 0.008</td>
<td>1.508 -</td>
<td>3.849</td>
</tr>
<tr>
<td>2B</td>
<td>wsnp_Ex_c5239_9272511</td>
<td>117.14</td>
<td>0.549</td>
<td>0.853 0 - 0.001</td>
<td>2.585 -</td>
<td>4.56</td>
</tr>
<tr>
<td>2B</td>
<td>wPt-741530</td>
<td>126.91</td>
<td>1.035</td>
<td>1.151 0 - 0.002</td>
<td>2.004 -</td>
<td>2.99</td>
</tr>
<tr>
<td>2B</td>
<td>wsnp_Ku_c48694_54811376</td>
<td>219.85</td>
<td>0.855</td>
<td>0.99 0 - 0.0</td>
<td>3.521 -</td>
<td>4.004</td>
</tr>
<tr>
<td>2D</td>
<td>wPt-9780</td>
<td>43.74</td>
<td>0.575</td>
<td>0.828 0 - 0.001</td>
<td>2.328 -</td>
<td>5.109</td>
</tr>
<tr>
<td>Genotype</td>
<td>Feature</td>
<td>Value_1</td>
<td>Value_2</td>
<td>Value_3</td>
<td>Value_4</td>
<td>P-value_1</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>W9A10COLE_IRR, W9A10COLE_RFD, W9A10YANA_RFD</td>
<td>wPt-6847</td>
<td>80.31</td>
<td>-0.953</td>
<td>-</td>
<td>0</td>
<td>2.2</td>
</tr>
<tr>
<td>W9A09COLE_IRR, W9A10COLE_IRR</td>
<td>wPt-733227</td>
<td>84.54</td>
<td>-1.076</td>
<td>-</td>
<td>0</td>
<td>2.0</td>
</tr>
<tr>
<td>W9A10COLE_IRR, W9A10YANA_RFD</td>
<td>wsnp_Ex_c9619_15913632</td>
<td>159.28</td>
<td>-0.725</td>
<td>-</td>
<td>0</td>
<td>1.2</td>
</tr>
<tr>
<td>W9A09YANA_RFD, W9A10COLE_RFD, W9A10YANA_IRR</td>
<td>wsnp_Ku_c4319_7847988</td>
<td>168.02</td>
<td>0.657</td>
<td>0.729</td>
<td>0</td>
<td>0.8</td>
</tr>
<tr>
<td>Lines</td>
<td>Catalog Number</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
<td>Value 4</td>
<td>Value 5</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>----------------------</td>
<td>-----------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A09YANA_RFD, W9A10COLE_RFD, W9A10YANA_IRR, W9A10YANA_RFD</td>
<td>wsnp_Ex_c15674_24005648</td>
<td>27.55</td>
<td>0.562 - 0.899</td>
<td>0 - 0.003</td>
<td>0.7 - 2.4</td>
<td>1.886 - 5.306</td>
</tr>
<tr>
<td>W9A09YANA_IRR, W9A09YANA_RFD, W9A10COLE_RFD, W9A10YANA_IRR, W9A10YANA_RFD</td>
<td>wPt-9634</td>
<td>40.72</td>
<td>0.607 - 0.875</td>
<td>0 - 0.002</td>
<td>1.2 - 1.9</td>
<td>2.107 - 3.394</td>
</tr>
<tr>
<td>W9A10YANA_IRR, W9A10YANA_RFD, W9A10COLE_RFD, W9A10YANA_IRR, W9A10YANA_RFD</td>
<td>wsnp_CAP11_rep_c8581_3702222</td>
<td>160.5</td>
<td>0.594 - 0.684</td>
<td>0 - 0</td>
<td>0.8 - 1.4</td>
<td>3.017 - 4.353</td>
</tr>
<tr>
<td>W9A10YANA_IRR, W9A10YANA_RFD, W9A10COLE_RFD, W9A10YANA_IRR, W9A10YANA_RFD</td>
<td>wsnp_Ex_c8386_14127329</td>
<td>14.49</td>
<td>0.671 - 0.831</td>
<td>0 - 0</td>
<td>1.1 - 1.9</td>
<td>2.927 - 4.12</td>
</tr>
<tr>
<td>W9A10YANA_IRR, W9A10YANA_RFD, W9A10COLE_RFD, W9A10YANA_IRR, W9A10YANA_RFD</td>
<td>wPt-1158</td>
<td>57.77</td>
<td>0.668 - 1.086</td>
<td>0 - 0.01</td>
<td>0.8 - 2.1</td>
<td>1.434 - 4.046</td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A09YANA_IRR, W9A09YANA_RFD, W9A10COLE_RFD, W9A10YANA_IRR, W9A10YANA_RFD</td>
<td>wPt-741571</td>
<td>65.46</td>
<td>0.708 - 0.994</td>
<td>0 - 0.001</td>
<td>1.3 - 2.0</td>
<td>2.389 - 3.799</td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A09YANA_IRR, W9A09YANA_RFD, W9A10COLE_RFD, W9A10YANA_IRR, W9A10YANA_RFD</td>
<td>wsnp_JD_c2146_2955972</td>
<td>68.71</td>
<td>-1.383 -</td>
<td>0 - 0</td>
<td>2.5 - 2.7</td>
<td>4.24 - 5.265</td>
</tr>
<tr>
<td>WT</td>
<td>Allele</td>
<td>Genotype</td>
<td>Genotype 1</td>
<td>Genotype 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>----------</td>
<td>----------</td>
<td>------------</td>
<td>------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09COLE_IRR, W9A09YANA_IRR, W9A10COLE_IRR, W9A10COLE_RFD, W9A10YANA_IRR, W9A10YANA_RFD</td>
<td>3B</td>
<td>wsnp_Ex_c56591_58653455</td>
<td>178.61</td>
<td>-0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genotypes</td>
<td>Marker</td>
<td>Peak</td>
<td>p-Value (tSNP)</td>
<td>Mean</td>
<td>95% CI</td>
<td>Median</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>-------------</td>
<td>---------</td>
<td>----------------</td>
<td>----------</td>
<td>----------</td>
<td>---------</td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A09YANA_IRR, W9A09YANA_RFD</td>
<td>wPt-2780</td>
<td>90.6</td>
<td>0.035 - 0.792</td>
<td>0.001</td>
<td>0.5 - 1.1</td>
<td>1.326</td>
</tr>
<tr>
<td>W9A10YANA_RFD</td>
<td>wPt-4280</td>
<td>14.98</td>
<td>-0.917 - -</td>
<td>0 -</td>
<td>1.6 - 2.5</td>
<td>2.203</td>
</tr>
<tr>
<td>W9A09YANA_RFD, W9A10COLE_RFD, W9A10YANA_IRR</td>
<td>wsnp_CAP7_c1723_854530</td>
<td>24.8</td>
<td>0.706 - 0.89</td>
<td>0 - 0</td>
<td>1.4 - 1.6</td>
<td>3.628</td>
</tr>
<tr>
<td>W9A09COLE_IRR, W9A10COLE_RFD, W9A10YANA_IRR, W9A10YANA_RFD</td>
<td>wPt-1400</td>
<td>38.17</td>
<td>0.877 - 1.113</td>
<td>0 - 0</td>
<td>1.7 - 4.1</td>
<td>3.39</td>
</tr>
<tr>
<td>W9A09YANA_IRR, W9A09YANA_RFD, W9A10COLE_RFD, W9A10YANA_RFD, W9A10YANA_IRR</td>
<td>wsnp_BE422566B_Ta_1_1</td>
<td>43.3</td>
<td>1.059 - 1.42</td>
<td>0 -</td>
<td>2.1 - 3.6</td>
<td>1.927</td>
</tr>
<tr>
<td>Chromosome</td>
<td>Marker</td>
<td>Name</td>
<td>Value1</td>
<td>Value2</td>
<td>Value3</td>
<td>Value4</td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>------------------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>4B</td>
<td></td>
<td>wsnp_Ex_c23638_32875196</td>
<td>74.59</td>
<td>0.94 - 1.609</td>
<td>0 - 0</td>
<td>2.9 - 6.1</td>
</tr>
<tr>
<td>4B</td>
<td></td>
<td>wPt-3917</td>
<td>104.93</td>
<td>-1.031 - -</td>
<td>0 -</td>
<td>1.0 - 1.8</td>
</tr>
<tr>
<td>4B</td>
<td></td>
<td>wsnp_BE403378B_Ta_2_1</td>
<td>116.41</td>
<td>0.884 - 1.456</td>
<td>0 - 0</td>
<td>1.7 - 5.8</td>
</tr>
<tr>
<td>4D</td>
<td></td>
<td>wsnp_BF470532D_Ta_2_1</td>
<td>47.88</td>
<td>0.657 - 1.236</td>
<td>0 - 0</td>
<td>0.9 - 2.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>---</td>
<td>-------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>
| W9A09YANA_IRR,    | 4D| wPt-3058 | 72.85 | -0.996 | 0 | 1.4 - 3.3 | 2.05 -
| W9A09YANA_RFD,    |   |         | 0.659 | 0.002 |   | 4.472 |
| W9A10COLE_IRR,    |   |         |       |       |   | 2.748 |
| W9A10COLE_RFD,    |   |         |       |       |   |       |
| W9A10YANA_IRR,    |   |         |       |       |   | 1.949 |
| W9A10YANA_RFD,    |   |         |       |       |   | 5.125 |
| W9A09COLE_RFD,    | 5A| wsnp_Ku_rep_c72362_72059764 | 6.98 | -0.759 | 0 | 0.7 - 0.8 | 2.259 -
| W9A09YANA_IRR,    |   |         | 0.635 | 0.001 |   | 2.748 |
| W9A09COLE_IRR,    | 5A| wsnp_Ex_c26968_36188964 | 15.6 | -0.823 | 0 | 0.6 - 1.9 | 1.949 -
| W9A09YANA_IRR,    |   |         | 0.436 | 0.003 |   | 5.125 |
| W9A10COLE_IRR,    |   |         |       |       |   |       |
| W9A10COLE_RFD,    |   |         |       |       |   |       |
| W9A10YANA_IRR,    |   |         |       |       |   |       |
| W9A10YANA_RFD,    |   |         |       |       |   |       |
| W9A09COLE_IRR,    | 5A| wsnp_Ex_c5978_10478584 | 57.95 | 0.613 - 0.843 | 0 | 0.9 - 1.2 | 3.556 -
<p>| W9A10COLE_RFD,    |   |         | 3.779 |
| W9A09COLE_IRR,    | 5A| wsnp_Ex_c31799_40545376 | 69.07 | 1.571 - 3.171 | 0 | 7.7 - |
| W9A09YANA_IRR,    |   |         | 14.3  | 24.38 - |
| W9A09YANA_RFD,    |   |         | 54.182 |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>W9A09COLE_RFD,</td>
<td>5A</td>
<td>wsnp_Ex_c9927_16346100</td>
<td>85.01</td>
<td>-0.791</td>
<td>-</td>
<td>0</td>
<td>0.7 - 1.1</td>
</tr>
<tr>
<td>W9A09YANA_IRR,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09YANA_RFD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10COLE_IRR,</td>
<td>5A</td>
<td>wPt-1903</td>
<td>112.42</td>
<td>0.52 - 0.739</td>
<td>0</td>
<td>0.7 - 1.4</td>
<td>1.996 - 4.07</td>
</tr>
<tr>
<td>W9A10COLE_RFD,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10YANA_IRR,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10YANA_RFD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09COLE_RFD,</td>
<td>5A</td>
<td>wsnp_Ex_rep_c67292_65834396</td>
<td>127.25</td>
<td>-0.726</td>
<td>-</td>
<td>0</td>
<td>0.7 - 1.1</td>
</tr>
<tr>
<td>W9A09YANA_IRR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10YANA_IRR,</td>
<td>5B</td>
<td>wPt-9925</td>
<td>17.3</td>
<td>-0.705 - -</td>
<td>0</td>
<td>0</td>
<td>1.2 - 1.6</td>
</tr>
<tr>
<td>W9A10YANA_RFD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09YANA_IRR,</td>
<td>5B</td>
<td>rPt-6127</td>
<td>27.31</td>
<td>-1.393 - -</td>
<td>0</td>
<td>0</td>
<td>3.3 - 3.4</td>
</tr>
<tr>
<td>W9A09YANA_RFD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09YANA_IRR,</td>
<td>5B</td>
<td>wPt-665267</td>
<td>28.23</td>
<td>1.093 - 1.253</td>
<td>0</td>
<td>0</td>
<td>2.1 - 3.6</td>
</tr>
<tr>
<td>W9A09YANA_RFD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09COLE_IRR,</td>
<td>5B</td>
<td>wPt-1149</td>
<td>28.42</td>
<td>0.542 - 1.236</td>
<td>0</td>
<td>0</td>
<td>0.9 - 2.6</td>
</tr>
<tr>
<td>W9A09COLE_RFD,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09YANA_IRR,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09YANA_RFD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10COLE_IRR,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10COLE_RFD,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10YANA_IRR,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10YANA_RFD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accession Numbers</td>
<td>Chromosome</td>
<td>Value</td>
<td>Description</td>
<td>0.581 - 0.719</td>
<td>0 - 0.001</td>
<td>0.8 - 1.3</td>
<td>2.321 - 3.433</td>
</tr>
<tr>
<td>---------------------------</td>
<td>------------</td>
<td>-------</td>
<td>-------------</td>
<td>----------------</td>
<td>-----------</td>
<td>-----------</td>
<td>---------------</td>
</tr>
<tr>
<td>W9A09YANA_IRR, W9A10COLE_RFD, W9A10YANA_IRR</td>
<td>5B</td>
<td>31.03</td>
<td>wsnp_BE446509B_Ta_2_6</td>
<td>_</td>
<td>0.581 - 0.719</td>
<td>0 - 0.001</td>
<td>0.8 - 1.3</td>
</tr>
<tr>
<td>W9A10COLE_IRR, W9A10YANA_IRR</td>
<td>5B</td>
<td>44.84</td>
<td>wPt-9467</td>
<td>0.462 - 0.816</td>
<td>0 - 0.003</td>
<td>0.7 - 2.0</td>
<td>1.929 - 5.279</td>
</tr>
<tr>
<td>W9A09COLE_IRR, W9A09COLE_RFD, W9A10COLE_IRR, W9A10YANA_IRR, W9A10YANA_RFD</td>
<td>5B</td>
<td>86.06</td>
<td>wsnp_Ex_c210_411604</td>
<td>-1.535 - -</td>
<td>0 - 0.792</td>
<td>0 - 0.013</td>
<td>1.5 - 3.3</td>
</tr>
<tr>
<td>W9A09COLE_IRR, W9A09YANA_IRR, W9A09YANA_RFD</td>
<td>5B</td>
<td>93.64</td>
<td>wPt-1733</td>
<td>-1.172 - -</td>
<td>0 - 1.071</td>
<td>0 - 0.071</td>
<td>2.0 - 2.6</td>
</tr>
<tr>
<td>W9A09COLE_IRR, W9A09YANA_IRR, W9A09YANA_RFD</td>
<td>5B</td>
<td>93.74</td>
<td>wPt-5896</td>
<td>-1.295 - -</td>
<td>0 - 0.566</td>
<td>0 - 0.005</td>
<td>0.7 - 2.8</td>
</tr>
<tr>
<td>W9A10COLE_IRR, W9A10YANA_IRR, W9A10YANA_RFD</td>
<td>5B</td>
<td>98.54</td>
<td>wPt-2707</td>
<td>-1.079 - -</td>
<td>0 - 0.814</td>
<td>0 - 0.005</td>
<td>2.0 - 3.9</td>
</tr>
<tr>
<td>Genotypes</td>
<td>Chromosome</td>
<td>Accession</td>
<td>Score</td>
<td>Effect Size</td>
<td>p-value</td>
<td>CI Low</td>
<td>CI High</td>
</tr>
<tr>
<td>-----------</td>
<td>------------</td>
<td>-----------</td>
<td>-------</td>
<td>-------------</td>
<td>---------</td>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A10YANA_RFD</td>
<td>5B</td>
<td>wsnp_Ex_rep_c70120_69070056</td>
<td>109.16</td>
<td>0.71 - 1.1</td>
<td>0.001</td>
<td>1.2 - 2.5</td>
<td>2.357 - 4.606</td>
</tr>
<tr>
<td>W9A10YANA_RFD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10YANA_RFD</td>
<td>5B</td>
<td>wsnp_BE495277B_Ta_2_5</td>
<td>125.3</td>
<td>-1.021 - -</td>
<td>0.737</td>
<td>1.3 - 3.0</td>
<td>2.967 - 6.364</td>
</tr>
<tr>
<td>W9A10YANA_RFD</td>
<td>5B</td>
<td>wsnp_ID_c1845_2564931</td>
<td>210.71</td>
<td>-0.783 - -</td>
<td>0.54</td>
<td>0.7 - 0.9</td>
<td>2.778 - 3.441</td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A09YANA_RFD, W9A10YANA_RFD</td>
<td>5D</td>
<td>wsnp_Ex_c508_1008029</td>
<td>61.02</td>
<td>0.863 - 1.645</td>
<td>0.002</td>
<td>1.9 - 3.7</td>
<td>3.933 - 10.456</td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A09YANA_RFD, W9A10YANA_RFD</td>
<td>5D</td>
<td>wPt-5505</td>
<td>85.34</td>
<td>-0.897 - -</td>
<td>0.774</td>
<td>1.8 - 2.7</td>
<td>2.121 - 3.253</td>
</tr>
<tr>
<td>Sample Group</td>
<td>Type</td>
<td>Gene Accession</td>
<td>Value</td>
<td>Base Value 1</td>
<td>Base Value 2</td>
<td>Base Value 3</td>
<td>Base Value 4</td>
</tr>
<tr>
<td>--------------</td>
<td>------</td>
<td>----------------</td>
<td>-------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
</tr>
<tr>
<td>W9A09YANA_RFD, W9A10COLE_IRR, W9A10COLE_RFD, W9A10YANA_IRR</td>
<td>6A</td>
<td>wPt-741630</td>
<td>3.22</td>
<td>0.783 - 1.148</td>
<td>0 - 0</td>
<td>1.9 - 3.5</td>
<td>3.199 - 5.685</td>
</tr>
<tr>
<td>W9A09COLE_IRR, W9A09YANA_RFD, W9A10COLE_IRR, W9A10YANA_IRR</td>
<td>6A</td>
<td>wPt-7616</td>
<td>14.89</td>
<td>1.007 - 1.368</td>
<td>0 - 0</td>
<td>2.1 - 4.3</td>
<td>2.645 - 4.706</td>
</tr>
<tr>
<td>W9A09COLE_IRR, W9A09YANA_RFD, W9A09COLE_RFD, W9A10YANA_IRR</td>
<td>6A</td>
<td>wPt-732263</td>
<td>48.29</td>
<td>0.812 - 2.099</td>
<td>0 - 0</td>
<td>1.5 - 7.5</td>
<td>1.485 - 6.837</td>
</tr>
<tr>
<td>W9A09COLE_IRR, W9A09YANA_RFD</td>
<td>6A</td>
<td>wsnp_Ex_c21688_30847181</td>
<td>211.14</td>
<td>-0.922 - 0.753</td>
<td>0 - 0</td>
<td>1.0 - 1.1</td>
<td>3.434 - 4.984</td>
</tr>
<tr>
<td>W9A09COLE_IRR, W9A09YANA_RFD</td>
<td>6B</td>
<td>wPt-7642</td>
<td>6.57</td>
<td>0.87 - 1.078</td>
<td>0.001 - 0.004</td>
<td>1.3 - 2.0</td>
<td>1.818 - 2.536</td>
</tr>
<tr>
<td>W9A10YANA_RFD</td>
<td>6B</td>
<td>wPt-8015</td>
<td>7.72</td>
<td>-0.961 - 0.841</td>
<td>0 - 0</td>
<td>2.3 - 2.8</td>
<td>4.753 - 5.62</td>
</tr>
<tr>
<td>W9A09YANA_RFD</td>
<td>6B</td>
<td>wPt-1756</td>
<td>12.85</td>
<td>-0.87 - 0.792</td>
<td>0 - 0</td>
<td>1.6 - 2.5</td>
<td>4.07 - 5.934</td>
</tr>
<tr>
<td>Accession</td>
<td>Chromosome</td>
<td>Feature Description</td>
<td>Start</td>
<td>End</td>
<td>Length</td>
<td>Score</td>
<td>Significance</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>---------------------</td>
<td>-------</td>
<td>-----</td>
<td>--------</td>
<td>-------</td>
<td>--------------</td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A09YANA_RFD, W9A10COLE_RFD, W9A10YANA_RFD</td>
<td>6B</td>
<td>wsnp_Ex_c2936_5416717</td>
<td>30.68</td>
<td>0.761 - 1.629</td>
<td>0.002</td>
<td>1.2 - 4.1</td>
<td>2.112 - 6.908</td>
</tr>
<tr>
<td>W9A09YANA_RFD, W9A10COLE_RFD, W9A10YANA_RFD</td>
<td>6B</td>
<td>wPt-665704</td>
<td>49.32</td>
<td>-1.61</td>
<td>0</td>
<td>2.2 - 8.6</td>
<td>1.536 - 5.868</td>
</tr>
<tr>
<td>W9A09YANA_RFD, W9A10COLE_RFD, W9A10YANA_RFD</td>
<td>6B</td>
<td>wPt-729979</td>
<td>51.9</td>
<td>0.489 - 0.918</td>
<td>0.003</td>
<td>0.8 - 1.9</td>
<td>1.952 - 5.105</td>
</tr>
<tr>
<td>W9A09YANA_RFD, W9A10YANA_RFD</td>
<td>6B</td>
<td>tPt-4990</td>
<td>64.46</td>
<td>1.018 - 1.09</td>
<td>0.0</td>
<td>1.8 - 2.1</td>
<td>3.113 - 3.917</td>
</tr>
<tr>
<td>W9A09YANA_RFD, W9A10YANA_RFD</td>
<td>6B</td>
<td>wsnp_Ex_c29387_38429452</td>
<td>94.36</td>
<td>-0.722</td>
<td>0</td>
<td>0.7 - 1.4</td>
<td>2.624 - 3.04</td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A10COLE_RFD</td>
<td>6D</td>
<td>wPt-665675</td>
<td>108.56</td>
<td>-0.649</td>
<td>0.003</td>
<td>0.6 - 0.7</td>
<td>1.781 - 1.952</td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A10COLE_RFD</td>
<td>7A</td>
<td>wPt-1080</td>
<td>65.06</td>
<td>0.573 - 0.626</td>
<td>0.004</td>
<td>0.6 - 0.7</td>
<td>1.432 - 1.876</td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A10COLE_RFD</td>
<td>7A</td>
<td>wPt-6620</td>
<td>122.34</td>
<td>0.618 - 0.831</td>
<td>0.002</td>
<td>1.0 - 1.2</td>
<td>2.067 - 2.096</td>
</tr>
<tr>
<td></td>
<td></td>
<td>tPt-9518</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>----</td>
<td>-----------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>W9A09COLE_IRR,</td>
<td>7A</td>
<td>W9A09COLE_RFD,</td>
<td>153.32</td>
<td>0.649 - 1.226</td>
<td>0 - 0.001</td>
<td>1.4 - 2.6</td>
<td>2.407 - 5.292</td>
</tr>
<tr>
<td>W9A09COLE_IRR,</td>
<td>7A</td>
<td>W9A09COLE_RFD,</td>
<td>153.43</td>
<td>-0.996 - -</td>
<td>0 - 0</td>
<td>1.6 - 1.7</td>
<td>3.686 - 4.018</td>
</tr>
<tr>
<td>W9A09COLE_IRR,</td>
<td>7A</td>
<td>W9A09COLE_RFD,</td>
<td>158.36</td>
<td>-1.259 - -</td>
<td>0 - 0</td>
<td>1.8 - 3.2</td>
<td>1.9 - 3.618</td>
</tr>
<tr>
<td>W9A09COLE_IRR,</td>
<td>7A</td>
<td>W9A09COLE_RFD,</td>
<td>12.9</td>
<td>-0.898 - -</td>
<td>0 - 0</td>
<td>0.8 - 2.0</td>
<td>2.57 - 6.032</td>
</tr>
<tr>
<td>W9A09COLE_IRR,</td>
<td>7B</td>
<td>W9A09COLE_RFD,</td>
<td>26.03</td>
<td>-0.893 - -</td>
<td>0 - 0</td>
<td>0.7 - 1.7</td>
<td>1.662 - 4.514</td>
</tr>
<tr>
<td>W9A09COLE_IRR,</td>
<td>7B</td>
<td>W9A09COLE_RFD,</td>
<td>59.2</td>
<td>-0.892 - -</td>
<td>0 - 0</td>
<td>1.2 - 1.9</td>
<td>3.659 - 4.652</td>
</tr>
<tr>
<td></td>
<td>Chromosome</td>
<td>Description</td>
<td>Depth</td>
<td>Min - Max</td>
<td>Peak Width</td>
<td>Peak Width Start - Peak Width End</td>
<td>Peak Width Start</td>
</tr>
<tr>
<td>-------------------------</td>
<td>------------</td>
<td>-----------------------------------------------</td>
<td>-------</td>
<td>-----------</td>
<td>------------</td>
<td>-----------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>W9A09COLE_IRR,</td>
<td>7B</td>
<td><code>wsnp_Ex_rep_c109138_92064554</code></td>
<td>78.98</td>
<td>0.85 - 1.144</td>
<td>0 - 0</td>
<td>1.7 – 2.0</td>
<td>3.252 - 1.144</td>
</tr>
<tr>
<td>W9A10COLE_IRR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10COLE_RFD,</td>
<td>7B</td>
<td><code>wsnp_ID_c2701_3626787</code></td>
<td>95.35</td>
<td>0.564 - 0.818</td>
<td>0 - 0.02</td>
<td>0.7 - 2.2</td>
<td>1.173 - 0.818</td>
</tr>
<tr>
<td>W9A10YANA_IRR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09COLE_IRR,</td>
<td>7B</td>
<td><code>wsnp_RFL_Contig3817_4141066</code></td>
<td>105.21</td>
<td>0.759 - 1.074</td>
<td>0 - 0</td>
<td>1.2 - 1.8</td>
<td>3.42 - 1.074</td>
</tr>
<tr>
<td>W9A09COLE_RFD,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09YANA_IRR,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09YANA_RFD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09YANA_IRR,</td>
<td>7B</td>
<td><code>wPt-3093</code></td>
<td>201.61</td>
<td>0.684 - 0.742</td>
<td>0.008 - 0.01</td>
<td>1.0 - 1.1</td>
<td>1.423 - 0.742</td>
</tr>
<tr>
<td>W9A09YANA_RFD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10COLE_IRR,</td>
<td>7D</td>
<td><code>wPt-2565</code></td>
<td>1.37</td>
<td>-1.27 - -</td>
<td>0 - 0</td>
<td>4.1 - 4.8</td>
<td>4.007 - 1.27</td>
</tr>
<tr>
<td>W9A10COLE_RFD,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10YANA_IRR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table A.3: Markers significant for WSCC at more than one site from the GAPIT analysis for the DArT marker set. No SNP markers were significant for a FDR adjusted p-value <0.10 for site by site GAPIT analyses. Chromosome and position are from the consensus map. Nearby genes are from CMAP GrainGene database (http://wheat.pw.usda.gov/cmap/) searches. GAPIT analyses returned significant markers for the two water deficit environments 2009 Coleambally and Yanco Rainfed, and not any of the well-watered environment experiments.

<table>
<thead>
<tr>
<th>Experiments</th>
<th>Marker</th>
<th>Chr</th>
<th>Dist(cM)</th>
<th>P-value Range</th>
<th>Minor Allele Frequency</th>
<th>FDR Adjusted P-values Range</th>
<th>Nearby Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>W9A09COLE_RFD, W9A09YANA_RFD</td>
<td>wPt-9592</td>
<td>1A</td>
<td>68.2502</td>
<td>0.000520 - 0.000789</td>
<td>0.491</td>
<td>0.0725 - 0.0999</td>
<td></td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A09YANA_RFD</td>
<td>wPt-665784</td>
<td>1A</td>
<td>69.1966</td>
<td>0.000557 - 0.0006277</td>
<td>0.474</td>
<td>0.0725 - 0.0923</td>
<td></td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A09YANA_RFD</td>
<td>wPt-7359</td>
<td>1B</td>
<td>11.2844</td>
<td>6.97×10⁻⁵ - 9.73×10⁻⁵</td>
<td>0.107</td>
<td>0.0228 - 0.0323</td>
<td></td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A09YANA_RFD</td>
<td>wPt-666719</td>
<td>1D</td>
<td>83.0816</td>
<td>1.56×10⁻⁵ - 2.97×10⁻⁵</td>
<td>0.278</td>
<td>0.00784 - 0.0193</td>
<td>Glu-D1</td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A09YANA_RFD</td>
<td>wPt-3743</td>
<td>1D</td>
<td>83.2698</td>
<td>1.58×10⁻⁵ - 3.88×10⁻⁵</td>
<td>0.275</td>
<td>0.00784 - 0.0193</td>
<td>Glu-D1</td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A09YANA_RFD</td>
<td>wPt-733835</td>
<td>1D</td>
<td>86.4719</td>
<td>0.000205 - 9.16E×10⁻⁵</td>
<td>0.285</td>
<td>0.0228 - 0.0510</td>
<td></td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A09YANA_RFD</td>
<td>wPt-797974</td>
<td>2D</td>
<td>41.1402</td>
<td>0.000519 - 0.000607</td>
<td>0.378</td>
<td>0.0725 - 0.0923</td>
<td>Ppd-D1</td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A09YANA_RFD</td>
<td>wPt-800147</td>
<td>4A</td>
<td>62.655</td>
<td>0.000804 - 0.000527</td>
<td>0.327</td>
<td>0.0725 - 0.0999</td>
<td></td>
</tr>
</tbody>
</table>
Table A.4: Markers significant for ZAD at more than one site from GAPIT analysis for the DArT and SNP marker datasets. Chromosome and position are from the consensus map. Nearby genes are from CMAP GrainGene database (http://wheat.pw.usda.gov/cmap/) searches.

<table>
<thead>
<tr>
<th>Experiments</th>
<th>Marker</th>
<th>Chr</th>
<th>Dist(cM)</th>
<th>P-value Range</th>
<th>Minor Allele Frequency</th>
<th>FDR Adjusted P-values Range</th>
<th>Nearby Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>W9A10COLE_IRR, W9A10COLE_RFD, W9A10YANA_IRR</td>
<td>woPt-730744</td>
<td>2D</td>
<td>73.0</td>
<td>4.85×10⁻⁵ - 7.91×10⁻⁵</td>
<td>0.161</td>
<td>0.0482 - 0.0787</td>
<td>Ppd-D1</td>
</tr>
<tr>
<td>W9A09COLE_IRR, W9A09COLE_RFD, W9A09YANA_IRR, W9A09YANA_RFD, W9A10COLE_IRR, W9A10COLE_RFD, W9A10YANA_IRR, W9A10YANA_RFD</td>
<td>wsnp_CAP12_c812_428290</td>
<td>2D</td>
<td>57.9</td>
<td>3.91×10⁻¹⁶ - 1.414×10⁻¹⁴</td>
<td>0.134</td>
<td>1.50×10⁻¹² - 5.41×10⁻¹¹</td>
<td>Ppd-D1</td>
</tr>
<tr>
<td>W9A09YANA_IRR, W9A09YANA_RFD</td>
<td>wsnp_BE422566B_Ta_1_2</td>
<td>4B</td>
<td>43.3</td>
<td>0.000169 - 0.000171</td>
<td>0.0582</td>
<td>0.0498 - 0.0507</td>
<td></td>
</tr>
<tr>
<td>W9A09COLE_IR, W9A09COLE_RFD, W9A09YANA_IRR, W9A09YANA_RFD, W9A10COLE_IRR, W9A10COLE_RFD, W9A10YANA_IRR, W9A10YANA_RFD</td>
<td>wsnp_AJ612027A_Ta_2_1</td>
<td>5A</td>
<td>66.2</td>
<td>6.30×10⁻⁹ - 3.74×10⁻⁶</td>
<td>0.360</td>
<td>1.20×10⁻⁵ - 0.00410</td>
<td>Vrn-A1</td>
</tr>
<tr>
<td>W9A09COLE_IRR, W9A09COLE_RFD, W9A09YANA_IRR, W9A09YANA_RFD, W9A10COLE_IRR, W9A10COLE_RFD, W9A09YANA_IRR, W9A09YANA_RFD, W9A10YANA_IRR, W9A10YANA_RFD</td>
<td>wsnp_AJ612027A_Ta_2_5</td>
<td>5A</td>
<td>66.2</td>
<td>1.29×10⁻⁵ - 0.000219</td>
<td>0.151</td>
<td>0.00619 - 0.0764</td>
<td>Vrn-A1</td>
</tr>
<tr>
<td>W9A09COLE_IRR, W9A09COLE_RFD, W9A09YANA_IRR, W9A09YANA_RFD, W9A10COLE_IRR, W9A10COLE_RFD, W9A09YANA_IRR, W9A09YANA_RFD, W9A10YANA_IRR, W9A10YANA_RFD</td>
<td>wsnp_BE404341A_Ta_2_3</td>
<td>5A</td>
<td>66.7</td>
<td>6.07×10⁻⁵ - 0.000105</td>
<td>0.136</td>
<td>0.0258 - 0.0446</td>
<td>Vrn-A1</td>
</tr>
<tr>
<td>W9A09COLE_IRR, W9A09COLE_RFD, W9A09YANA_IRR, W9A09YANA_RFD, W9A10COLE_IRR, W9A10COLE_RFD, W9A09YANA_IRR, W9A09YANA_RFD, W9A10YANA_IRR, W9A10YANA_RFD</td>
<td>wsnp_BF293620A_Ta_2_1</td>
<td>5A</td>
<td>66.2</td>
<td>1.19×10⁻⁶ - 0.000215</td>
<td>0.150</td>
<td>0.000909 - 0.0937</td>
<td>Vrn-A1</td>
</tr>
<tr>
<td>W9A09COLE_IRR, W9A09COLE_RFD, W9A09YANA_IRR, W9A09YANA_RFD, W9A10COLE_IRR, W9A10COLE_RFD, W9A09YANA_IRR, W9A09YANA_RFD, W9A10YANA_IRR, W9A10YANA_RFD</td>
<td>wsnp_BJ224975A_Ta_2_1</td>
<td>5A</td>
<td>66.2</td>
<td>1.53×10⁻⁸ - 4.29×10⁻⁶</td>
<td>0.357</td>
<td>1.47×10⁻⁵ - 0.00410</td>
<td>Vrn-A1</td>
</tr>
<tr>
<td>Sample Name</td>
<td>Snp ID</td>
<td>%A</td>
<td>Vrn Values</td>
<td>Vrn Values</td>
<td>SNP Location</td>
<td>Notes</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>--------------</td>
<td>------</td>
<td>-------------------------------------</td>
<td>-------------------------------------</td>
<td>--------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A09YANA_IRR, W9A09YANA_RFD</td>
<td>wsnp_Bf224975A_Ta_2_2</td>
<td>5A</td>
<td>66.2</td>
<td>7.20×10^-5 - 0.000118</td>
<td>0.140</td>
<td>0.0276 - 0.0450</td>
<td>Vrn-A1</td>
</tr>
<tr>
<td>W9A09COLE_IRR, W9A09COLE_RFD, W9A09YANA_IRR, W9A09YANA_RFD</td>
<td>wsnp_Ex_c22727_31934296</td>
<td>5A</td>
<td>66.7</td>
<td>9.26×10^-10 - 6.31×10^-7</td>
<td>0.359</td>
<td>1.77×10^-6 - 0.00121</td>
<td>Vr-A1</td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A09YANA_IRR, W9A09YANA_RFD</td>
<td>wsnp_Ex_c31799_40545376</td>
<td>5A</td>
<td>69.1</td>
<td>0.000154 - 0.000243</td>
<td>0.369</td>
<td>0.0490 - 0.0581</td>
<td>Vrn-A1</td>
</tr>
<tr>
<td>W9A09COLE_IRR, W9A09COLE_RFD, W9A09YANA_IRR, W9A09YANA_RFD</td>
<td>wsnp_Ex_c31799_40545478</td>
<td>5A</td>
<td>69.1</td>
<td>1.74×10^-6 - 0.000195</td>
<td>0.379</td>
<td>0.0009498 - 0.0747</td>
<td>Vrn-A1</td>
</tr>
<tr>
<td>W9A09COLE_RFD, W9A09YANA_IRR, W9A09YANA_RFD</td>
<td>wsnp_Ex_c7729_13177883</td>
<td>5A</td>
<td>69.1</td>
<td>1.43×10^-6 - 0.000152</td>
<td>0.379</td>
<td>0.000911 - 0.0647</td>
<td>Vrn-A1</td>
</tr>
<tr>
<td>Genotypes</td>
<td>SNP Number</td>
<td>Chromosome</td>
<td>Genotype</td>
<td>Effect Size</td>
<td>p-Value</td>
<td>OR</td>
<td>95% CI</td>
</tr>
<tr>
<td>--------------------</td>
<td>-------------------</td>
<td>------------</td>
<td>----------</td>
<td>-------------</td>
<td>---------</td>
<td>-----------</td>
<td>--------------</td>
</tr>
<tr>
<td>W9A09COLE_RFD,</td>
<td>wsnp_Ex_rep_c66689_65010988</td>
<td>5A</td>
<td>66.7</td>
<td>0.000126 - 0.000203</td>
<td>0.400</td>
<td>0.0438 - 0.0518</td>
<td>Vrn-A1</td>
</tr>
<tr>
<td>W9A09YANA_IRR,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09YANA_RFD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A09YANA_IRR,</td>
<td>wsnp_Ex_c12048_19288999</td>
<td>5B</td>
<td>71.1</td>
<td>3.83×10^-5 - 0.000223</td>
<td>0.266</td>
<td>0.0291 - 0.0533</td>
<td>Vrn-B1</td>
</tr>
<tr>
<td>W9A09YANA_RFD,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10COLE_RFD,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10YANA_IRR,</td>
<td>wsnp_Ra_c20970_30293078</td>
<td>5B</td>
<td>71.1</td>
<td>0.000332 - 0.000433</td>
<td>0.272</td>
<td>0.0747 - 0.0976</td>
<td>Vrn-B1</td>
</tr>
<tr>
<td>W9A09YANA_RFD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10YANA_IRR,</td>
<td>wsnp_Ra_c20970_30293227</td>
<td>5B</td>
<td>71.1</td>
<td>6.16×10^-5 - 0.000183</td>
<td>0.262</td>
<td>0.0383 - 0.0499</td>
<td>Vrn-B1</td>
</tr>
<tr>
<td>W9A10YANA_RFD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9A10YANA_IRR,</td>
<td>wsnp_Ex_c508_1008029</td>
<td>5D</td>
<td>61.0</td>
<td>3.90×10^-5 - 0.000332</td>
<td>0.202</td>
<td>0.0258 - 0.0978</td>
<td>Vrn-D1</td>
</tr>
<tr>
<td>W9A10YANA_RFD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


Cockerham CC. (1954). An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics, 39*(6), 859.


Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, & Sham PC. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics, 81*(3), 559-575.


Technow F, Bürger A, & Melchinger AE. (2013). Genomic prediction of northern corn leaf blight resistance in maize with combined or separated training sets for heterotic groups. G3: Genes|Genomes|Genetics, 3(2), 197-203.


